

Ritu Bhalla · Kothandaraman Narasimhan ·  
Sanjay Swarup

## Metabolomics and its role in understanding cellular responses in plants

Received: 13 May 2005 / Revised: 22 July 2005 / Accepted: 25 July 2005 / Published online: 12 October 2005  
© Springer-Verlag 2005

**Abstract** A natural shift is taking place in the approaches being adopted by plant scientists in response to the accessibility of systems-based technology platforms. Metabolomics is one such field, which involves a comprehensive non-biased analysis of metabolites in a given cell at a specific time. This review briefly introduces the emerging field and a range of analytical techniques that are most useful in metabolomics when combined with computational approaches in data analyses. Using cases from *Arabidopsis* and other selected plant systems, this review highlights how information can be integrated from metabolomics and other functional genomics platforms to obtain a global picture of plant cellular responses. We discuss how metabolomics is enabling large-scale and parallel interrogation of cell states under different stages of development and defined environmental conditions to uncover novel interactions among various pathways. Finally, we discuss selected applications of metabolomics.

**Keywords** Functional genomics · Metabolic profiling · Pathways · Systems biology

---

This special review article is dedicated to the commemoration of the retirement of Dr. Oluf L. Gamborg after 25 years of service as Founding Managing Editor of Plant Cell Reports. RB and KN have contributed equally to this review.

---

Communicated by P. P. Kumar

---

R. Bhalla  
Temasek Life Sciences Laboratory, National University of Singapore,  
Singapore

K. Narasimhan  
Department of Obstetrics and Gynecology, National University Hospital,  
Singapore

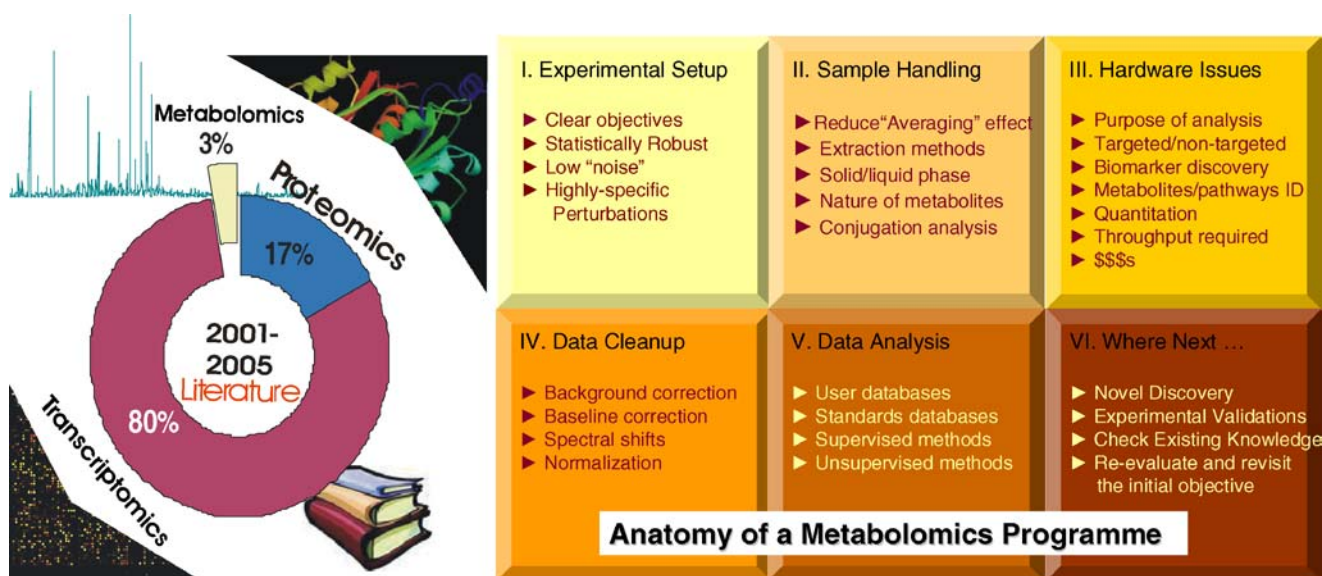
S. Swarup (✉)  
Department of Biological Sciences, National University of Singapore,  
Singapore  
e-mail: dbsss@nus.edu.sg

Nature produces an astonishing wealth of metabolites with important biological functions from both primary and secondary metabolic pathways in plants. Metabolites are the small molecules produced as intermediates and end-products of all metabolic processes. It is estimated that more than 100,000 secondary metabolites are produced by plants, while the total number is estimated to exceed over 500,000 (Hadacek 2002). To assess this diversity of structurally complex chemical compounds, various approaches have been initiated in the last decade largely due to the tremendous advances in the instrumentation and data handling capabilities. Large scale analysis of small molecules is called *metabolomics* (See Box 1 for terminologies and types of metabolomics analyses). In this review, we briefly describe some of the metabolomics technologies, data handling considerations, metabolomics as a tool for studying biochemical phenotypes of cells—either singly or in combination with other genomics approaches, its use in understanding cellular responses and in uncovering silent phenotypes, metabolomics of selected pathways and finally other applications of metabolomics. Along the way, readers are referred to many excellent reviews on topics not covered in detail here.

---

### Box 1 “Metabolomics” and Related Terminologies

Most commonly adopted interpretations of terms associated with this field are presented (Fiehn 2002; Hall et al. 2002; Sumners et al. 2003). *Target analysis* is restricted to the substrate and/or the direct product of a specific metabolic step. This form of metabolite analysis avoids interfering compounds and utilizes specialized protocols. A second approach is *metabolite* (or *metabolic*) *profiling*, where analysis is restricted to the identification and quantification of a selected number of pre-defined metabolites in a biological sample. Sample preparation and cleanup are focused on the chemical properties of these compounds. Instead of separating individual metabolites by physical parameters, focus is on collecting and analyzing



**Fig. 1** Emergence of metabolomics as a functional genomics platform and the anatomy of a robust metabolomics programme. Relative number of PubMed abstracts (out of 1,340 total) on the three major functional genomics platforms (as of April, 2005). Components

of a well planned and systematically performed metabolomics programme are listed. Most components are customized based on the objectives, scope and resources of the programme

data from crude metabolite mixtures to rapidly classify samples. Lastly, the complete form of metabolite analysis is called *metabolomics*, where the ultimate goal is of unbiased identification and quantification of all metabolites present in a sample from an organism grown under defined conditions. Among all these forms of analyses, metabolomics is the most promising approach in investigating metabolic networks and unraveling cellular processes, as it focuses on quantifying several metabolites individually hence defining their quantity and quality. Another term, metabonomics is used frequently in the biomedical (toxicology) literature and for methods involving nuclear magnetic resonance spectroscopy. However, "metabolomics" is a preferred term for unbiased metabolite analyses. A combination of metabolomics analysis with transcriptomics and proteomics gives the holistic view of the complex interactions between genes and metabolites (Fig. 1). Advances in sampling, analyzing and computational technologies will lead to a greater understanding of metabolism than was possible using other profiling approaches limited to a particular class of compounds. When combined with theoretical considerations and a modeling-based prediction of outcomes of changes, metabolomics can provide a pillar for *systems biology*. This field holds promise for applications such as metabolic engineering.

### Metabolomics technologies: Sample preparation and instrumentation

Please refer to items in Box 1 for some of the techniques available for metabolomics analyses. It is a pre-requisite for metabolomics analysis that careful consideration be made of the methods employed for tissue extraction, sample preparation, data acquisition, and data mining (Fig. 1).

Some practical steps that need to be kept in mind for the sample preparation include (i) the importance of rapidly stopping the inherent enzymatic activity of biological samples, (ii) the state of the tissue at the time of extraction, (iii) reduction of "averaging effects" due to mixtures of tissues or cell-types, (iv) the storage conditions of the tissue, although it is believed that under cryopreservation conditions the metabolic state of the samples can be maintained, but this may not hold for more labile or volatile compounds, (v) the choice of extraction buffers has to be compatible with the final objectives of the analysis, and (vi) development of robust procedures combined with minimal handling or pre-processing of samples before chromatography and/or mass spectrometry analyses.

Recent advances in metabolomics analysis owe primarily to improvements in the mass spectrometry (MS) technology that has resulted in formats that are more user-friendly and amenable to biologists. Additionally, combination of mass spectrometry with in-line gas or liquid chromatography has increased the efficiency of separation of molecules. Such multivariate spectrometric detection seems advantageous for quantifying and identifying as many individual compounds as possible in mixtures of metabolites. Gas chromatography-MS (GC-MS) and HPLC-MS remain the popular choice for quantitative and qualitative metabolite profiling. However, most of the metabolites found in plant extracts are too nonvolatile to be analyzed directly by GC methods. The compounds have to be converted to less polar, more volatile derivatives before they are applied to the GC column. Metabolomics analysis by gas chromatography/mass spectrometry (GC/MS) is complementary to approaches using infrared spectroscopy (Johnson et al. 2000). An improved chromatographic resolution can be achieved by using monolithic silica capillary columns. For example, C-18 monolithic silica capillary columns in HPLC

coupled with ion trap mass spectrometry when used for the detection of metabolome of *Arabidopsis* showed reduced ionization suppression by enhanced chromatographic resolution (Tolstikov et al. 2003).

Nuclear Magnetic Resonance (NMR) is another potentially very useful technique, since in-principle any chemical species that contains protons gives rise to signals (Krishnan et al. 2005). A crude biological extract, however, has many peaks, leading to overlapped signals and complex profiles or “fingerprints.” Two-dimensional NMR is often necessary to assign the signals, but one-dimensional <sup>1</sup>H-NMR spectra have been of great value as shown in many medical applications by Lindon and Nicholson (1997). Also NMR is used for metabolite fingerprinting method for plants, where the aim is to look for compositional similarities and explore the overall natural variability (Fiehn 2002). Some metabolite selection occurs in all methodologies, from initial solvent extraction through chromatography to MS ionization. Nevertheless, new advances have been made such as the use of scanning time-of-flight (TOF)-MS coupled with GC separation and integrated with peak deconvolution software. This technique can increase the number of metabolites detectable by GC-MS in crude plant extracts from 500 to 1,000. In one example of the use of this technique, the biochemical mode of action for herbicides and other bioactive compounds was rapidly and simultaneously classified by automated pattern recognition of the metabolome that is embodied in the <sup>1</sup>H NMR spectrum from crude plant extracts (Ott et al. 2003). The spectra were classified by artificial neural network analysis to discriminate herbicide modes of action. Such combination of NMR metabolite profiling and neural network classification is expected to be similarly relevant to other metabolomic applications discussed here.

A recently introduced technique called Fourier Transform Ion Cyclotron Mass Spectrometry (FTMS) can be used to study the phenotypic changes associated with metabolism (Aharoni et al. 2002; Baud-Camus et al. 2001; see review by Brown et al. 2005). FTMS is capable of non-targeted metabolic analysis and suitable for rapid screening of similarities and dissimilarities in large collections of biological samples, e.g., plant mutant populations and metabolite compositions in wine (Cooper and Marshall 2001). Separation of the metabolites is achieved solely by ultra-high mass resolution. Identification of the putative metabolite or class of metabolites to which it belongs can then be achieved by determining the elemental composition of the metabolite based upon the accurate mass determination. Relative quantitation is achieved by comparing the absolute intensities of each mass using internal calibration. Integrating the observed metabolic alterations into hypotheses about changes in biochemical pathways and gene expression levels are the next goal. In one such application, tobacco plants expressing a petunia myb regulator with altered petal color showed metabolite differences and the FTMS analysis revealed the metabolite to be cyanidin-3-rhamnoglucoside, a known flower pigment (Aharoni et al. 2002). In the same study, the authors applied the FTMS technology to study development stage-specific

accumulation of metabolites in strawberry. This technology requires specialized skills and equipments not easily accessible to many researchers; however, based on the literature it seems to be gradually gaining accessibility partly due to the adoption of the FTMS technology in proteomics analyses.

---

### Computational tools for metabolomics data analysis

Due to the large datasets arising from metabolomics analyses, a number of tools are available for computational needs and some others are under development. The metabolomics spectral formatting, alignment and conversion tool (MSFACTs) is one such example (Duran et al. 2003). This program is used for automated import, reformatting, alignment, and export of large chromatographic data sets to allow more rapid visualization and interrogation of metabolomics data. MSFACTs has been used in the processing of GC/MS metabolomics data from different tissues of the model legume plant, *Medicago truncatula*, the various tissues such as roots, stems, and leaves from the same plant could easily be differentiated based on metabolite profiles. Furthermore, similar types of tissues within the same plant, such as the first to eleventh internode of stems could also be differentiated based on metabolite profiles and by using this tool. Another computational tool, COMSPARI (COMparison of SPectral Retention Information) has been developed to facilitate the identification of minor compounds in complex mixtures by GC/MS and LC/MS (Katz et al. 2004). The processed data from metabolomics analyses can then be taken for statistical analyses. A number of excellent tools are available, such as—multivariate packages of SAS (SAS Institute, Cary, NC), Pirouette (Infometrix, Woodinville, WA), or MATLAB (Mathworks, Inc., Natick, MA). These packages allow both processing and display of relationships in the datasets through hierarchical cluster analysis (HCA), two-dimensional, and three-dimensional (2D/3D) or principal component analysis (PCA). Setting-up of the data sets can, therefore, allow biomarkers to be discovered as well as allow key differentiating molecules to be identified.

Arising from the need to handle large datasets together with experimental details, development of ArMet (www.armet.org) was reported, which is a framework for the description of plant metabolomics experiments and their results (Jenkins et al. 2004). This is a data handling tool for plant metabolomics and allows formal data descriptions, which specify the full experimental context, enable principled comparison of data sets, allow proper interpretation of experimental results, permit the repetition of experiments and provide a basis for the design of systems for data storage and transmission. It is aimed at providing a starting point for the development of community data standards for the metabolomics researchers.

Several multivariate methods for classifying and modeling analytical data have been used to evaluate metabolite spectral databases. Among these, simple unsupervised clustering algorithms and principal components based analyses enable visualization of biological data sets based on

the inherent similarity/dissimilarity of samples with respect to their biochemical composition. PCA-based methods usually constitute the first step in evaluating metabolomics data. It works extremely well for detecting patterns, trends, and groups among samples in multivariate data sets as well as giving an explanation to these sample distributions in terms of variable influence and efficient means for detecting and explaining deviating samples. Recently, a weighted PCA (WPCA) model has been put forth to translate spectra of repeated measurements into weights describing experimental errors thus focusing more on the natural variation in the data (Jansen et al. 2004). Once a relationship between the metabolic profile and phenotype has been identified from the initial analysis, supervised approaches such as partial least squares (PLS) algorithm for data reduction based on multiple component analysis (SIMCA) and partial least square—discriminant analysis (PLS-DA) may be implemented with a view to maximizing the separation between classes and identifying robust markers. Another commonly used algorithm for understanding metabolite relationships is “Self Organizing Maps” (SOM) (Kohonen et al. 1997). SOM analysis can be done to cluster complex data efficiently, which has proven to be an excellent tool for analyzing global characteristics of genome sequences and for revealing key combinations of oligonucleotides representing individual genomes (Abe et al. 2003). SOM analysis is widely used as a data mining and visualization method for complex data sets, image processing and speech recognition, process control, economical analysis, and diagnostics in pharmaceuticals, medicine and more recently in plant metabolomics (Kohonen et al. 1997; Markey et al. 2003).

### Text-mining for metabolomics researchers and integration with genomics information

A novel angle presented to the metabolomics researchers is how to deal with the tremendous increase in the information generated from the above procedures and link it with the knowledge available in the existing literature. Till recently, there were no text mining systems dedicated for plant biology literature. We have recently developed one such comprehensive system called the Dragon Plant Biology Explorer (DPBE) for knowledge extraction from PubMed documents that integrates information from literature with the genome and metabolome based information to produce interactive networks of associations (Bajic et al. 2005) (<http://research.i2r.a-star.edu.sg/DRAGON/ME2/>). It can be used to interrogate various types of cellular and other plant processes. The tool has been set up such that it can also help uncover the pharmacological effects of plant natural products. Users provide the direct output of their PubMed abstracts arising from their keyword searches. The DPBE tool can handle tens of thousands of abstracts in routine analyses. Output is presented both in tabular and interactive networks format.

Co-occurrence of terms in controlled lists of vocabularies is used to develop all the relationships. This tool uses a combination of lists of gene ontology terms and three

new lists developed for metabolome analyses as part of its controlled vocabularies. The “Metabolome Explorer” module contains the lists based on metabolic pathways, enzymes and metabolites derived from AraCyc, BRENDA, KEGG, and other metabolism databases. Metabolomics researchers have two options for analyses (i) use of the “Metabolome Explorer” module as described above or (ii) use of the functional genomics input module. In the latter case, PubMed abstracts can be collated based on keyword searches for each of the key metabolite uncovered from the metabolomics study. Lists can be then selected from the ones listed above and combined with others on genes, mutations, gene functions or cellular processes or plant parts. Hence, a comprehensive picture is obtained linking metabolites to gene products and other features of plant form and function. An example of an analysis by DPBE of PubMed abstracts resulting from keyword searches of the flavonoids “genistein,” or “rutin” is shown in Fig. 2. A part of the network is shown to highlight how the network display allows visualization of multiple aspects of the research field to be linked. Each node can be clicked in the online version to display the set of corresponding abstracts. Vocabulary terms found in the abstracts are also appropriately highlighted in color coded forms to allow rapid visual screening and quickly summarize the relevant information.

### Metabolomics as a link between genotype and biochemical phenotype

Even though the *Arabidopsis* genome has been completely sequenced, over 30% of its genes cannot be annotated by homology to genes in other organisms, and only 9% have been experimentally characterized (Haas et al. 2005) (Table 1). Moreover, of the nearly 25% genes believed to be involved in plant metabolism, most functional characterizations are not based upon rigid biochemical testing. Genome analysis has allowed uncovering of multiple forms of enzymes, which are important for generating the biochemical diversity as discussed earlier. Such a role of

**Table 1** Genes encoding enzymes and their regulators in the *Arabidopsis* genome: A summary

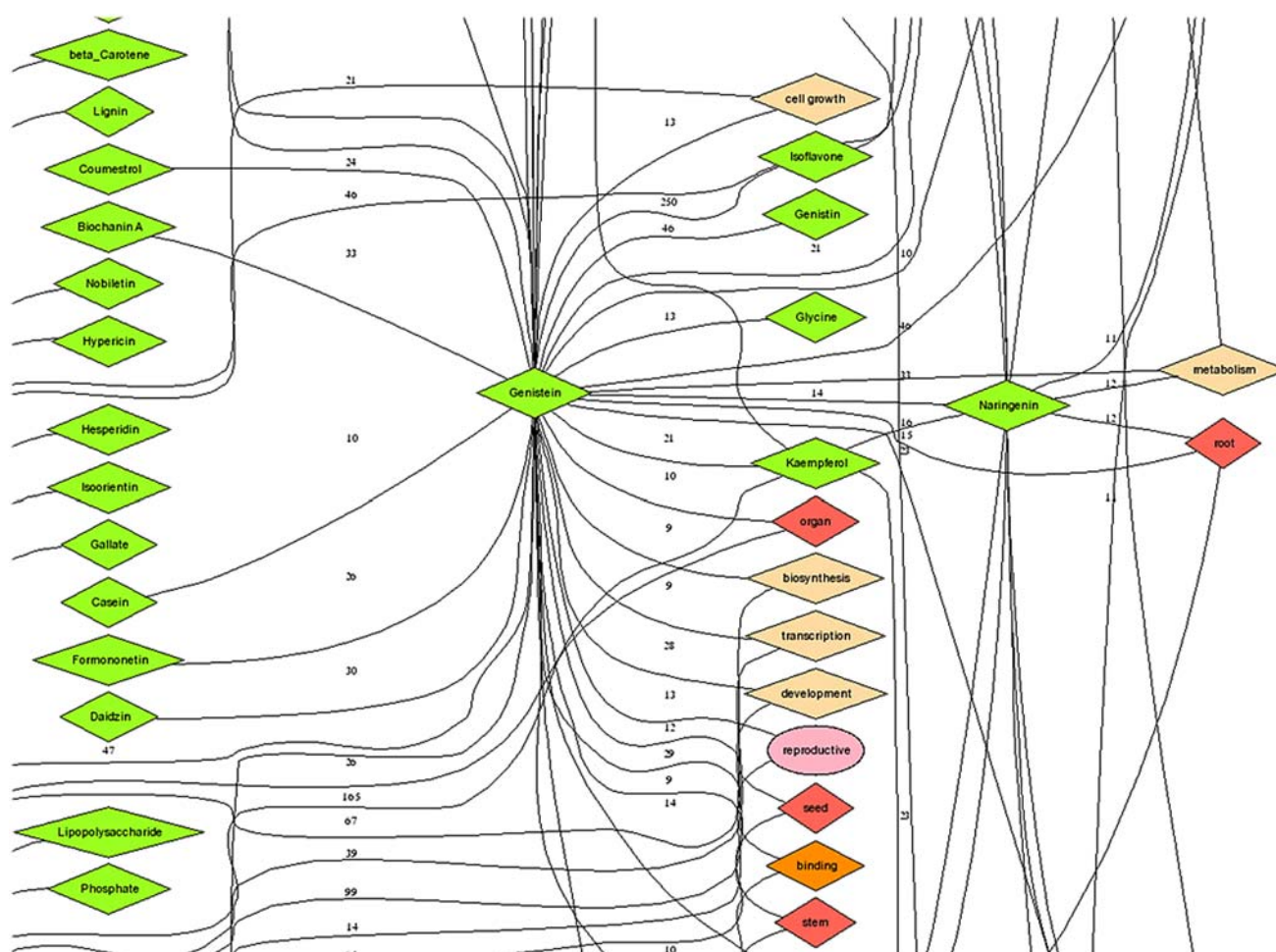
| Category <sup>a</sup>     | Gene ontology category | Number of genes <sup>b</sup> | Percentage of total <sup>c</sup> |
|---------------------------|------------------------|------------------------------|----------------------------------|
| Catalytic activity        | 0003824                | 6,604                        | 25.2%                            |
| Hydrolase                 | 0016787                | 2,281                        | 8.7%                             |
| Kinase                    | 0016301                | 1,281                        | 4.9%                             |
| Transferase               | 0016740                | 1,688                        | 6.4%                             |
| Others                    | –                      | 1,354                        | 5.2%                             |
| Enzyme regulator activity | 0030234                | 185                          | 0.7%                             |

Adapted from Haas et al. (2005)

<sup>a</sup>Selected categories of molecular function related to metabolism

<sup>b</sup>Number of genes corresponding to the selected categories shown in column 1

<sup>c</sup>Percentage values are based on a total of 26,207 protein coding genes



**Fig. 2** An association network resulting from the analysis of PubMed abstracts retrieved using keywords for the flavonoids, “genistein” or “rutin,” by the Dragon Plant Biology Explorer (<http://research.i2r.a-star.edu.sg/DRAGON/ME2/>). The “Metabolome Ex-

plorer” module was chosen and three metabolism-related lists as well as one for anatomy were used for the analysis. A part of the network is shown

metabolomics in linking genotype with phenotype has been reviewed earlier (Fiehn 2002). Together with metabolite analyses, substrate specificities and catalytic activities can be studied for the isoforms of various enzymes. Such overlapping catalytic activities towards different substrates have been reported for the glucosyltransferase genes in *Arabidopsis* (Lim et al. 2000) and *O*-methyltransferase genes in *Thalictrum tuberosum* (Frick and Kutchan 1999). Overlapping expression pattern of the genes for various isoforms has been reported for numerous cases, including the alcohol dehydrogenase gene family members in *Vitis vinifera* (Tesniere and Verries 2001). In many cases, gene duplications give rise to redundant (or partially redundant) functions, leading to silent phenotypes. Mutational approaches also may not uncover the function of such genes. Hence, adoption of metabolomics approaches together with other functional genomics approaches is highly desirable to provide a more comprehensive view of the cellular processes.

The primary aim of “omics” technologies is the non-targeted identification of all gene products produced directly or indirectly (i.e. transcripts, proteins and metabo-

lites) present in a specific biological sample (Fridman and Pichersky 2005). In one of the early examples of comprehensive analyses of plant metabolites, GC-MS analysis was used to probe the metabolism of *Arabidopsis* leaves and close to 326 distinct compounds were distinguished (Fiehn et al. 2000a). A chemical structure could be assigned to roughly half of these compounds. Metabolomics approach was also used in distinguishing the profiles in compounds of two mutants—one metabolic and one developmental—in different ecotype backgrounds. The *dgd1* mutant that has reduced digalactosyldiacylglycerol accumulation in the Col-2 background and the *sddl-1* mutant that has reduced stomatal density in the C24 background were profiled (Fiehn et al. 2000b). The *Arabidopsis* ecotypes were found to be highly divergent and were distinguishable from their respective mutants. An interesting finding was that “metabolic phenotypes” of the two ecotypes were more divergent, as judged from the principle component analysis, than the mutants were from their respective parental ecotypes. A detailed statistical evaluation of the data set showed that there were 41 significant changes in the *sddl-1* mutant; two of the most significant changes were in

unknown hydrophilic substances. In the case of the *dgd1* mutation, the changes were even more pleiotropic with a total of 153 significant changes being recorded in comparison to the parental ecotype.

A further example of the ability of metabolic profiling to reveal unexpected changes associated with genes has come from efforts to produce a biodegradable plastic, polyhydroxybutyrate, in *Arabidopsis* (Bohmert et al. 2000). One of the aims of the study was to achieve high levels of polyhydroxybutyrate in *Arabidopsis* leaves. A novel pathway was introduced for polyhydroxybutyrate synthesis consisting of three genes from the plant-associated bacterial species, *Ralstonia eutropha*. The approach was successful in generating the highest levels of polyhydroxybutyrate hitherto recorded in plants, up to 4% of the leaf fresh weight; however, routine metabolic profiling of the *Arabidopsis* leaves revealed rampant pleiotropic changes in organic acids, amino acids, sugars and sugar alcohols. The origin and full significance of these changes currently remain a mystery; however, it is unlikely that they would have been identified if nor for the broad metabolic profiling strategies available currently.

---

### **Integrated functional genomics approaches in metabolic pathways studies**

Combinations of various functional genomics platforms have been used to obtain more comprehensive views of metabolic pathways networks and their regulation by gene or protein expression. One of the earliest examples was in the case of the yeast galactose utilization (GAL) pathway, where a combination of genomics and proteomics approaches were used to study the effects of metabolic perturbations using a series of metabolic mutants (Ideker et al. 2001). Several new control points and interactions with other pathways were detected. In another example, the metabolic pathway for the production of the cholesterol lowering (polyketide-derived secondary metabolite) drug, lovastatin, was studied in *Aspergillus terreus* to engineer its pathway (Askenazi et al. 2003). A combination of microarray and metabolomics methods was used to identify novel controls of lovastatin biosynthetic pathway, based on which, genetic tools were designed used for strain improvement.

Integrated approaches have begun to be used in plants more recently mainly to study the responses to nutritional or biotic stresses. In one such case, gene-to-metabolite networks were described for regulation of sulfur and nitrogen nutrition and secondary metabolism in *Arabidopsis* (Hirai et al. 2004). In a follow-up paper, the group used batch-learning SOMs for the purpose of identifying novel functions (Hirai et al. 2005). Using their approach, novel desulfoglucosinolate sulfotransferases were identified to be affected by nutritional stress. Biosynthetic pathways leading to plant defense response and mechanism towards biotic stresses have been studied using combined metabolomics and transcriptomics approaches as discussed below (Kant et al. 2004).

One of the early and more intensive studies using integrated approaches in ornamental plants was on the regulation of floral scent production in petunia revealed by targeted metabolomics (Verdonk et al. 2003). This study was carried out by applying solid phase microextraction (SPME) techniques coupled to GC-MS analysis. Volatile emission was monitored *in vivo* using a targeted metabolomics approach. Mature flowers released predominantly benzenoid compounds of which benzaldehyde, phenylacetaldehyde, methylbenzoate, phenylethylalcohol, iso-eugenol and benzylbenzoate were most abundant. DNA-microarray analysis revealed that genes of the pathways leading to the production of volatile benzenoids were upregulated late during the day, preceding the increase of volatile emission. RNA-gel blot analyses confirmed that the levels of phenylalanine ammonia lyase (PAL) and S-adenosyl methionine (SAM) synthase transcripts increased towards the evening.

To take integrated approaches one step further, a set of protocols have been described to handle sequential extraction of metabolites; proteins and RNA from the same samples, allowing the data to become suitable for multivariate data analysis. A detection of 652 metabolites, 297 proteins and clear RNA bands in a single *Arabidopsis thaliana* leaf sample was validated by the authors (Weckwerth et al. 2004a). Development of such standardized protocols is going to help researchers integrate multiple functional genomics platforms.

---

### **Use of metabolomics in studying plant stress responses**

Metabolomics is being increasingly used for understanding the cellular phenotypes in response to various types of stresses- biotic or abiotic. In addition to the examples discussed in the above Section on integrated approaches, some others are discussed here. In one recent study of sulfur deficiency response, general metabolic readjustment was found (Nikiforova et al. 2005). Mutual influences were found between sulfur assimilation, nitrogen imbalance, lipid breakdown, purine metabolism, and enhanced photorespiration. A general reduction of metabolic activity was seen under conditions of depleted sulfur supply. These observations together with those of Hirai et al. (2005) discussed in the previous Section are likely to advance the field of nutritional stress response further. Metabolomics has also been applied to the case of cold stress response especially in the pathway involving the central regulator CBF (Cook et al. 2004). A total of 325 metabolites were upregulated in cold-treated *Arabidopsis* ecotype Ws-2 plants. Of these, 256 (79%) also increased in non-acclimated Ws-2 plants in response to overexpression of C-repeat/dehydration responsive element-binding factor (CBF)3. As in the case of sulfur deprivation response, in the cold response also, extensive reconfiguration was seen. In the case of biotic stress, biosynthetic pathways involved in plant defense response and mechanism have been studied using a combination of metabolomics and transcriptomics approaches (Kant et al.

2004). Study of induced direct and indirect defense responses in tomato plants to spider mite infestation were done using these approaches and were able to detect differential timing of the two responses at very early stages.

---

### Uncovering silent phenotypes of mutations

Metabolome data can be used to reveal the phenotype of silent mutations. Plant genome contains thousands of genes, of which many produce silent phenotypes upon mutation. Intercellular concentrations of metabolites can reveal phenotypes of proteins active in metabolic regulation. Quantification of several metabolite concentrations relative to the concentration of one selected metabolite can reveal the site of action in the metabolic network of a silent gene. Similarly comprehensive analysis of metabolite concentrations in mutants, providing “metabolic snapshots” can reveal functions when snapshots from strains deleted for unstudied genes are compared to those deleted for known genes. This strategy has been successfully applied in yeast by Raamsdonk et al. (2001) and they have named it as “FANCY”—Functional analysis by co-response in yeast. In plants, connections of metabolic networks were analyzed for a silent potato plant line suppressed in expression of sucrose synthase isoform II (Weckwerth et al. 2004b). Despite the silent phenotype, metabolic perturbations were identified in carbohydrate and amino acid metabolic pathways even when no differences in average metabolite levels were found.

---

### Targeted metabolomics analyses of selected pathways

We briefly discuss a selected secondary metabolism pathway based on the diversity of its metabolites in the model plant *Arabidopsis*. The term “phenylpropanoid” is mostly used to refer to any compound bearing a 3-carbon chain attached to 6-carbon aromatic ring (C<sub>6</sub>–C<sub>3</sub> compounds). Majority of the phenylpropanoids are derived from cinnamic or *p*-coumaric acids. Plants have the unique ability to divert large amounts of carbon from aromatic amino acid from shikimate pathway metabolism into the biosynthesis of natural products based on a phenylpropane skeleton. These diverse phenylpropanoid compounds which include flavonoids, lignin, coumarins and many small phenolic molecules, have multiplicity of functions in structural support, pigmentation, defense and signaling. Biosynthesis of phenylpropanoid compounds is not only activated in specific tissues and cell types, but also in response to environmental stresses such as by wounding, pathogen infection, and/or UV irradiation.

Recent work has shown that plants exude large amounts of secondary metabolites of which phenylpropanoids form the majority. Metabolomics analysis was recently described for the root exudates of *Arabidopsis* (Narasimhan et al. 2003). Of the 149 hydrophobic compounds identified in the exudates, 125 were secondary metabolites and 76% of these were phenylpropanoid compounds. Secondary metabolite

profiles of some of the metabolic mutants in the study were significantly different from that of the wild type plants (Fig. 3). These observations taken together with those from stress responses discussed above point to the fact that metabolic pathways can undergo system level readjustments in response to genetic or external perturbations.

A noteworthy development in the field of metabolomics has been its integration with the quantitative genetics field. In one such series of studies, Kliebenstein et al. (2001a,b) carried out investigations on the inheritance patterns of glucosinolates in *Arabidopsis* and found it to be under quantitatively inherited. It will be interesting to note further developments in the merger of these two fields as it will have a bearing on the nutritional improvement of many crops.

While we have mostly used examples drawn from the model plant *Arabidopsis*, some other plant systems have emerged as highly suitable models for genomics and metabolomics work. Legumes are a good source of isoflavonoids and other valuable phenylpropanoid compounds. These compounds include isoflavones such as genistein, diadzein flavanones such as hesperetin and others. They are also important in agriculture as key metabolites involved plant-microbe interaction. *Medicago truncatula* is the model plant for molecular and genetic studies of legumes. Legumes are important agricultural protein crops, and contribute to the biological fixation of nitrogen. Legumes offer unique opportunities for the study of plant-microbe interactions. Furthermore, legumes synthesize several unique (pharmaceutical interesting) compounds, such as iso-flavonoids, triterpenes, alkaloids, and phytosterols.

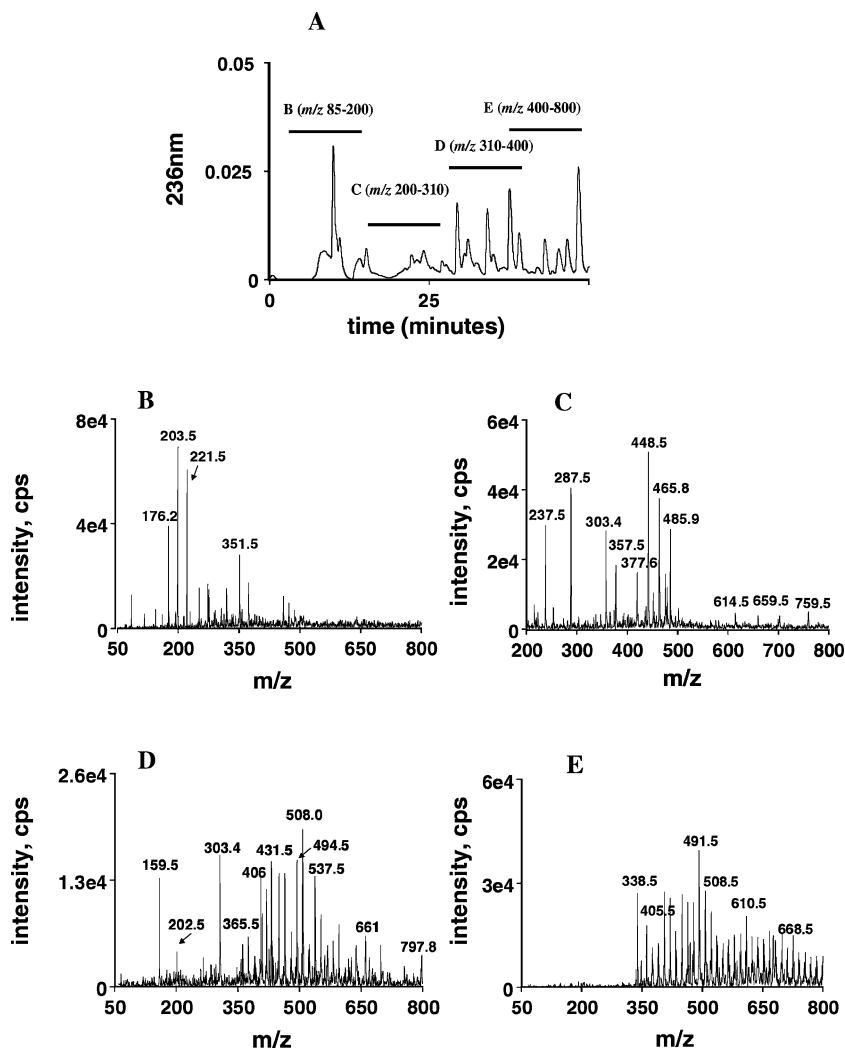
One major application of metabolomics in studies associated with legumes is on nodule cell differentiation. Unlike roots, which acquire a range of essential mineral nutrients from the soil, nodules are specialized in nitrogen acquisition and are produced only when mineral nitrogen is limiting. Such specialization might be expected to be reflected by a more streamlined metabolism, carried out by fewer proteins and orchestrated by fewer genes. A GC-MS based approach was used to profile metabolites present in nodules and other organs of *Lotus* (Raamsdonk et al. 2001). Detailed analysis of nodule and root metabolite profiles will help to identify novel aspects of nodule metabolism that underpin symbiotic nitrogen fixation. For example comparison of root and nodule transcript levels in *Lotus* uncovered nodule-induced genes encoding enzymes that have not previously been associated with symbiotic nitrogen fixation. Quantification of the reactants and products of these enzymes may provide some insight into their importance in nodules. Metabolomics is also likely to prove useful in analyzing the nodule phenotypes of symbiotic interactions formed with either mutants of bacteria or plants.

---

### Other applications of metabolomics approaches

A number of applications of metabolomic analyses, where genetics could play a major role, can be imagined. Some are more obvious—such as (i) metabolic engineering of

**Fig. 3** Metabolomics analysis of root exudates from *Arabidopsis Landsberg erecta* ecotype (wild type) using a combination of HPLC and mass spectrometry (MS). **A** Distribution of phenolic compounds in the HPLC profiles is uncovered by mass spectrometry analysis of individual fractions. Related compounds generally elute in HPLC in adjacent fractions **B** indoles and lignin monomers; Indole-3-acetic acid ( $m/z$  175.18), syringaldehyde ( $m/z$  182.18), camalexin ( $m/z$  202.5) **C** flavonols and flavonoid aglycones; kaempferol ( $m/z$  286.24) quercetin ( $m/z$  302.4) **D** indole and glucosinolate conjugates; methyl IAA glucose ( $m/z$  365), sinapoyl glucose ( $m/z$  386) and **E** flavonol and cyanidin glucosides: cyanidin glucoside (507.5) and quercetin 3-O-rutinoside ( $m/z$  633). Spectra were taken in the positive mode of an electrospray ionization triple-quadrupole mass spectrometer



valuable biochemical pathways; (ii) enhancing the nutritional value of foods; (iii) decreasing the need for pesticide or fertilizer application or (iv) engineering of pathways needed for the production of pharmaceuticals in plants (Giddings et al. 2000). Other fields of applications are less obvious. For example, metabolomics studies could be applied for assessing the substantial equivalence of genetically modified organisms if the metabolic phenotypes of a variety of well-known cultivars (that are commonly believed to be safe) are compared to transgenic plants. In addition, metabolomics can have a deep impact in understanding metabolism, for example, for the prediction of novel metabolic pathways, and to describe cellular networks in vivo. Three such applications are briefly discussed here.

#### Identification of uncommon phytochemicals

Metabolomics can be used to identify uncommon plant metabolites. In one case, GC/MS method was used for qualitative and quantitative detection of 150 compounds in *Arabidopsis* (Fiehn et al. 2000a). Fifteen uncommon plant metabolites were identified later from these compounds

(Fiehn et al. 2000b). A study of selected ecotypes and mutants showed the power of this method for functional genomics (Fiehn et al. 2000a). The number of metabolites that can be identified in a single analysis has been growing and more than 500 compounds can be identified currently using techniques such as FT-MS as described above. Novel transferases in the glucosinolate pathway were recently described using the FTMS approach (Hirai et al. 2005).

#### Classification and quality control of phytomedicines

Recently, phytomedicines have acquired great importance in the drug industry. To improve the accuracy and consistency of control phytomedicines worldwide new analytical methods for their stricter standardization are being employed. Such methods are both objective and robust, addressing the reproducibility of the content of the chemical profiles. These requirements can be met by NMR-based metabolomics which combines high resolution  $^1\text{H-NMR}$  spectroscopy with chemometrics analyses. In one such application, chamomile flower extracts were analyzed from three different



geographical regions and used to study the effects of regional differences on composition of the extracts (Wang et al. 2004). This metabolomics strategy can prove to be an efficient tool for the quality control and authentication of phytomedicines.

#### Assessment of the substantial equivalence of genetically modified organisms

One distinct field of applications includes assessment of the substantial equivalence of genetically modified organisms (World Health Organization 2000). Metabolic phenotypes of a variety of well-known cultivars that are commonly believed to be safe are compared with those of transgenic plants to demonstrate substantial equivalence. Chemical fingerprinting has been used for the evaluation of unintended secondary metabolic changes in transgenic food crops. An off-line combination of 400 MHz proton (<sup>1</sup>H)-NMR spectroscopy and liquid chromatography was used for the multi-component comparison of low molecular weight compounds in complex plant matrices (Noteburn et al. 2000). Metabolomics techniques involving NMR spectroscopy were used to identify any unintended effects of modifications (Le Gall et al. 2003). In a recent review, a detailed discussion is provided of unintended effects and their detection in genetically modified plants using recent approaches such as metabolomics (Cellini et al. 2004).

#### Conclusions

Recent advances in analytical chemistry instrumentation combined with data handling capabilities has led to the adoption of metabolomics by a growing number of biologists, including plant scientists. Metabolomics, therefore, has emerged as the third component of functional genomics, following the prior starts in the fields of genomics and proteomics. It is helping plant biologists to understand a number of plant processes and study responses by combining genomics and biochemical phenotyping capabilities. Such integrated approaches are not only helpful in assigning functions to a large class of function-unknown (or *FUN*) genes and their interactions with other pathways, but are also useful in applications such as metabolic engineering and assessment of genetically modified plants. As part of a more recent emerging area, robust and carefully collected data generated from metabolomics can be combined with computationally-intensive approaches based on modeling of pathways to steer this field towards systems biology, which promises to provide an integrated view of the cellular processes.

#### References

- Abe T, Kanaya S, Kinouchi M, Ichiba Y, Kozuki T, Ikemura T (2003) Informatics for unveiling hidden genome signatures. *Genome Res* 13:693–702
- Aharoni A, Ric de Vos CH, Verhoeven HA, Maliepaard CA, Kruppa G, Bino RJ, Goodenowe DB (2002) Non-targeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry OMICS 6:217–234
- Askenazi M, Driggers EM, Holtzman DA, Norman TC, Iverson S, Zimmer DP, Boers ME, Blomquist PR, Martinez EJ, Monreal AW, Fiebelman TP, Mayorga ME, Maxon ME, Sykes K, Tobin JV, Cordero E, Salama SR, Trueheart J, Royer JC, Madden KT, Royer JC, Madden KT (2003) Integrating transcriptional and metabolite profiles to direct the engineering of lovastatin-producing fungal strains. *Nature Biotechnol* 21:150–156
- Bajic VB, Veronika M, Veladandi PS, Meka A, Heng M-K, Rajaraman K, Pan H, Swarup S (2005) DPBE: A Text-mining Tool for Integrating Associations between Genetic and Biochemical Entities with Genome Annotation and Biochemical Terms Lists. *Plant Physiol* 138 [In Press]
- Baud-Camus F, Soursac MP, Davrinche M, Farinotti R (2001) Determination of N-acetylation phenotype using caffeine as a metabolic probe and high-performance liquid chromatography with either ultraviolet detection or electrospray mass spectrometry. *J Chrom B Biomed Sci Appl* 760:55–63
- Bohmert K, Balbo I, Kopka J, Mittendorf V, Nawrath C, Poirier Y, Tischendorf G, Trethewey RN, Willmitzer L (2000) Transgenic *Arabidopsis* plants can accumulate polyhydroxybutyrate to up to 4% of their fresh weight. *Planta* 211:841–845
- Brown SC, Kruppa G, Dasseux JL (2005) Metabolomics applications of FT-ICR mass spectrometry. *Mass Spectrom Rev* 24:223–231
- Cook D, Fowler S, Fiehn O, Thomashow MF (2004) A prominent role for the CBF cold response pathway in configuring the low-temperature metabolome of *Arabidopsis*. *Proc Natl Acad Sci USA* 101:15243–15248
- Cellini F, Chesson A, Colquhoun I, Constable A, Davies HV, Engel KH, Gatehouse AM, Karenlampi S, Kok EJ, Leyvay JJ, Lehesranta S, Noteburn HP, Pederson J, Smith M (2004) United effects and their detection in genetically modified crops. *Food Chem Toxicol* 42:1089–1125
- Cooper HJ, Marshall AG (2001) Electrospray ionization Fourier transform mass spectrometric analysis of wine. *J Agri Food Chem* 49:5710–5718
- Duran AL, Yang J, Wang L, Sumner LW (2003) Metabolomics spectral formatting, alignment and conversion tools (MSFACTS). *Bioinformatics* 19:2283–2293
- 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
- Fiehn O (2002) Metabolomics: The link between genotypes and phenotypes. *Plant Mol Biol* 48:155–171
- Frick S, Kutchan TM (1999) Molecular cloning and functional expression of *O*-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. *Plant J* 17:329–339
- Fridman E, Pichersky E. (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. *Curr Opin Plant Biol* 8:242–248
- Giddings G, Allison G, Brooks D, Carter A (2000) Transgenic plants as factories for biopharmaceuticals. *Nature Biotechnol* 18:1151–1155
- Haas BJ, Wortman JR, Ronning CM, Hannick LI, Smith RK Jr, Maiti R, Chan AP, Yu C, Farzad M, Wu D, White O, Town CD (2005) Complete reannotation of the *Arabidopsis* genome: Methods, tools, protocols and the final release. *BMC Biol* 3:7
- Hadacek F (2002) Secondary metabolites as plant traits: Current assessments and future perspectives. *CRC Crit Rev Plant Sci* 27:273–322
- Hall R, Beale M, Fiehn O, Hardy N, Sumner L, Bino R (2002) Plant metabolomics: The missing link in functional genomics strategies. *Plant Cell* 14:1437–1440

- 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 and metabolite-to-gene networks in *Arabidopsis* by integration of metabolomics and transcriptomics. *J Biol Chem* 280:25590–25595
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 101:10205–10210
- Ideker T, Thorsson V, Ranish J, Christmas R, Buhler J (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292:929–934
- Jansen JJ, Hoefsloot HC, Boelens HF, van der Greef J, Smilde AK (2004) Analysis of longitudinal metabolomics data. *Bioinformatics* 20(15):2438–2446
- Jenkins H, Hardy N, Beckmann M, Draper J, Smith AR, Taylor J, Fiehn O, Goodacre R, Bino RJ, Hall R, Kopka J, Lane GA, Lange BM, Liu JR, Mendes P, Nikolau BJ, Oliver SG, Paton NW, Rhee S, Roessner-Tunali U, Saito K, Smedsgaard J, Sumner LW, Wang T, Walsh S, Wurtele ES, Kell DB (2004) A proposed framework for the description of plant metabolomics experiments and their results. *Nat Biotechnol* 22:1601–1606
- Johnson HE, Gilbert RJ, Winson MK, Goodacre R, Simht AR, Rowland JJ, Hall MA, Kell DB (2000) Explanatory analysis of the metabolome using genetic programming of simple, interpretable rules. *Genet Prog Evol Machines* 1:243–258
- Kant MR, Ament K, Sabelis MW, Haring MA (2004) Schuurink RC. Differential timing of spider mite-induced direct and indirect defenses in tomato plants. *Plant Physiol* 135:483–495
- Katz JE, Dumlaio DS, Clarke S, Hao J (2004) A new technique (COMSPARI) to facilitate the identification of minor compounds in complex mixtures by GC/MS and LC/MS: tools for the visualization of matched datasets. *J Am Soc Mass Spectrom* 15: 580–584
- Kliebenstein DJ et al. (2001a) Comparative quantitative trait loci mapping of aliphatic, indolic and benzylic glucosinolate production in *Arabidopsis thaliana* leaves and seeds. *Genetics* 159:359–370
- Kliebenstein DJ et al. (2001b) Genetic control of natural variation in *Arabidopsis* glucosinolate accumulation. *Plant Physiol* 126: 811–825
- Kohonen T, Kaski S, Lappalainen H, Salojärvi J (1997) The adaptive-subspace self-organizing map (ASSOM). In: Proceedings of WSOM'97, Workshop on Self-Organizing Maps, Espoo, Finland, June 4–6, pp. 191–196. Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland
- Krishnan P, Kruger NJ, Ratcliffe RG (2005) Metabolite fingerprinting and profiling in plants using NMR. *J Exp Bot* 56:255–265
- Le Gall G, Colquhoun IJ, Davis AL, Collins GJ, Verhoeven ME (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using <sup>1</sup>H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J Agri Food Chem* 51:2447–2456
- Lim EK, Li Y, Parr A, Jackson R, Ashford DA, Bowles DJ (2000) Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in *Arabidopsis*. *J Biol Chem* 276:4344–4349
- Lindon JC, Nicholson JK (1997) Recent advances in high-resolution NMR spectroscopic methods in bioanalytical chemistry. *TRAC-Trends Anal Chem* 16:190–200
- Markey MK, Lo JY, Tourassi GD, Floyd CE (2003) Self-organizing map for cluster analysis of a breast cancer database. *Artif Intell Med* 27:113–127
- 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
- Nikiforova VJ, Kopka J, Tolstikov V, Fiehn O, Hopkins L, Hawkesford MJ, Hesse H, Hoefgen R (2005) Systems Rebalancing of Metabolism in Response to Sulfur Deprivation, as Revealed by Metabolome Analysis of *Arabidopsis* Plants. *Plant Physiol* (In press)
- Noteburn HP, Lommen A, Jagt RCV, Weseman JP (2000) Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J Biotechnol* 77:103–114
- Ott KH, Aranibar N, Singh B, Stockton GW (2003) Metabonomics classifies pathways affected by bioactive compounds. Artificial neural network classification of NMR spectra of plant extracts. *Phytochemistry* 62(6):971–985
- Raamsdonk LM, Teusink B, Broadhurst D, Zang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, Westerhoff HV, van Dam K, Oliver SG (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* 19:45–50
- Summers LW, Mendes P, Dixon RA. (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* 62:817–836
- Tesniere C, Verries C (2001) Molecular cloning and expression of cDNAs encoding alcohol dehydrogenases from *Vitis vinifera* L. during berry development. *Plant Sci* 157:77–88
- Tolstikov VV, Lommen A, Nakaniski K, Tanaka N, Fiehn O (2003) Monolithical silica-based capillary reversed-phase liquid chromatography/electrospray mass spectrometry for plant metabolomics. *Anal Chem* 75:6737–6740
- Verdonk JC, Ric de Vos CH, Verhoeven HA, Haring MA, van Tunen AJ, Schuurink RC (2003) Regulation of floral scent production in petunia revealed by targeted metabolomics. *Phytochemistry* 62:997–1008
- Wang Y, Tang H, Nicholson JK, Hylands PJ, Sampson J, Whitcombe I, Stewart CG, Caiger S, Oru I, Holmes E (2004) Metabolomic strategy for the classification and quality control of phytochemistry: A case study of chamomile flower (*Matricaria recutita* L.). *Planta Med* 70(3):250–255
- Weckwerth W, Wenzel K, Fiehn O (2004a) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. *Proteomics* 4:78–83
- Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004b) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci USA* 101:7809–7814
- World Health Organization. (2000) Safety aspects of genetically modified foods of plant origins. In: Report of a joint FAO/WHO expert consultation on foods derived from biotechnology, held in Geneva, Switzerland, 29 May–2 June 2000. World Health Organization, Geneva, pp 1–35