PROFILING OF GENETICALLY MODIFIED ORGANISMS USING OMICS TECHNOLOGIES

Alberto Valdés, C. Simó, C. Ibáñez and Virginia García-Cañas

Laboratory of Foodomics, CIAL, CSIC, Nicolas Cabrera 9, 28049 Madrid, Spain

*Corresponding e-mail: virginia.garcia@csic.es

Abstract

Strict regulations including risk assessment, labeling, traceability and marketing have been established due to the controversial safety aspects of genetically modified organisms (GMOs). One of the main polemic issues associated with GMOs safety are the possible unintended effects, defined as effects that go beyond the primary expected effects of the genetic modification. In order to effectively investigate the potential adverse effects on the human health, including the existence or not of unintended effects, new analytical tools are needed to facilitate comprehensive compositional studies of GMOs. In this context, profiling technologies have the potential to provide valuable information regarding GMOs composition that can be useful for characterization, traceability and even GMO detection. In this chapter, the application of the main -omics technologies (transcriptomics, proteomics, and metabolomics) in combination with bioinformatics and chemometrics tools to GMO profiling are discussed.

Keywords: GMO, omics, gene-expression profiling, metabolomics, proteomics, transcriptomics, substantial equivalence, unintended effects, transgenic crops.
TABLE OF CONTENTS

1. INTRODUCTION

2. DEBATED SAFETY ISSUES ON GMOs

3. OMICS PROFILING IN GMO ANALYSIS
   3.1 Transcriptomics
   3.2 Proteomics
   3.3 Metabolomics
      3.3.1 NMR
      3.3.2 MS-based technologies
      3.3.3 Multi-platform strategies
   3.4 Cross-omics studies

4. FUTURE OUTLOOK AND CONCLUSIONS
1. INTRODUCTION

For centuries, the production of foods with the desired quality has been a major goal in agriculture. To that aim, classical plant breeding has been applied to improve plant varieties with different techniques, such as plant crossing and selection, cell tissue culture and mutagenesis based on irradiation, among others. On the other side, genetic engineering (or recombinant DNA technology) allows to transfer selected individual gene sequence from one organism into another, where the acceptor can be from the same species or not. Recombinant technology represents one of the most technological advances in the past decades in modern biotechnology, and the organism derived from this technology is termed genetically modified organism (GMO). In this sense, genetically modified (GM) foods are foods derived from organisms whose genetic material (DNA) has been modified in a way that does not occur naturally, e.g. through the introduction of a gene from a different organism [1]. The modifications incorporated through the recombinant technology generally represent some advantages in terms of agronomic productivity and industrial processing of the GM crops over their non-modified counterparts.

The first GM crop was commercialized in 1996, and since then, 170 million hectares have been approved by 2012. At this time, 30 countries have approved GMO crops [2]. Soybean, maize, cotton and rapeseed are the most represented crops in terms of cultivated area, and by itself, soybean accounts for more than 50% of the GM crop production (mostly used for high protein animal feed) [3]. Beside these major crops, other minor GM crops that can be also found in the market include canola, potatoes, eggplant, carrots, etc. Regarding the genetically modified traits in GMOs, the most frequent are herbicide resistance, insect resistance and resistance to viral pathogens [4]. Other important traits are resistance to virus, resistance to severe environmental conditions or enhanced nutritional properties. Some GM products that are in the pipeline of commercialization in a near future include plants enriched in b-carotene [5], vitamin E [6] or omega-3 fatty acids [7], which are considered as second generation GMOs.

2. DEBATED SAFETY ISSUES ON GMOs

Development, release into the environment, and commercialization of GMO have been greatly debated, since the first GMO was sown, more than three decades ago [8]. The main questionable aspects regarding GMOs have been centered on four areas, namely, environmental concerns [9, 10], potential harm to human health [11, 12], ethical concerns interferences with nature and individual choice [13], and patent issues [14, 15].
Due to the above mentioned controversial safety aspects of GMOs, the European Union and other countries have established strict regulations including risk assessment, labeling, traceability and marketing of GMOs. One commonly established concept for the evaluation of safety assessment of GM foods is the substantial equivalence [16]. This approach is based on the assumption that traditional crop-plant varieties currently in the market that have been consumed for decades have gained a history of safe use [17]. Consequently, they can be used as comparators for the safety assessment of new GMO crop varieties derived from established plant lines. However, as these regulations not been established for all the countries, there is an “asynchronous approval” of these GM crops. In the same manner, GMO labeling and traceability differ between countries with different national legislation. For example, by the Regulation 1829/2003 of the European Union, it is mandatory to label as GMO containing when any food contain more than 0.9% of an authorized GMO, and the threshold is established at 0.5% when the GMO is nonauthorized. Meanwhile, in Australia and Japan, the threshold for labeling has been established at 1% and 5% respectively.

Although recombinant DNA technology is considered highly accurate for genetic modification, one of the main controversial issues associated with GMOs safety are the possible unintended effects, which might occur during GMO development. The unintended effects can be defined as effects that go beyond the primary expected effects of the genetic modification, and represent statistically significant differences in a phenotype compared with an appropriate phenotype control [18]. Unintended effects can be originated by rearrangements, insertion or deletions during the genetic transformations or during the tissue culture stages of GMO development [19, 20]. Alterations linked to secondary and pleiotropic effects of gene expression are some examples of unintended changes, and they could be somehow explained considering the function of a transgene, the site of integration in the genome or based on our current knowledge of plant metabolism [21, 22]. In some cases, unintended effects will be observed if they result in a distinct phenotype, including compositional alterations. Thus, the comprehensive characterization of the plant at the molecular level would therefore facilitate detection and description of the potential unintended effects in GMOs [23].

3. OMICS PROFILING IN GMOs ANALYSIS

As it has pointed out above, any GMO or its derived products have to pass through an approval system before entering on the market [24]. In consequence, there is a need for analytical tools that facilitate comprehensive compositional studies of GMOs in order to effectively investigate substantial equivalence and the potential adverse effects on the human health, including the existence or not of unintended effects. Compositional equivalence between GM crops and
conventional non-GM comparators is considered to provide an equal or increased assurance of the safety of foods derived from GM plants. To this regard, the selection of the comparator is a crucial aspect, and many questions about the need for comparing varieties grown in different areas and seasons, the key components to be analyzed, among others have been raised [25]. Following the recommendations from the Organization of Economic Cooperation and Development (OECD), compositional equivalence between GM and non GM lines can be achieved by targeted analysis focused on macronutrients and micronutrients, antinutrients and natural toxins for each crop variety [18, 26]. With the application of this strategy, 95% of the crop composition is covered [27], and some studies has proven that unintended effects could be identified [5, 28]. However, it has been pointed out that this approach is biased and some unintended effects derived from genetic transformation may remain undetected [29]. To solve this problem, a comprehensive study of GMO composition would help in the recognition of unintended effects that could not be detected using targeted analysis. For this purpose, the development and use of profiling technologies such as genomics, transcriptomics, proteomics and metabolomics have been recommended by the European Food Safety Agency (EFSA) [24]. Moreover, application of profiling analysis has been suggested by a panel of experts on risk assessment and management in those comparison studies where the most scientifically isogenic and conventional comparator would not grow, or not grow as well, under the relevant stress condition [30].

The combination of transcriptomics, proteomics and metabolomics technologies could provide a great coverage of genes, proteins and metabolites. In this context, Foodomics, recently defined as “a new discipline that studies the food and nutrition domains through the application of advanced omics technologies in order to improve consumers’ well-being and confidence” [31, 32] can provide valuable information that could be essential for GMOs detection, traceability and characterization (Fig. 1). Also, Foodomics offers unprecedented opportunities to study the molecular mechanism leading to a particular phenotype or the mechanism operating in important cellular processes, such as the response to different stresses [33, 34].

In spite of the recommendations and the great opportunities offered by omics technologies, there are some criticisms about the usefulness of molecular profiling in GMO risk assessment [35]. The main argument against their use is based on the problems of their routine use among laboratories, due to the lack of standardized and validated procedures. Another significant issue against profiling relies on the limited predictive capacity of the profiles for safety evaluation. Although molecular profiling can effectively measure relative differences in compounds between two varieties with high sensitivity, the biological relevance of such differences cannot be determined without previous knowledge of the natural variability of the crop composition [36]. Even knowing the natural variability of a compound, it is difficult to decode the biological
meaning of the detected differences in terms of food safety risk [35]. Moreover, when comparing a GM crop to its non-GM isogenic variety, it is important to grow both varieties under identical conditions to avoid the influence of other variability factors such as location, climate, season and farming practices. Furthermore, all these factors along with the natural variation of the chemical composition of a crop have to be taken into account to make this overview as complete as possible [37]. Considering all the advantages and drawbacks of molecular profiling for GMO risk assessment, their use has been generally accepted by the scientific community, as shown by the number of works reported on profiling approaches for comparative profiling analysis and/or the investigation of unintended effects in GMOs [33, 38]. In the published omics profiling studies on GMOs, it is interesting to note that some differences can be linked to genetic transformation. However, these studies demonstrate than differences between conventional varieties are in general more pronounced than the divergences observed between GM and non-GM crops. In addition, variations have been found when the same variety is grown in different environmental conditions. Representative examples of omics techniques, namely, transcriptomics, proteomics and metabolomics, will be discussed in the following sections.

3.1 Transcriptomics

In the last years gene-expression profiling techniques, such as DNA microarray, have matured and experienced a great development in terms of high-throughput, sensitivity and automatism. Linked to this aspect, the extensive optimization and standardization of gene-expression microarray has put this technique at the forefront of transcriptomic techniques. Gene expression microarray technique is based on the hybridization of specific nucleic acids and this feature can be used to measure the relative quantities of specific mRNA sequences in two or more conditions for thousands of genes simultaneously. Microarray technology is helpful for the identification of differences in comparative transcriptomics analysis, and for instance, it has been applied by Van Dijk et al. as a holistic approach to discover changes present in the natural variation of specific genes in different conditions [39]. In addition, transcriptomic analysis has also demonstrated to be a valuable profiling method to assess possible unintended effects of genetic transformation in wheat, maize, soybean, potato and rice crops (Table 1).

Gene-expression microarray technology has been applied by different research groups in comparative transcriptional studies between the transcriptional profiles in GM wheat and its untransformed counterpart. For example, Gregersen et al. analyzed the gene expression profile in developing seeds of wild type wheat and wheat transformed for endosperm-specific expression of an *Aspergillus fumigatus* phytase [40]. In their study, authors concluded that the


expression of the codon-modified *A. fumigatus* phytase gene in the wheat seed had no significant effects on the overall gene expression patterns in the developing seed. Later, Baudo et al. analyzed transgenic and conventionally bred wheat lines expressing additional genes encoding high molecular weight subunits of glutenin, suggesting that the presence of the transgenes did not significantly alter gene expression [41].

Transcriptomic profiles of GM maize have been also obtained using microarray technology. In a series of articles, Coll et al. reported the comparative study of different MON810 varieties with their near-isogenic counterpart [42, 43]. In the first study, after in vitro culturing of the maize plantlets under highly controlled experimental conditions (to avoid changes due to environmental factors), high-density Affymetrix microarray technology was used to analyze the gene expression [42]. Of the 13,339 genes represented in the microarray, 307 and 25 differentially expressed genes (DEGs) were found when comparing two MON810 varieties and their near-isogenic counterparts. However, 693 and 832 DEGs were found when comparing between non-GM varieties and between GM varieties respectively. These results suggest that the genetic background of each variety has a great influence when assessing the substantial equivalence of GM crops. In a later study, a similar methodology was applied to assess the effect of different field environments and cultural conditions over MON810 maize varieties and their counterparts [43]. Microarray data revealed a deregulation of 0.07-0.2% of the maize transcriptome when growing the plants under low-nitrogen and control conditions. The expression of 13 and 23 genes were altered between the transgenic and non-transgenic in control and low-level nitrogen conditions, respectively. However, a higher number of genes (31) were deregulated when comparing the transgenic line in low-nitrogen and control conditions, suggesting that the environmental conditions has higher influence than the genetic modification. As a common technique applied when handling microarray expression data because of its high sensitivity, real-time quantitative polymerase chain reaction (RT-qPCR) was used to confirm the gene expression. A total of 37 amplifying systems were designed and successfully applied, getting a 71.1% degree of coincidence between microarray and RT-qPCR data. The transcriptional profiles of these sequences were subjected to principal component analysis (PCA), and the results indicate that the natural variation of gene expression between the varieties and conditions is larger than the variation due to the genetic modification.

A comparative transcriptomic study on glyphosate resistant soybean and its near isogenic line were carried out by Zhu et al. [44]. Both lines were treated with glyphosate, and gene expression profiles were obtained at 1, 4, and 24 hours post-treatment using a cDNA microarray representing 27,513 genes. After the treatment, 170 genes were rapidly altered in glyphosate sensitive soybean, while transcript changes in glyphosate-resistant soybeans were minor or negligible due to the empirical false discovery rate (FDR). In addition, two genes out of 27,513
were found altered when comparing transcriptional profiles between cotyledons from resistant and sensitive lines, so the authors concluded that there were not unexpected consequences at the transcriptome level associated with the genetic modification. In a separate report, Cheng et al. investigated the genotype of five soybean cultivars, two transgenic and three conventional, grown in the same conditions [45]. In their study, authors used Affymetrix Soybean GeneChip to analyze 25 soybean samples, and the resultant gene expression profiles were subjected to PCA and unsupervised hierarchical clustering. Both multivariate analyses demonstrated that the GM cultivars did not cluster into a separate group from traditional cultivars. These results suggest that transgene insertion had negligible effects on global gene expression. Pairwise comparison between the GM cultivars and the non-GM counterparts showed that the number of DEGs identified was lower than those obtained in the comparison between non-GM varieties. Furthermore, only the genes cysteine protease inhibitor and dihydroflavonol-4-reductase were down-regulated in both transgenic cultivars compared with their non-GM counterpart. However, it could not be concluded if these changes proceed from the natural variation of the parent genotype, an effect of the transgenic product or an effect of the insertion event.

In addition to these studies, Baroja-Fernández et al. have focused their studies on the characterization of GM potatoes [46]. Agilent microarray slides containing 46,345 genes were used to obtain the transcriptomic profiles from potatoes with modified sucrose synthase (SuSy) gene. In SuSy-antisensed tubers, the expression of 357 genes was found to be dysregulated compared with its non-GM counterpart; however, 118 genes were deregulated in SuSy-overexpressing tubers. In spite of these results, the SuSy-overexpressing tubers exhibited a substantial increase in starch, UDPglucose and ADPglucose content when compared with controls. Nevertheless, there were no changes in the expression of genes that encode for enzymes directly involved in starch and sucrose metabolism.

Rice plants with genetic modifications have been studied several times by various groups. For instance, Batista et al. studied the extent of transcriptome modification through transgenesis and mutation breeding [48]. Gene expression of two stable transgenic plants, two γ-irradiated stable mutants, and the corresponding unmodified original genotypes were analyzed by Affymetrix GeneChip Rice Genome Array, which covers 51,279 genes. Hierarchical clustering of gene expression profiles showed that modified genotypes always grouped with the respective unmodified controls. Although the authors concluded that the use of either mutagenesis or transgenesis may cause alterations in the expression of untargeted genes, the alterations were more extensive in mutagenized than in transgenic plants for all the cases studied. Affymetrix GeneChip Rice Genome Array was also applied by Beatty et al. to investigate the transcriptional profiles of roots and shoots of GM rice over-expressing alanine aminotransferase (AlaAT) [49]. A higher number of DEGs were found in roots (55) than in shoots (36) in the transgenic line,
corresponding to 0.11% and 0.07%, respectively, of the rice genome. Although authors could not found genes directly related with aminotransferase activity, nitrogen transport and assimilation, higher level of some amino acids, total nitrogen content and grain yield were found in transgenic lines compared with control plants. Later, Montero et al. analyzed the transcriptome profile of GM rice expressing the AFP antifungal protein, and its non-GM counterpart using the same microarray platform [50]. Of the 51,279 genes represented in the microarray chip, 0.4% DEGs with at least a two-fold increase or decrease were found in the GM variety over the near isogenic variety. The expression of 34 gene sequences was confirmed by RT-qPCR with 82% of agreement between the RT-qPCR and the microarray data. The analysis of the expression of the confirmed genes suggested that 35% and 15% of the detected differences could be attributed to procedure used to obtain GM plants and the transformation event, respectively. Thus, only around a 50% of the transcriptional unintended effects could be associated to the transgene itself. More recently, an insect-resistant rice variety has been investigated by Liu et al. to detect potential unintended effects (susceptibility to rice brown spot mimic lesion and sheath blight disease) of the insect-resistant transgenic KDM rice [51]. Using the Affymetrix GeneChip Rice Genome Array, 680 DEGs were found when comparing gene expression profiling of the GM and its non-GM counterpart. To know the pathways and biological functions more altered after the genetic modification, DEGs were subjected to functional enrichment analysis using the bioinformatics tool Plant MetGenMAP. Among the 17 significantly changed pathways, 8 were directly implicated in plant stress and defense responses, and the other 9 were directly associated with plant amino acid metabolism. Amino acid profiling using isobaric tags for relative and absolute quantification (iTRAQ) and liquid chromatography (LC) coupled to mass spectrometry (MS) (iTRAQ®-LC-MS/MS) technique was performed to confirm the transcriptomics results. These analyses showed changes in 10 amino acids and in γ-amino-n-butyric acid, a typical stress response amino acid in plants.

3.2 Proteomics

The study of proteins is especially interesting in food safety because they may act as toxins, antinutrients or allergens [52]. Proteomics, a high-throughput technology able to quantify hundreds of proteins simultaneously, has become very important in comparative studies of GM plants and their non-modified counterparts [53]. Two conceptually different strategies can be followed in comparative proteomics: the shotgun and the bottom-up approaches.

In the shotgun proteomics approach, protein digestion is performed without any pre-fractionation or separation of the proteome. The resulting peptides from the protein hydrolysis are generally separated by LC or capillary electrophoresis (CE) followed by MS analysis,
providing rapid and automatic identification of proteins in the sample. Although this strategy has already been demonstrated to be a suitable strategy for protein profiling, it has been barely applied in the field of GMOs. Simó et al. investigated the unintended proteomics effects in herbicide resistant-GM soybean by the application of CE coupled to a time-of-flight (TOF) mass analyzer and electrospray ionization (ESI) [54]. Optimization of several parameters affecting the CE-ESI-TOF MS separation and detection were carried out in this work in the first stages of the method development. Once the conditions were reached, 151 peptides were automatically detected for each soybean line, but no statistically differences between the samples were found. Luo et al. quantified differences in protein profiles between GM rice and its near-isogenic line combining the shotgun approach with the iTRAQ technique [55]. Four different digested samples were treated with four independent isobaric reagents, designed to react with all primary amines of protein hydrolyzates. Rice-endosperm treated samples were subsequently pooled and analyzed by tandem MS. The analyses revealed significant differences in 103 proteins out of the 1,883 identified between GM and wild-type rice.

Unlike the shotgun approach, bottom-up proteomics approach has been widely applied to investigate the substantial equivalence and potential unintended effects in GMOs. This strategy involves the application of two-dimensional gel electrophoresis (2-DGE) followed by image analysis. Excision of the proteins from the gel spots and hydrolysis with trypsin is mandatory prior to the identification of the differentially expressed proteins (DEPs) by MS. 2-DGE analysis have some advantages and drawbacks: on the one hand, it provides the highest protein resolution capacity with a low cost of instrumentation; on the other hand, highly hydrophobic proteins, with extreme isoelectric points or high molecular weight are difficult to analyze using this methodology. In addition, the gel-to-gel variation is the one of the major sources of error. For the protein identification, matrix-assisted laser desorption/ionization (MALDI) coupled to a TOF mass spectrometer, or different variants of LC-MS are used. 2-DGE separation technique has been applied for comparing the proteome between the GM and the non-GM counterpart of different crops, such as, maize, soybean, potato and tomato among others.

Albo et al. compared the proteomic profiles obtained from insect-resistant GM MON810 maize and its non-GM counterpart using 2-DGE/MALDI-TOF/TOF MS [56]. Some unintended effects were found among the GM and the non-GM maize, such as the glucose and ribitol dehydrogenase spots uniquely found the GM maize, or the endochitinase A spot only found in non-GM maize. Also, some spots were overexpressed (triosephosphate isomerase 1 and globulin-1 S) and other spots were down-regulated (cytosolic 3-phosphoglycerate kinase and aldose reductase) in the GM maize respect to the non-GM counterpart. In a different study, Zolla et al. used 2-DGE/LC coupled to an ion trap (IT) mass spectrometer for the analysis of two subsequent generations of the MON810 maize variety and their counterparts in different
environmental conditions [57]. The comparison between the non-GM and GM maize grown in different environmental conditions revealed 100 differentially expressed proteins. On the other side, only the expression level of 43 proteins was altered in transgenic seeds with respect to their controls when controlled growing conditions were used. With these results, the authors concluded that environment plays the main influence on proteomic profiles of transgenic seeds.

In a later work, 2-DGE/LC-ESI-IT MS was also used by Coll et al. to analyze the proteome profiles of two different MON810 maize varieties [58]. A small number (≤1.2% analyzed proteins) of quantitative differential spots were observed between the GM and non-GM varieties. In addition, all differences were variety-specific and thus could not directly be attributed to the MON810 modification.

Comparative protein profiling analysis has also been carried out in potatoes using different varieties, environmental conditions and genetically modified lines [59]. 2-DGE/LC-ESI-IT MS analysis showed statistically significant differences in 1,077 of 1,111 protein spots between different varieties and landraces, respectively. Comparing GM lines and their non-GM counterparts, only 9 proteins out of 730 exhibited significant differences. Multivariate PCA indicated clear separation between several genotypes, but GM potatoes could not be separated from their non-GM counterparts.

Brandão et al. also applied the bottom-up approach to compare seed-protein profiles of herbicide-tolerant soybean and its near-isogenic line [60]. Some 2-DGE parameters such as the loaded mass of the proteins, the pH separation range, and manual/automatic image editing were optimized prior to the evaluation (Fig. 2). Of the 10 proteins with at least 90% variation found to be differentially expressed between the GM soybean and its counterpart, 8 proteins were successfully identified by MALDI coupled to a quadrupole (Q)-TOF MS.

2-DGE has been sometimes substituted by differential in gel electrophoresis (DIGE) technique to prevent gel-to-gel irreproducibility. In DIGE, different samples are labeled with ultrahigh-sensitive fluorescent dyes, typically Cy5 and Cy3, and then loaded in the same gel. After separation, gel images obtained are processed with specific software for comparative analysis using two different detection channels, allowing the simultaneous detection of protein spots labeled with the two fluorescent dyes. DIGE in combination with MALDI-Q/TOF MS was used by Barbosa et al. to compare profiling proteomes of GM soybean and its non-GM control (Fig. 3) [61]. The proteomic data obtained could be correlated with results from enzymatic determination of catalase, superoxide dismutase, glutathione reductase and ascorbate peroxidase, suggesting higher oxidative stress in the transgenic soybean seeds. DIGE and MALDI-TOF/TOF MS techniques have also been combined to study the effect of transgenes, conventional genetic breeding, and natural genetic variation in the proteome of two insect-
resistant transgenic rice [62]. Multivariate PCA of protein profiles could not differentiate between the GM varieties and their non-GM counterparts, but it could separate the non-GM varieties. The highest differences were found between the indica and japonica cultivars, followed by the three indica varieties, and finally between the GM rice and their counterparts. Univariate analysis of variance (ANOVA) of the 6 rice proteomes allowed the detection of 443 DEPs and the identification of 234 proteins by the use of MALDI-TOF/TOF MS technology. Most of the identified proteins were related with metabolism, protein folding and modification, and defense response.

CE and LC techniques can be good substitutes of 2-DGE for protein separation. These techniques can be coupled directly to a mass spectrometer, require lower amount of starting material, provide full automation, high-throughput capabilities and have better reproducibility than 2-DGE in terms of qualitative and quantitative analysis. Based on this concept, CE-ESI MS with different mass spectrometers (TOF and IT) were compared for protein profiling of insect-resistant transgenic maize [63]. The performance of both analyzers showed similar sensitivity and repeatability when analyzing intact zein-proteins fraction extracted from three different maize varieties and their near-isogenic lines. Although the CE-ESI-TOF MS provided more identified proteins, the differences between the GM and no-GM maize were not statistically significant. In another report, García-López et al. developed a LC-ESI-IT MS method to compare albumin, globulin, prolamin, and glutelin proteins fractions isolated from several insect-resistant maize varieties and their non-transgenic lines [64]. Some differences could be found between maize from diverse regions, but not between the GM and non-GM lines.

3.3 Metabolomics

Metabolomics involves identification and quantification of a high number of metabolites that are substrates, intermediates and end products of cellular activities. Metabolite profiling has been used to characterize the biological variation of the metabolic composition in commercial maize hybrids by the environment and/or genotype [65]. Also, this technology could be used to investigate the effectiveness of the genetic engineering procedure, as it is frequently used to obtain optimal production of plant metabolites, which may directly benefit human health and plant growth [66]. In addition, metabolomics has the potential to play an important role in GMO analysis, allowing the detection of intended or not intended effects, which may occur due to the genetic transformation [26].

A typical metabolite profiling analysis involves the following steps: i) metabolite extraction that often has to be adapted on a case-by-case basis depending on the type of sample and analytical
platform chosen; ii) sample preparation which may include partial purification and derivatization steps; iii) instrumental analysis of sample; iv) detection and quantification of metabolite signals to generate a data matrix that summarizes the detected signals and their intensity data; v) statistical analysis of metabolite profiles [67].

Owing to the extraordinary diversity of the chemical structure and physicochemical properties of metabolites, there is no a single analytical platform or methodology capable to detect, quantify, and identify all metabolites in the same analysis. Two major analytical platforms are currently used for metabolomics analysis: nuclear magnetic resonance (NMR) and MS. These techniques, stand alone or combined with separation techniques (LC-NMR, gas chromatography (GC)-MS, LC-MS and CE-MS), are complementary and frequently, used in parallel in metabolomics research [68].

3.3.1 NMR

NMR has been applied for metabolite profiling in different crops, such as maize, potato, wheat and lettuce. For instance, Manetti et al. studied the different accumulation of metabolites in hydro-alcoholic extracts of GM MON810 maize seeds and their non-GM control. Authors of this work found that the levels of choline, asparagine, histidine and trigonelline were lower in GM maize than in their controls [69]. Hydro-alcoholic extracts of GM MON810 maize seeds were also analyzed by Piccioni et al. with 1-D and 2-D NMR techniques [70]. In this case, 40 water-soluble metabolites were identified, and ethanol, lactic acid, citric acid, lysine, arginine, raffinose, trehalose, R-galactose and adenine were identified for the first time in the 1H-NMR spectrum in maize seeds. Also, PCA carried out with the metabolite profile, enabled the discrimination of the transgenic seeds from the non-transgenic ones.

In another report, Defernez et al. applied a similar approach to study different lines of GM and control potato samples [71]. Metabolite profiles were firstly subjected to a multivariate analysis for an initial exploration, followed by univariate analysis to confirm which compounds were mainly responsible for the differences found. Proline, trigonelline, and other phenolic compounds were statistically different between the GM and control potato samples. However, the most obvious differences were seen between the non-GM varieties studied. Transgenic potatoes expressing human beta amyloid, curdlan synthase or glycogen synthase were also analyzed by Kim et al. using 1H-NMR [72]. The data obtained were submitted to PCA, and no differences were obtained between transgenic and control lines. In the case of GM pea plant, 1H-NMR profiles, analyzed by PCA and other multivariate tools, failed on providing an acceptable classification of the GM pea plant, the null segregant control without the transgene,
and the parental line [73]. Similar results were found by Sobolev et al. in transgenic lettuce with enhanced growth properties [74]. The comparison of 24 hydro-soluble metabolites detected by NMR in 180 samples could not differentiate between the GM and the non-GM counterparts.

3.3.2 MS-based technologies

Compared to NMR, MS-based procedures have been more used in metabolite profiling of GMOs. The main advantage of MS is its higher sensitivity, and when coupled to GC, LC or CE, higher resolution and sensitivity could be achieved for low-abundance metabolites [75].

GC-MS combines high separation efficiency and reproducibility due to the stable ionization achieved by electron impact (EI), and is one of the most popular profiling techniques to study the metabolome of GMOs. The first report using this technique was carried out by Roessner et al. that characterized the metabolite profile of a transgenic potato tuber variety with modified sugar or starch metabolism [76]. After polar extraction followed by a methoximation and silylation, 77 out of 150 compounds detected were identified by comparison of the obtained spectra with commercially available spectra from MS libraries. The identified compounds provided valuable information of the altered metabolic pathways and the unexpected changes in the GM potatoes. When using the same methodology for the analysis of transgenic potatoes with altered sucrose catabolism, the same group reported increased levels of amino acids [77].

PCA was also applied in other studies of the same group, allowing the discrimination of the GM crop lines from the respective non-GM line [78]. Multivariate analysis has been also applied for the statistical analysis of GC-MS data obtained from a transgenic potato designed to contain high levels of inulin-type fructans [79]. Flow-injection analysis (FIA)-ESI MS was used to analyze 600 potato extracts and 2,000 tuber samples were analyzed by GC-TOF MS. Among the 242 metabolites detected, the chemometrics analysis could not differentiate between the GM and the non-GM potatoes.

The derivatization and analysis GC-MS method developed by Roessner et al. [76] was also applied by other authors to investigate the metabolomics profiling of a tryptophan (Trp)-enriched GM soybean [80]. It was reported that out of the 37 total organic acids, sugars, alcohols and phenolic compounds identified, fructose, myo-inositol and shikimic were found in higher concentration in GM leaves. Also, the metabolomic analyses of embryogenic cultures exhibited higher levels of malonic acid and urea and lower levels of β-hydroxybenzoic acid and galactose in GM soybean.

Other GC-MS methods have been developed to investigate transgenic rice. For instance, Zhou et al. applied a GC-flame-ionization detection (FID) and GC-MS to study the unintended effects
of different insect-resistant transgenic rice [81]. The resulted data were analyzed by partial least squares-discriminant analysis (PLS-DA) and PCA, suggesting that both the environment and the gene manipulation had remarkable impacts on the contents of different compounds. In addition, the levels of sucrose, mannitol and glutamic were increased in GM rice. In another study, carotenoid biofortified GM rice and five conventional rice cultivars were analyzed by GC-TOF MS [82]. PCA carried out with the 52 identified metabolites could separate the pigmented and non-pigmented rice samples (Fig. 4). However, transgenic rice could not be distinguished from the non-transgenic counterpart, suggesting that natural variation between varieties is higher than the differences between GM and non-GM isogenic lines.

Owing to its relevancy in metabolite profiling analysis, several groups have investigated the suitability of different metabolite-extraction procedures for GMO analysis. Selective extraction techniques, such as supercritical fluids or accelerated solvents, have been applied to investigate unintended effects in GM soybean [83] and maize [84]. These techniques enabled the selectively extraction of amino acids and fatty acids, and the combination with GC-MS enabled their quantification. More recently, Frank et al. designed a complex extraction scheme to obtain four fractions containing major lipids; minor lipids; sugars; sugar alcohols and acids; and amino acids and amines [85]. The extraction method was applied to insect-resistant and herbicide-tolerant GM maize and their non-modified counterparts, grown at distinct locations and in different seasons. The fractions obtained were independently analyzed using GC-EI-Q MS, and PCA of the data indicated that environmental influences had more impact on the maize metabolite profiles than the genetic modification.

LC-MS has been also used for metabolite profiling of GM crops providing advantages such as versatility, wide dynamic range and reproducible quantitative analysis. LC-MS is able to separate and to analyze complex samples. LC-MS is frequently used to profile polar/non-volatile, large and thermolabile compounds, demonstrating good performance on profiling secondary metabolites and complex lipids. In addition, LC-MS can resolve and quantify multiple components in crude biological extracts typically down to the nanomolar or picomolar range.

Some interesting examples of LC-MS application in this field are the studies on transgenic rice with different modifications: endosperm flavonoids production [86], enhanced starch synthesis [87], increased tryptophan production [88], or insect resistance [89]. In the latter work, a LC-ESI-Q/TOF MS method was developed to compare acetonitrile/water, acetone/water and methanol/water in terms of metabolite extraction [89]. The PLS-DA enabled the classification of the GM and non-GM samples, being 15 metabolites the responsible of the separation. However, when including samples grown in different sites and dates, it was showed that
environmental factors played a greater role than gene modification for most of the metabolites. Also, LC-Q/IT MS coupling with different interfaces (ESI or atmospheric-pressure chemical ionization), have been used to characterize wheat overexpressing genes that confer increased fungal resistance [23]. Similarly, LC-ESI-Q MS has been used to investigate GM tomato overexpressing a grapevine gene that encodes the enzyme stilbene synthase [90].

CE coupled to mass spectrometry has been also successfully applied for metabolite profiling of GMOs. It is considered complementary to LC-MS and GC-MS, as ionic and polar thermolabile compounds can be analyzed. Although high efficiency, speed and resolution can be achieved, only moderate sensitivity is reached due to the minimum amount of sample injected. Rice overexpressing dihydroflavonol-4-reductase, that enhances H2O2 tolerance, submergence and infection by *Magnaporthe grisea*, was explored by CE-ESI-Q MS [91]. Identification of chemical compounds was performed by comparison of their m/z values and migration times with standard metabolites. Cis-aconitate, isocitrate and 2-oxoglutarate were higher in GM leaves, whereas fructose-1,6-bisphosphate and glyceraldehyde-3-phosphate were lower in GM roots. CE-ESI-TOF MS has been used for metabolite profiling of GM maize [92] and soybean [93, 94]. Levandi et al. carried out a multivariate statistical analysis of the maize metabolic profiles, finding statistically differences between the GM and non-GM lines [92]. In their study, 2 metabolites (L-Carnitine and stachydrine) out of the 27 tentatively identified were found statistically significant. A method developed by Garcia-Villalba et al. enabled the tentative identification of 45 metabolites (including isoflavones, amino acids, and carboxylic acids) in herbicide-resistant transgenic soybean [93]. Interestingly, 4-hydroxy-l-threonine seemed to disappear in the transgenic soybean compared to its parental non-transgenic line. In a separate report, herbicide-resistant transgenic soybean was characterized by Giufrida et al. using a novel chiral CE-ESI-TOF-MS method [94]. In that research, the obtained D/L-amino acid profiles were very similar for conventional and GM soybean.

Fourier transform ion-cyclotron (FT-ICR-MS) provides the highest mass resolution and accuracy, and enables the determination of the elemental compositions of metabolites, which facilitates annotation procedures for unknown compounds [95]. Direct infusion analysis of plant extract without a previous separation and/or derivatization can be achieved, however, its use is very restricted due to the equipment cost, the difficulties in hardware handling and the extremely large amount of data generated. Takahashi et al. applied this technique to elucidate the effects of the over-expression of the YK1 gene in stress-tolerant GM rice [96]. More than 850 metabolites could be determined, and the metabolomics fingerprint in callus, leaf and panicle were significantly different from one another.
3.3.3 Multi-platform strategies

The combination of more than one analytical platform for metabolomics profiling generally provides complementary results, which enables the comprehensive analysis of GMOs metabolome. As an example of this, León et al. combined FT-ICR-MS with CE-TOF-MS for the metabolic profiling of six varieties of maize, three transgenic insect-resistant lines and their corresponding near-isogenic lines [97]. The spectral data obtained in both positive and negative ESI modes with FT-ICR-MS were uploaded into MassTRIX server in order to identify maize-specific metabolites annotated in the KEGG database [98]. Interestingly, electrophoretic mobilities and m/z values provided by CE-TOF MS were very helpful in the identification of those compounds that could not be unequivocally identified by FT-ICR-MS, such as isomeric compounds. LC-MS and GC-MS have been used for the comparative analysis of grapevine varieties with enhanced response to abiotic stress and their non-modified counterparts [99]. Differences in hydroxycinnamic acid, quercetin-3-glucoside, quercetol-3-glucuronide and in the degree of polymerization in proanthocyanidins were found when comparing the profiles of phenolic compounds carried out by LC-ESI-IT MS. However, volatile secondary metabolites that belong to the classes of monoterpenes, C12-norisoprenoids and shikimates were profiled by GC-EI-IT MS, and no differences were found between GM and non-GM lines. Two transgenic ringspot virus-resistant papaya varieties and their non-modified counterparts have also been analyzed by LC-MS and GC-MS [100]. LC-ESI-Q MS enabled the detection of organic acids, carotenoids, and alkaloids, whereas GC-EI-IT MS was applied for the detection of volatile organic compounds and sugar/polyals. GM and non-GM lines could not be differentiated with the multivariate analysis of the both data platforms; however it could separate papaya samples from different harvesting times (Fig. 5). Kusano et al. applied a broader approach to compare two GM tomato varieties over-expressing miraculin glycoprotein [101]. The multivariate analysis using orthogonal PLS-DA of LC-ESI-Q/TOF MS, GC-EI-TOF MS and CE-ESI-Q/TOF MS data provided lower differences between the transgenic lines and the controls than differences observed among ripening stages and traditional cultivars.

3.4 Cross-omics studies

The information obtained by the different omics techniques can be combined to generate a broader view of GMO composition. This approach may enhance the opportunities to identify potential unintended effect, and also, the inter-relationships between the different levels of information. In the studies where several omics technologies are applied, the general trend is to perform statistical analysis on each independent omics dataset. Thus, Scossa et al. investigated GM wheat overexpressing a low-molecular-weight glutenin sub-unit at the transcriptome and
proteome levels [102]. The glutenin overexpression coincided with the down-regulation of other classes of storage proteins, evidencing the complementary potential of cDNA microarrays and 2-DGE to assess the concordance between the RNA and protein levels.

A more complete study accomplished by Barros et al. involved the use of gene, protein and metabolite profiling of two transgenic insect-resistant and herbicide-tolerant maize varieties [103]. Gene expression microarray and 2-DGE analyses were performed for transcriptome and proteome profiling, respectively, whereas 1H-NMR and GC-MS, were used for metabolome profiling. Univariate analysis of individual variables (year of harvest, agricultural practices and location) for the factor genotype (GM and non-GM lines) was applied for individual omics datasets. Interestingly, the gene expression level of maize allergen Zeam14 was lower in the GM varieties, whereas the glucose and fructose were increased in the insect-resistant maize, and the γ-tocopherol and inositol were decreased in the herbicide-tolerant line. Multivariate analysis indicated that growing seasons as well as locations had a stronger overall influence in the three levels of information of the three maize genotypes than the genetic modification.

4. FUTURE OUTLOOK AND CONCLUSIONS

The application of the different profiling technologies combined with suitable chemometrics tools (including multivariate and univariate statistical analysis) are providing valuable information about the possible effects of genetic modification at transcriptomics, proteomics and metabolomics levels. However, the lack of common standardized experimental protocols is still limiting its use in any stage of the safety assessment of GMOs. Nonetheless, implementation of these techniques in control laboratories will diminish the highly costs, and the unification and validation of analytical platforms and protocols will enable the comparison of experiments performed in laboratories around the world. In the near future, the development and application of novel methodologies such as next-generation sequencing in gene-profiling, advances of protein coverage in protein profiling, or the application of multidimensional techniques such as GC x GC or LC x LC in metabolite profiling, could make the analysis of GMOs more reliable, and in turn, more challenging. Besides, there is a demand for the availability of advanced bioinformatics tools capable of the integrating and interpreting of the huge amount of data provided by these high-throughput technologies.

Acknowledgments
This work was supported by AGL2011-29857-C03-01 project (Ministerio de Economía y Competitividad, Spain), and CSD2007-00063 FUN-C-FOOD (Programa CONSOLIDER, Ministerio de Educación y Ciencia, Spain). A.V. thanks the Ministerio de Economía y Competitividad for his FPI pre-doctoral fellowship.
REFERENCES


[24] European Food Safety Agency (2006). Guidance document of the scientific panel on genetically modified organisms for the risk assessment of genetically modified plants and derived food and feed. EFSA Communications Department, Parma, Italy


[40] P.L. Gregersen, H. Brinch-Pedersen and P.B. Holm, Transgenic Res. 14:887-905, 2005


[68] V. Shulaev, Brief Bioinform. 7:128-139, 2006


FIGURE LEGENDS

Figure 1. Ideal foodomics platform to analyze genetically-modified organism (GMOs).

Figure 2. 2-D gels of soybean seeds in the optimized conditions of the applied mass of protein at pH 3–10 range (A) and pH 4–7 range (B). Reproduced from reference [60].

Figure 3. DIGE analysis of soybean seed proteins. Spots with expression variation among samples of transgenic (T) and nontransgenic (NT) soybean seeds. Down-regulated proteins (2 and 4) and overexpressed proteins (1 and 3) are shown. Reproduced from reference [61].

Figure 4. Selected ion chromatograms of metabolites extracted from non-transgenic rice (cv. NDB) (a) and transgenic rice (PAC) (b) as Methoxime/Trimethylsilyl derivatives separated on a 30 m 9 0.25-mm I.D. fused silica capillary column coated with 0.25-lm CP-SIL 8 CB low bleed. Peak identification: 1 pyruvic acid, 2 lactic acid, 3 valine, 4 alanine, 5 oxalic acid, 6 glycolic acid, 30 valine, 7 serine, 8 ethanolamine, 9 glycerol, 10 leucine, 11 isoleucine, 12 proline, 13 nicotinic acid, 14 glycine, 15 succinic acid, 16 glyceraldehyde, 17 fumaric acid, 70 serine, 18 threonine, 19 b-alanine, 20 malic acid, 21 salicylic acid, 22 aspartic acid, 23 methionine, 24 pyroglutamic acid, 25 4-aminobutyric acid, 26 threonine acid, 27 arginine, 28 glutamic acid, 29 phenylalanine, 30 p-hydroxybenzoic acid, 31 xylose, 32 asparagine, 33 vanillic acid, 34 glutamine, 35 shikimic acid, 36 citric acid, 37 quinic acid, 38 fructose, 380 fructose, 39 galactose, 40 glucose, 41 syringic acid, 42 mannose, 43 mannitol, 44 ferulic acid, 45 p-coumaric acid, 46 inositol, 440 ferulic acid, 47 tryptophan, 48 sinapic acid, 49 sucrose, 50 cellobiose, 51 trehalose, 52 raffinose, IS internal standard (ribitol). Reproduced from reference [82].

Figure 5. Scores plot for PC1 and PC2 from the Principal Component Analysis model of the total data set of cv. MZH colored in 4 different ways: (A) volatile organic compounds; (B) sugars; (C) organic acids; (D) carotenoids; (E) alkaloids. (□) Transgenic in May, (*) non-transgenic in May, (△) transgenic in September, (○) non-transgenic in September. Reproduced from reference [100].
### Table 1. Transcriptomics profiling of GMOs using microarray technique

<table>
<thead>
<tr>
<th>GM crop</th>
<th>Phenotype</th>
<th>Tissue</th>
<th>Donor specie</th>
<th>Genetic modification</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wheat</td>
<td>Endosperm-specific phytase expression</td>
<td>Seed</td>
<td><em>Aspergillus fumigatus</em></td>
<td>PhyA</td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td>Nutritionally enhanced</td>
<td>Seed and leaf</td>
<td><em>Triticum aestivum</em></td>
<td>Glu-A1, Glu-D1</td>
<td>[41]</td>
</tr>
<tr>
<td>Maize</td>
<td>Insect resistance</td>
<td>Seed</td>
<td><em>Bacillus thuringiensis</em></td>
<td>Cry1Ab</td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>Insect resistance</td>
<td>Leaf</td>
<td><em>B. thuringiensis</em></td>
<td>Cry1Ab</td>
<td>[43]</td>
</tr>
<tr>
<td>Soybean</td>
<td>Herbicide tolerance</td>
<td>Leaf</td>
<td><em>Agrobacterium tumefaciens</em></td>
<td>CP4 EPSPS</td>
<td>[44]</td>
</tr>
<tr>
<td></td>
<td>Herbicide tolerance</td>
<td>Leaf</td>
<td><em>A. tumefaciens</em></td>
<td>CP4 EPSPS</td>
<td>[45]</td>
</tr>
<tr>
<td>Potato</td>
<td>Starch metabolism</td>
<td>Tuber</td>
<td><em>Solanum tuberosum</em></td>
<td>Sus4</td>
<td>[46]</td>
</tr>
<tr>
<td>Rice</td>
<td>Free tryptophan accumulation</td>
<td>Seed</td>
<td><em>Oryza sativa</em></td>
<td>OASA1D</td>
<td>[47]</td>
</tr>
<tr>
<td></td>
<td>Control stress-inducible genes</td>
<td>Seed</td>
<td><em>Hordeum vulgare</em></td>
<td>BCBF1</td>
<td>[48]</td>
</tr>
<tr>
<td></td>
<td>Alanine aminotransferase over-expression</td>
<td>Root and shoot</td>
<td><em>H. vulgare</em></td>
<td>AlaAT</td>
<td>[49]</td>
</tr>
<tr>
<td></td>
<td>Fungal resistance</td>
<td>Leaf</td>
<td><em>Aspergillus giganteus</em></td>
<td>afp</td>
<td>[50]</td>
</tr>
<tr>
<td></td>
<td>Insect resistance</td>
<td>Leaf</td>
<td><em>B. thuringiensis</em></td>
<td>Cry1Ab</td>
<td>[51]</td>
</tr>
</tbody>
</table>