Review article

Plant Omics Journal
1(1):26-36 (2008)
ISSN 1836-3644

Southern Cross Journals© 2008
www.pomics.com

Linking transcript profiles to metabolites and metabolic pathways: A systems biology approach to transgene risk assessment

***Kiambi DK[1], Fortin M[2], Stromvick M[2]**

[1*] *Corresponding author- International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), c/o ILRI, P.O.Box 30709-00100 Nairobi, Kenya*
[2] *Department of Plant Science, McGill University, 21,111 Lakeshore, Ste-Anne-de-Bellevue, Quebec Canada H9X 3V9*

Abstract

In recent years, questions related to molecular composition and its implications for nutrition and health have been raised as advances in technology speed up the introduction of new diversity into breeding programs, either via transgenic technology or by using molecular markers in combination with wide crosses. Metabolite profiling offers great opportunities for characterization of this diversity phenotypically with respect to its metabolite composition. It provides a powerful resource to guide breeding programs and to alert researchers to positive or detrimental traits at an early stage. The power of this approach will be vastly increased by combining it with transcript profiling and a systematic survey of the metabolite composition of the plant products that are already on the market. This integrated approach and holistic profiling within a systems biology approach enables the careful tracking of the response of the organism to conditional perturbations at different molecular and genetic levels using available databases. This approach to profiling will not only provide a baseline for comparison of plants with novel traits (PNTs) with traditional comparators that are 'generally recognized as safe', but also provide a rational framework for risk assessment via 'substantial equivalence'. It also provides important inputs into nutritional research and contributes to the public debate about the acceptability of changes in food-production chains and development of science based regulation of plants with novel traits.

Key words: biological systems; functional genomics; integrated approach; pleiotropic effects; profiling

Introduction

Plant biotechnology and genetic modifications offer significant potential in increasing crop production and diversification of the nutritional base. However, one of the major concerns is the possibility of unintended effects caused by transgene integration. Upon random insertion of specific DNA sequences into the plant genome (intended effect), the disruption, modification or silencing of active genes or the activation of silent genes may occur. This could result in the formation of new metabolites, altered levels of existing metabolites, modified metabolism, novel fusion proteins, or other pleiotropic effects that could compromise safety, such as production of new allergens or toxins (Kuiper et al., 2001; Cellini et al., 2004). Unintended effects may be partly predictable on the basis of knowledge of the place of the transgenic DNA insertion, the function of the inserted trait, or its involvement in metabolic pathways; while other effects are unpredictable due to the limited knowledge of gene regulation and gene−gene interactions. Pleiotropic effects such as multiple metabolic changes in tocopherol, chlorophyll, fatty acids and phytoene have been reported by (Shewmaker et al., 1999) while engineering Canola for over-expression of phytoene-synthase. Similarly, in the process of manipulating potato to express yeast invertase, Engel et al.(1998) reported up to 48% reduction in glycoalkaloid levels while Momma et al.

(1999) reported a 50% increase in vitamin B6-content in their work on expression of soybean glycin in rice. Pleiotropic effects have also been demonstrated through gain-function analysis. As reported by Fernie et al. (2004), the analysis of a gene of known function that was introduced into *A. thaliana* confirmed the expected function but also revealed new effects on the metabolic network. This included the up-regulation of the methionine pathway with up to 2–4-fold increases and the down-regulation of the isoleucine pathway, with isoleucine decreasing to 15% compared to levels in the wild-type. However, it should be emphasized that the occurrence of unintended effects is not specific to genetically modified organisms as it also occurs frequently in conventional breeding as reported in Thomas et al. (1998); Coulston and Kolbye (1990) and Beir (1990). In this paper, we briefly review molecular approaches to transgene safety assessment and also provide an update of the on-going work in linking transcript profiling, metabolite profiling and metabolic pathways as a systems biology approach to studying risks associated with transgenes.

Assessing genetic changes

The comparison of the chemical composition of the genetically modified plant to that of a traditionally obtained counterpart has been a key element in the safety assessment of genetically modified crops. Such a comparative approach reveals similarities as well as differences between the transgenic crop and the selected comparator and will thus provide information on the status of 'substantial equivalence' (König et al., 2004; Ye et al., 2000). Through different platforms, it is possible to compare two types of samples, a control sample and a treated or genetically modified sample, to identify individual components showing differential behavior and to therefore account for the responses of the system to the applied perturbation. This comparative analysis generally relies on the statistically significant detection of genetic differences between sample groups as a result of gene function at the level of protein activity and the consequences of introducing a new protein into the metabolic network.

Different approaches and strategies have been applied in the identification of potential secondary effects of the genetic modification. Traditionally, unintended effects have been identified through the targeted approach where an analysis of the agronomical and morphological characteristics of the new plant is followed by an extensive proximate or chemical analysis of key nutrients, anti-nutrients and toxicants typical for the plant. Limitations of this analytical, comparative approach are the possible occurrence of unknown natural toxins and anti-nutrients, particularly in food plant species with no history of safe use, and the availability of adequate detection methods. In addition, there are no generally accepted and harmonized guidelines that define the full extent of the analyses required to fulfill statuary risk assessment procedures. Furthermore, the targeted approach is considered to be biased and focuses more on known compounds and expected or predictable changes (Millstone et al., 1999).

To avoid biases, non-targeted methods are now being used as an alternative approach for the detection of unintended effects, using profiling techniques. This type of comprehensive screening for potential changes in the characteristics of the genetically modified organisms becomes even more important since the next generation of GM crops is likely to include varieties with improved nutritional properties. In the development of this new generation of transgenic crops, there is a possibility of more far reaching effects on metabolic processes due to complexities associated with insertion of large DNA fragments or clusters of genes, increased metabolic perturbations and generation of new biosynthetic pathways. This could lead to the occurrence of unpredictable unintended effects not revealed by a targeted approach and new methods are therefore being used. These methods are genomic based and they include transcript profiling, proteomics and metabolomics. They allow for the screening of potential physiological, cyclical, developmental or environmental changes of the modified host organism at different cellular integration levels: at the genome level during gene expression and protein translation and at the level of metabolic pathways (Kuiper et al., 2003; Kuiper et al., 2001; Cellini et al., 2004). When these genomics-based methods are integrated with bioinformatics technologies, it becomes possible to investigate the global unintended effects through the analysis of transcripts, proteins, and metabolite profiles using a systems biology approach. This global analytical approach allows studies of transgene effects, comparative analysis of genetically modified organisms and investigation of biological entities as integrated systems of genetic, protein, metabolic,

cellular, and pathway events that are in constant flux and interdependent (Jonsson et al., 2005).

Metabolomics as a profiling approach for transgene risk assessment

Metabolomics as a comprehensive analysis in which all the metabolites of an organism are identified and quantified (Trethewey et al., 1999; Fiehn et al., 2000; Sumner et al., 2003; Bino et al., 2004) has emerged as a functional genomics methodology that contributes to our understanding of the complex molecular interactions in biological systems (Roessner et al., 2001; Hall et al., 2002; Jenkins et al., 2004). It therefore represents the logical progression from large-scale analysis of RNA and proteins at the systems level.

In recent years, several reviews have been published and described the use of metabolomics in functional genomics research, including comparative analysis between genetically modified crops and their traditional comparators (Fiehn et al., 2000; Roessner et al., 2001; Catchpole et al., 2005; Lehesranta et al., 2005; Chen et al., 2003). In the context of functional genomics, metabolomics is now regarded as a viable counterpart to protein and transcript profiling technologies (Hall et al., 2002; Streeter and Strembu, 1998; Trethewey, 2001; Trethewey et al., 1999). Indeed, the integration of methods based on gas chromatography/mass spectrometry (GC/MS), liquid chromatography/ mass spectrometry (LC/MS), Fourier Transform Mass Spectrometry (FTMS) and NMR for the comprehensive identification and particularly, the accurate quantification of metabolites has attained a technical robustness that is comparable to or even better than conventional mRNA or protein profiling technologies (Aharoni et al., 2002; Fiehn et al., 2000; Roessner et al., 2001; Weckwerth et al., 2001; Kopka et al., 2004).

The accurate identification and the relative quantification of a high number of metabolites in a multitude of samples makes it possible to study dynamic metabolomics networks (Fiehn, 2003; Weckwerth et al., 2004a; 2004b) and also undertake comparative studies between genetically modified crops and their traditional comparators that are 'generally recognized as safe (Gras)' based on the extent of their natural variation (Roessner et al., 2001; Catchpole et al., 2005; Lehesranta et al., 2005). The data generated is fundamentally different from traditional biological measurements and thus the analysis is often restricted to rather pragmatic approaches, such as data mining tools to discriminate between different metabolic phenotypes. These analytical approaches include tools for data acquisition, transformation, validation, aligning, deconvolution and machine learning such as Metalign (www.metalign.nl), MSFACTS (**www.noble.org/ PlantBio/ MS/ MSFACTs/ MSFACTs**), AMDIS (www.amdis.net) and MASSLAB (Duran et al., 2003; Bino et al., 2004; Taylor et al., 2002). To enable the analysis of transgene effects at the metabolic level, the data obtained from metabolomic experiments can also be organized into metabolic correlation networks based on their pair-wise correlations but the key challenge is to deduce unknown pathways on the basis of observed correlations (Fiehn et al., 2003; Fiehn and Weckwerth, 2003; Fiehn, 2003; Steur et al., 2003a; 2003b; Fiehn, 2002; Weckwerth and Fiehn, 2002; Weckwerth et al., 2004). In addition, metabolome mass-spectral reference databases such as GMD (Kopka et al., 2005), other databases for visualizing biochemical and metabolic pathways and user-driven tools for displaying data onto diagrams of metabolic pathways and processes have also been similarly developed recently. These include AraCyc (http://www.Arabidopsis.org/ tools/aracyc/); ArMet (http://www.armet.org/); MetaCyc http://metacyc.org/ and MetNet (Kose et al., 2001; Wurtele et al., 2003; Bino et al., 2004; Yang et al., 2005).

Though metabolomics is developing as a reliable tool for transgene risk assessment, many critical parameters that complicate the interpretation of metabolite profiles are yet to be resolved. These include the discrepancy between the low number of detected metabolites versus the real number of possible metabolites in plants that is estimated at 200,000 (Pichersky and Gang, 2000), variations caused by the extraction process, the bias against compound classes and most importantly, the overlap of many compartmentalized metabolic processes in tissue samples. For it to become a more robust profiling technique, the integration of metabolomic data with other functional genomics information using a systems biology approach needs enhancing through the establishment of relational databases that store, compare, integrate and enable the determination of causal relationships between genes, transcripts, proteins and metabolites.

Transcript profiling in detecting differential gene expression for transgene risk assessment

Functional genomics refers to the study of direct expression products of genes, the mRNA transcripts and the related regulatory elements. It can therefore provide insight into the complex metabolic relationships within an organism including pathways that are relevant for the safety of food crops. It may also lead to an in-depth understanding of the natural variation in the expression of genes under different environmental conditions (Sommerville and Sommerville, 1999). The scale and resolution of DNA micro-arrays that are used to generate transcripts facilitates the detection of alterations in gene expression and the possible consequences for food safety, if the relevant pathways are known. The technology relies on using a large variety of individual identified probes that could be cDNA sequences or oligonucleotides in a single experiment by arraying the probes to a solid surface. The probes can either be synthesized on a solid support (oligonucleotides, especially gene chips from Affymetrix) or synthesized prior to spotting in array format (Lockhart and Winzeler, 2000; van Hal, 2000). All the probes are subsequently hybridized simultaneously to the labeled sample under investigation. This allows gene expression profiles to be established from individual or mixed tissue samples such as transgenic plant varieties and compared to the unmodified controls (wild types). Any differences in gene expression profiles that may be detected could be an indication of unintended effects of the genetic modification and may provide information for further investigations and implications for risk assessment (Kuiper et al., 2001; 2003; Cellini et al., 2004).

Micro-arrays have recently been used as an alternative to traditional analysis of differential gene expression (Alwine et al., 1977; Welsh et al., 1992; Liang and Pardee, 1992) due to the advantage of the parallel screening of a large number of identified gene sequences for differences in gene expression in different types of tissues (Van Hal et al., 2000). In relation to food safety or risk assessment of plant products, gene expression studies could focus on metabolic routes leading to the formation of anti-nutrients, including natural toxins, as well as on the metabolism of positive nutritional factors (micro- and macronutrients) to monitor for possible unintended changes. Moreover, other cDNAs can be spotted on the array to screen for alterations in gene expression in other metabolic systems of the plant that may be of relevance to the safety or nutritional value of the plant. Traditional methods of differential display have been applied in the detection of altered gene expression in genetically modified plant material (Kok et al., 2001; Kok et al.,1998; Liang and Pardee, 1992) but currently there are no published examples available on the application of genomics (DNA micro array technology) to the detection of unintended effects in GM products. However, within the EU Fifth Framework project GMOCARE (GMOCARE, 2003), the potential for analyzing differential gene expression using DNA micro-arrays as a means of contributing to future improved food safety evaluation strategies is currently being assessed. In addition our studies reveal that……..

Transript profiling has benefited from some of the most advanced genomics database for plants including The *Arabidopsis* Information Resource (TAIR) (Rhee et al.; 2003), GARNET. http://www.york.ac.uk/res/garnet/garnet.htm, KEGG, http://www.genome.ad.jp/kegg /pathway. html and TIGER. Recent advances have seen the emergence of databases that combine sequence information with information on genetics, gene expression, homology, regulation, function, interactions, biochemical pathways and phenotype information (Lockhart and Winzeler, 2000; Gerstein, 2000; Bassett et al., 1999; Steinhauser et al., 2004b). Such databases are now enabling the better understanding of observed differences in gene expression and related phenotypic alterations, and hence the subsequent consequences for food safety.

Integrating transcript profiling and metabolite profiling through a systems biology approach

Biological relevance

Both theoretical and experimental disciplines have seen the emergence of systems-based approaches to biology in the past few years as typified shifts from the more traditional reductionist approach towards more holistic approaches, with experimental strategies aimed at understanding interactions, such as links between transcripts and metabolites, across multiple molecular entities (Oksman-Caldentey et al., 2004). This holistic understanding of the biological behavior of a complex system enables the careful tracking of the response of an organism to conditional

perturbations at different molecular and genetic levels. Through holistic profiling within a systems biology approach it is possible to identify markers and mechanisms that are important to the function of the perturbed system, with the ultimate goal of developing computational models that enable the prediction of the response of the system to any given perturbation (Kitano, 2002; Sweetlove et al., 2003). This "systems" approach to biology involves the comprehensive characterization of the components of a biological system at the transcriptome, proteome and metabolome levels (Weckwerth, 2003; Fiehn et al., 2001). The three levels of expression profiling provide a complete picture of the RNAs, proteins and metabolites that allows the inference of relevant associations between macromolecules; identification of functional linkages between phenotypic expressions and construction of models that quantitatively describe the dynamics of the biological system. In addition, the linkage of functional metabolomic information to mRNA and protein expression data makes it possible to visualize the functional genomic repertoire of an organism (Bino et al., 2004), bringing us into a new era of gene discovery, understanding biological systems and how genes are connected to metabolites (Oksman-Caldentey et al., 2004; Trethewey, 2001). This knowledge has great potential for application, particularly in the development and engineering of crops that combine an attractive appearance and taste with improved levels of phytonutrients such as flavonoids and carotenoids (Jonsson et al., 2005).

Integration of metabolites and transcript profiles

The multi-parallel approach combining metabolites and transcripts profiling methods are known to provide an immediate insight into the behaviour of the whole metabolic network after modulation of a particular gene function (Fiehn et al., 2001; Oksman-Caldentey and Saito, 2005). It provides exciting opportunities for defining gene function at the level of metabolic networks and the overall phenotype in the context of a particular organism. Integration of metabolite and transcript profiling has shown that a statistically significant change in the steady-state level of any given metabolite will be triggered by an over-expression of 0.1–1.0% of the genes in a genome. In some cases, the genes will influence flux directly in a pathway and in other cases, they might trigger a host of regulatory changes that alter the atomic partitioning or the activity of metabolic networks (Fernie et al., 2005).

Due to the well-known connectivity between the molecules described by transcriptomic and metabolomic approaches, several studies have tried to correlate transcript and metabolite profiles to decipher metabolic networks, identify candidate genes and elucidate gene functions (Urbanczyk-Wochniak et al., 2005; Urbanczyk-Wochniak et al., 2004; Sharit et al., 2003; Lavid et al., 2002; Hirai et al., 2004; Goosens et al., 2003). For example, in the studies by Urbanczyk-Wochniak et al. (2004), metabolite–transcript correlations were revealed from large data sets collected throughout development in wild-type and transgenic tubers engineered to have enhanced sucrose metabolism. The transcript levels of approximately 280 transcripts that showed reproducible changes with respect to control samples were systematically plotted against changes in metabolite levels of paired samples. A total of 571 out of the 26,616 possible pairs showed significant correlation (at the $P < 0.01$ level). Most of the significant correlations were new and included the identification of several strong correlations between genes and nutritionally important metabolites. This approach has a high potential value in the identification of candidate genes for metabolic engineering and modifying the metabolite content of biological systems (Fernie et al., 2004; Trethewey, 2001). Another study by the same group (Urbanczyk-Wochniak et al., 2005) describes the parallel profiling of diurnal patterns of metabolite and transcript abundance in potato. The study revealed 56 significant differences observed in metabolic contents and 832 significant differences in transcript levels. The qualitative comparison of the combined data obtained from the parallel analysis of transcripts and metabolites suggested that relatively few changes in gene expression strongly correlate with changes in metabolite levels during a diurnal cycle.

Significant progress has also been made in exploring cellular processes by combining genome-wide transcriptomics and metabolomics as reported by Hirai et al. (2004). In this study, DNA array transcriptome analysis was combined with metabolite profiling and more specific targeted quantitative analysis resulting in a huge amount of data. Novel bioinformatics tools were developed to integrate the data sets and to generate gene-to-metabolite networks. In a similar study Goossens et al. (2003) combined cDNA-amplified fragment length polymorphism

(AFLP) transcript profiling with targeted metabolite analysis to map the biosynthetic genes involved in alkaloid metabolism. Since sequence information for many medicinal plants is very limited, the cDNA-AFLP transcript profiling provided a very powerful tool to identify many candidate genes involved in the production of secondary metabolites. Functional analysis of these candidate genes will generate a lot of data and might help to find not only the biosynthetic genes of a particular plant pathway but also master regulators such as transcription factors that are involved in plant secondary metabolism in general.

Development of integrated databases

Notable progress has been made in the establishment of integrated and comprehensive systems biology databases that will allow the development of biological systems networks by integrating transcriptome, metabolome and flux data. These include MapMan (**http://gabi.rzpd.de/ projects/ MapMan/**), which is a user-driven tool that displays large datasets onto diagrams of metabolic pathways or other processes. It is composed of multiple modules for hierarchical grouping of transcript and metabolite data that can be visualized using a separate user-guided module (Usadel et al., 2005; Thimm et al., 2004). Another more comprehensive tool is MetNet http:// www. public.iastate.edu/ ~mash/ MetNet/ homepage. html, which contains a suite of open-source software tools for systems biology and is designed to provide a framework for the formulation of testable hypotheses regarding the function of specific genes (Bino et al., 2004). Other systems biology oriented software have recently been developed and they include (CSB.DB) (http://csbdb.mpimp-golm.mpg.de/), an open access comprehensive systems-biology database that presents the results of bio-statistical analyses on gene expression data in association with additional biochemical and physiological knowledge. The database platform provides tools that support insight into life's complexity pyramid with a special focus on the integration of data from transcript and metabolite profiling experiments (Stenhauser et al., 2004). In addition, Ludeman et al. (2004) have developed PaVESy as a relational sequel data mining and managing system for editing and visualization of biological pathways. The database design allows storage of biological objects, such as metabolites, proteins, genes and respective relations, which are required to assemble metabolic and regulatory biological interactions. The database model accommodates highly flexible annotation of biological objects by user-defined attributes.

Some progress has been made in using cDNA libraries or EST databases in combination with metabolic and gene expression profiles to make biological inferences, identify genes and elucidate gene functions when complete genome sequences are not available (Mercke et al., 2004; Guterman et al., 2002; Martin et al., 2004; Ritchman et al., 2005). For instance, metabolic profiling performed by Martin et al. (2004) on young Norway spruce trees treated with methyljasmonate (MeJA) showed the emission of a large number of monoterpenes and sesquiterpenes, as well as the synthesis of non-volatile diterpenes. The group proceeded to screen a cDNA library of young spruce shoots and leaves by a combination of homology-based PCR and DNA-hybridization techniques, thereby isolating nine TPS cDNAs. Each cDNA was expressed in *E. coli* and tested with the appropriate substrate. The results of these assays indicated that four of the cDNAs encoded monoterpene synthases, three cDNAs encoded sesquiterpene synthases, and two cDNAs encoded diterpene synthases (Fridman and Pichersky, 2005).

Implications for PNT risk assessment and regulation

Despite the progress made one of the outstanding questions remains “What is the relative power of the two phenotyping technologies to discriminate biological systems that either differ in developmental state or show well-characterized changes in response to the expression of transgenes?”. These two main functional genomics approaches dealt with in this study are ‘information-rich’, and each method is vulnerable to various statistical caveats because the data generally originate from a few samples, yet each sample is characterized by several thousand features including genes, *m/z* values (mass-to-charge ratios of metabolites or metabolite fragments) and spectral intensities that might lead to difficulties in the interpretation and validation of resultant data (Bino et al., 2004). For these technologies to be even more effective tools for PNT risk assessment and regulation, further progress will need to be made in the validation of the vast information generation. This will allow the studying of the biological entity dynamics and analysis of fluxes in metabolic

pathways in order to decipher the biological relevance of transcripts to metabolites. More comprehensive bio-informatic tools need to be developed to extract relevant biological information from raw data sets using a systems biology approach that integrates data sets from both metabolomics and genomics platforms. The comprehensiveness of coverage given by these two major profiling techniques will also need to be improved. As this comprehensiveness increases and bioinformatic tools mature, functional metabolomic information can be linked to transcriptome datasets to allow a better understanding of organisms within a systems biology realm. This will make it easier to visualize and assess the effects of transgenes and perturbations resulting from their integration in biological entities. The combination of the new techniques of metabolic and gene expression profiling will also allow the identification of the function of the majority of the genes in plant genomes and also make tangible contributions towards comparisons of plants with novel traits with the traditional comparators that are generally recognized as safe. However, for this to be accomplished, the development of publicly available databases of crop composition and profiles is an absolute requirement in order to determine natural variation of compounds within and between given plant species. As information is gathered, evolving baselines and benchmarks with which to compare plants with novel traits could be envisaged. These databases would also greatly aid the robustness of targeted analyses.

Conclusions

There is little doubt that the existing profiling techniques when used in an integrated manner and using a systems biology approach provide sufficient basis for science based regulation of PNTs. They have proven successful in revealing unintended effects but it may be argued, however, that unintended effects do not automatically or necessarily infer health hazards. Ideally, only those parameters that fall outside the range of natural variation should be considered further in safety assessment. The main impediment however, is the lack of information on the natural variation within and between plant cultivars for all the parameters that can be measured. Safety assessments could be simplified if the identification and safety significance of any observed differences is known. The regulators need to develop guidelines on how different should a particular parameter be from its “Gras” comparator for it to be considered a risk. However, one major drawback is the lack of adequate toxicity databases to aid the interpretation of the safety significance of compounds with unknown identity and/or function. Major differences based on quantities and or novelty of un-intended effects may lead to the consideration of more extensive safety testing but this becomes a regulatory issue.

References

Aharoni A., de Vos CHR, Verhoeven HA, Maliepaard CA., Kruppa G, Bino R, and Goodenowe DB (2002) Non-targeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. OMICS 6: 217–234.

Alwine JC, Kemp DJ, Stark GR (1977) Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proceedings of the National Academy of Sciences U.S.A. 74: 5350–5354.

Bassett DE., Eisen MB, Boguski MS (1999) Gene expression informatics—it's all in your mine. Nature Genetics 21 Suppl.: 51–55.

Beier RC (1990) Natural pesticides and bioactive components in food. Rev. Environ. Contamin. Tox. 113: 47–137.

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, Trethewey RN, Lange BM, Wurtele ES, Sumner LW (2004) Potential of metabolomics as a functional genomics tool. Trends in Plant Science. Vol. 9 No.9: 418-425.

Catchpole GS, Beckmann M, Enot DP, Mondhe M, Zywicki B, Taylor J, Hardy N, Smith A, King RD, Kell DB, Fiehn O, Draper J (2005) Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proc Natl Acad Sci U S A.102(40):14458-62.

Cellini FA, Chesson I, Colquhoun A, Constable HV, Davies KH, Engel AMR, Gatehouse S, Kärenlampi EJ, Kok J, Leguay S, Lehesranta HP, Noteborn J, Pedersen J, Smith M (2004) Unintended effects and their detection in genetically modified crops. Food and Chemical Toxicology 42:1089-1125.

Chen F, Duran AL, Blount JW, Sumner LW, Dixon RA (2003) Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis. Phytochemistry. 64(5):1013-21.

Coulston F, Kolbye AC (1990) Biotechnologies and food: assuring the safety of foods produced by genetic modification. Reg Toxicol Pharmacol. 12, S1−S196

Duran AL, Yang J, Wang L, Sumner LW (2003) Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). Bioinformatics.19(17):2283-93.

Engel KH, Gerstner G, Ross A (1998) Investigation of glycoalkaloids in potatoes as example for the principle of substantial equivalence. In: Novel Food Regulation in the EU −Integrity of the Process of Safety Evaluation. Berlin: Federal Institute of Consumer Health Protection and Veterinary Medicine, pp. 197−209.

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nature Reviews Molecular Biology. Vol 5 No.9: 763-769.

Fiehn O (2003) Metabolic networks of *Cucurbita maxima* phloem.Phytochemistry. 62(6):875-86.

Fiehn O (2002) Metabolomics--the link between genotypes and phenotypes. Plant Mol Biol. 48(1-2):155-71

Fiehn O,Weckwerth, W (2003). Deciphering metabolic networks. Eur. J. Biochem 270: 579–588

Fiehn O, Kloska S, Altmann T (2001) Integrated studies on plant biology using multiparallel techniques. Curr Opin Biotechnol.12(1): 82-6

Fiehn O, Kopka J, Dormann P, Altmann T, Trethewey RN, Willmitzer L. 2000. Metabolite profiling for plant functional genomics. Nat Biotechnol.18(11):1157-61.

Fridman E, Pichersky E (2005) Metabolomics, genomics, proteomics, and the identification of enzymes and their substrates and products. Curr Opin Plant Biol. 8(3):242-8

Gerstein M (2000) Integrative database analysis in structural genomics. Nature Structural Biology 7 Suppl.: 960–963.

GMOCARE (2003) GMOCARE, New Methodologies for Assessing the Potential of Unintended Effects in Genetically Modified Food Crops. GMOCARE Secretariat. , RIKILT Institute of Food Safety, Wageningen http:// www. entransfood.com/RTDprojects/GMOCARE/default. htm .

Goossens A., Häkkinen ST, Laakso I, Seppänen-LaaksoT, Biondi S, De Sutter V, Lammertyn F, Nuutila AM, Söderlund H, Zabeau M (2003) A functional genomics approach towards the understanding of secondary metabolism in plant cells. Proc Natl Acad Sc USA 100, 8595-8600.

Guterman, I, M. Shalit, N. Menda, D. Piestun, M. Dafny-Yelin, G. Shalev, E. Bar, O. Davydov, M. Ovadis and M. Emanuel *et al.*,. 2002. Rose scent: genomics approach to discovering novel floral fragrance-related genes. *Plant Cell* 14: 2325–2338.

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(7):1437-40.

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.

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(12):1601-6.

Jonsson P, Johansson AI, Gullberg J, Trygg J, Grung B, Marklund S, Sjostrom M, Antti H, Moritz T (2005) High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. Anal Chem 77(17): 5635-42.

Kitano H (2002) Systems biology: a brief overview. Science 295: 1662-1664.

Kok EJ, van der Wal-Winnubst ENW, Van Hoef AMA, Keijer J (2001) Differential display of mRNA. In: Akkermans, ADL, van Elsas JD, de Bruijn FJ, (eds) Molecular Microbial Ecology Manual, Supplement 5, 8.1.1.

Kok EJ, Keijer J, Van Hoef AMA, Van der Wal-Winnubst ENW, Henkens MHC, Kuiper HA, (1998) mRNA fingerprinting of transgenic food crops. In: Report of the Demonstration Programme on Food Safety Evaluation of Genetically Modified Foods as a Basis for Market Introduction. Ministry of Economic Affairs, The Hague, 37-49.

König A, Cockburn A, Crevel RWR, Debruyne E, Grafstroem R, Hammerling U, Kimber I, Knudsen I, Kuiper HA, Peijnenburg AACM, Penninks AH, Poulsen M, Schauzu M, Wal JM, (2004)

Assessment of the safety of food derived from genetically modified (GM) crops. Food and Chemical Toxicology 42(7):1047-88

Kopka J, Fernie A, Weckwerth W, Gibon Y, Stitt M (2004) Metabolite profiling in plant biology: platforms and destinations. Genome Biol 5(6):109.

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, Dormann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D. 2005. GMD@CSB.DB: the Golm Metabolome Database. Bioinformatics 21(8):1635-8.

Kose F, Weckwerth W, Linke T, Fiehn O (2001) Visualizing plant metabolomic correlation networks using clique-metabolite matrices.Bioinformatics.17(12):1198-208

Kuiper HA (2001) Assessment of the food safety issues related to genetically modified foods. Plant J 27(6):503-28.

Kuiper HA, Kok EJ, Engel KH (2003) Exploitation of molecular profiling techniques for GM food safety assessment. Curr Opin Biotechnol 14(2):238-43

Lavid NJ, Wang M, Shalit I, Guterman E, Bar T, Beuerle N, Menda S, Shafir D, Adam Z (2002) O-methyltransferases involved in the biosynthesis of volatile phenolic derivatives in rose petals. Plant Physiol 129: 1899–1907.

Liang P, Pardee AB (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 255: 967–971.

Ludemann A, Weicht D, Selbig J, Kopka J (2004) PaVESy: Pathway Visualization and Editing System.

Lehesranta SJ, Davies HV, Shepherd LV, Nunan N, McNicol JW, Auriola S, Koistinen KM, Suomalainen S, Kokko HI, Karenlampi SO (2005) Comparison of tuber proteomes of potato varieties, landraces, and genetically modified lines. Plant Physiol 138(3):1690-9.

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature 405: 827–836.

Millstone E, Brunner E, Mayer S (1999) Beyond substantial equivalence. Nature 401: 525–526.

Martin DM, Faldt J, Bohlmann J (2004) Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. Plant Physiol 135:1908–1927.

Mercke P, Kappers IF,Verstappen FW, Vorst O, Dicke M, Bouwmeester HJ (2004) Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants. Plant Physiol 135:2012–2024.

Momma K, Hashimoto W, Yoon HJ, Ozawa S, Fukuda Y, Kawai S, Takaiwa F, Utsumi S, Murata K (2000). Safety assessment of rice genetically modified with soybean glycinin by feeding studies on rats. Biosci Biotechnol Biochem 64:1881−1886.

Oksman-Caldentey KM, Inzé D, Orešič M (2004) Connecting genes to metabolites by a systems biology approach. Proc Natl Acad Sci USA 101: 9949–9950.

Oksman-Caldentey KM, Saito K (2005) Integrating genomics and metabolomics for engineering plant metabolic pathways. Curr Opin Biotechnol 16(2):174-9

Pichersky E, Gang DR (2005) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective. Trends Plant Sci 5(10):439-45.

Rhee SY. *et al.* (2003) The *Arabidopsis* Information Resource (TAIR): a model organism database providing a centralized, curated gateway to *Arabidopsis* biology, research materials and community. Nucleic Acids Res 31:224–228.

Ritchman AS, Swanson A, Humphrey T, Chapman R, McGravey B, Pocs R, Brandle JE (2005) Functional genomics uncovers three glucosyltransferases involved in the synthesis of the major sweet glucosides of *Stevia rebaudiana*. Plant J 41: 56–67.

Roessner U, Luedemann A, Brust D, Fiehn O, Linke T, Willmitzer L, Fernie A (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. Plant Cell 13(1):11-29.

Roessner U, Wilmitzer L, Fernie AR (2001) Metabolic profiling and biochemical phenotyping of plant systems. Plant Cell Rep 21:189-196.

Scholz M, Gatzek S, Sterling A, Fiehn O, Selbig J (2004) Metabolite fingerprinting: detecting biological features by independent component analysis. Bioinformatics 15: 2447-54.

Shalit M, Guterman I, Volpin H, Bar E, Tamari T, Menda N, Adam Z, Zamir D, Vainstein A, Weiss D (2003) Volatile ester formation in roses. Identification of an acetyl-coenzyme A:geraniol/ citronellol acetyltransferase in developing rose petals. Plant Physiol 131: 1868–1876.

Shewmaker CK, Sheely JA, Daley M, Colburn S, Ke DY (1999) Seed-specific overexpression of

phytoene synthase: increase in carotenoids and other metabolic effects. Plant J 22: 401 –412.

Sommerville C, Sommerville S (1990) Plant functional genomics. Science 285: 380–383.

Steinhauser D, Usadel B, Luedemann A, Thimm O, Kopka J (2004a) CSB.DB: a comprehensive systems-biology database. Bioinformatics 20(18):3647-51.

Steinhauser D, Junker BH, Luedemann A, Selbig J, Kopka J (2004b) Hypothesis-driven approach to predict transcriptional units from gene expression data. Bioinformatics 20(12):1928-39.

Steuer R, Kurths J, Fiehn O, Weckwerth W (2003a) Interpreting correlations in metabolomic networks. Biochem Soc Trans 31(6):1476-8

Steuer R, Kurths J, Fiehn O, Weckwerth W.2003b. Observing and interpreting correlations in metabolomic networks. Bioinformatics. 19(8):1019-26.

Streeter JG, Strimbu CE (1998) Simultaneous extraction and derivatization of carbohydrates from green plant tissues for analysis by gas-liquid chromatography. Anal Biochem 259(2):253-7.

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry 62(6):817-36

Sweetlove, L *et al.* (2003) Predictive metabolic engineering: a goal for systems biology. Plant Physiol.132: 420–425.

Taylor J, King RD, Altmann T, Fiehn O (2002) Application of metabolomics to plant genotype discrimination using statistics and machine learning. Bioinformatics18 Suppl 2:S241-8.

Thimm O. *et al.*, (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J 37: 914–939.

Thomas WTB, Baird E, Fuller JD, Lawrence P, Young GR, Russell J, Ramsay L, Waugh R, Powell W (1998) Identification of a QTL decreasing yield in barley linked to Mlo powdery mildew resistance. Mol Breed 4: 381−393.

Trethewey RN (2004) Metabolic profiling as an aid to metabolic engineering in plants. Curr Opin Plant Biol 7:196-201.

Trethewey RN (2001) Gene discovery via metabolic profiling. Curr Opin Biotechnol.12(2):135-8.

Trethewey RN, Krotzky AJ, Willmitzer L (1999) Metabolic profiling: a Rosetta Stone for genomics? Curr Opin Plant Biol 2(2):83-5.

Urbanczyk-Wochniak E, Luedemann A, Kopka J, Selbig J, Roessner-Tunali U, Willmitzer L, Fernie AR (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. EMBO Rep 4(10):989-93

Urbanczyk-Wochniak E, Baxter C, Kolbe A, Kopka J, Sweetlove LJ, Fernie AR (2005) Profiling of diurnal patterns of metabolite and transcript abundance in potato (*Solanum tuberosum*) leaves. Planta 221: 891-903.

Usadel B, Nagel A, Thimm O, Redestig H, Blaesing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D, Scheible WR, Gibon Y, Morcuende R, Weicht D, Meyer S, Stitt M (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. Plant Physiol 138(3):1195-204.

van Hal NLW, Vorst O, van Houwelingen AMML, Kok EJ, Peijnenburg A, Aharoni A, van Tunen AJ, Keijer J (2000) The application of DNA microarrays in gene expression analysis. Journal of Biotechnology 78: 271–280.

Weckwerth W (2003) Metabolomics in systems biology. Annu Rev Plant Biol 54:669-89.

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O. 2004a. Differential metabolic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci U S A.101(20):7809-14.

Weckwerth W, Wenzel K, Fiehn O (2004b) Process for the integrated extraction, identification and quantification of metabolites, proteins and RNA to reveal their co-regulation in biochemical networks. Proteomics. 4(1):78-83.

Weckwerth W, Fiehn O (2002) Can we discover novel pathways using metabolomic analysis? Curr Opin Biotechnol 13(2):156-60.

Weckwerth W, Tolstikov V, Fiehn O (2001) Metabolomic characterization of transgenic potato plants using GC/TOF and LC/MS analysis reveals silent metabolic phenotypes. Proc ASMS Conf Mass Spectrom Allied Topics, 49th, pp. 1−2. Chicago: Am. Soc. Mass Spectrom.

Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, McClelland M (1992) Arbitrarily primed PCR fingerprinting of RNA. Nucleic Acids Research 2019: 4970–4976.

Wurtele E. (2003). MetNet: software to build and model the biogenetic lattice of *Arabidopsis*. Comp Funct Genomics 4:239–245.

Yang Y, Engin L, Wurtele ES, Cruz-Neira C, Dickerson JA (2005) Integration of metabolic networks and gene expression in virtual reality. Bioinformatics15: 21(18):3645-50.

Ye X, AlBabili S, Kloeti A, Zhang J, Lucca P, Beyer P, Potrykus I (2000) Engineering the provitamin A (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. Science 287: 303–305.