

- 1.
- 2.
- 3.
- 4.
- 5.
- 6.

## Molecular Profiling Techniques as Tools to Detect Potential Unintended Effects in Genetically Engineered Maize

*Eugenia Barros*

### Introduction

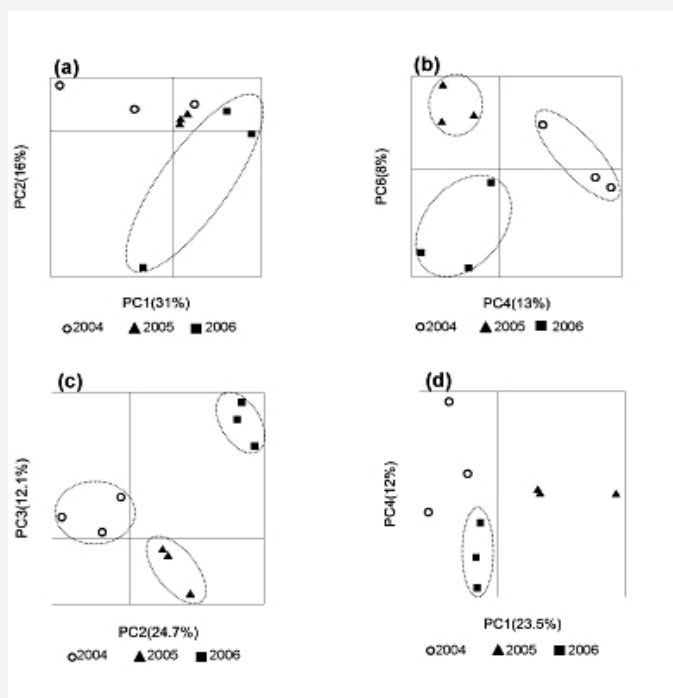
In the early stages of production and commercialization of foods derived from genetically engineered (GE) plants, international consensus was reached on the

principles of food safety evaluation. The concept of substantial equivalence became the starting point of a safety evaluation framework, based on the idea that GE foods can be compared with analogous existing foods.<sup>1</sup>

However, the controversy regarding GE plants and their potential impact on human health and the environment has necessitated the development of additional methods of risk assessment. Risk assessment focuses on adverse unintended effects that could potentially result from random transgene integration. Unintended effects can also occur in conventional breeding from mutagenesis or hybridization and backcrossing. Other factors totally unrelated to genetic engineering can contribute to alterations, including plant genetic characteristics (cultivar, isogenic lines), agronomic factors (soil, fertilizers), and environmental influences (location, weather, stress). These factors need to be considered during “GE versus non-GE” evaluations. The best way to detect unintended effects is through non-targeted analysis by using profiling technologies.

### Targeted versus non-targeted approach to detect potential unintended effects

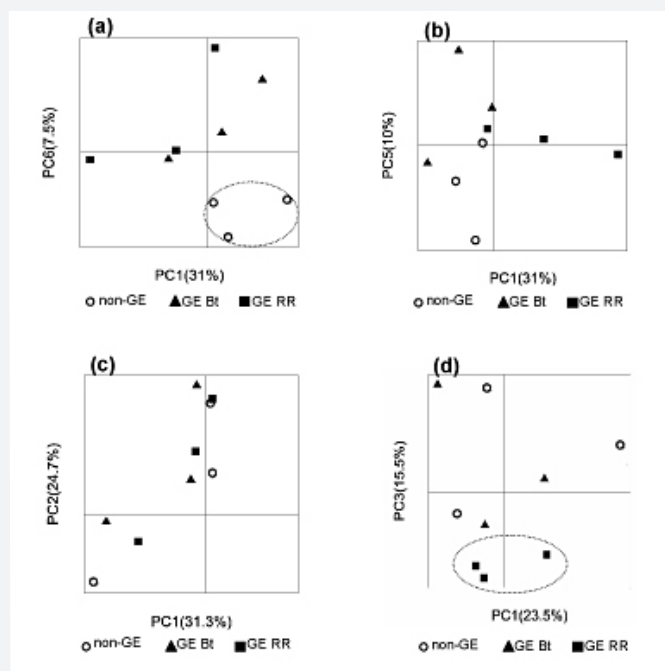
The evaluation of GE plants using targeted analysis looks at compositional variation in the GE plant compared to the non-GE counterpart, using a selection of typical crop analytes that are representative of the main metabolic pathways. These key compounds have been determined by international standards to form the basis of substantial equivalence. The substantial equivalence approach was adopted by regulatory bodies to ensure that GE plants/foods are as safe and nutritious as their conventional counterparts.<sup>2</sup> The analytes or key compounds included in the baseline analysis of targeted studies include proteins, carbohydrates, fats, vitamins, and other nutritional/anti-nutritional compounds that may affect the nutritional value and safety of the crop/food.<sup>1</sup> Limitations of targeted analysis include the inability to detect unknown anti-nutrients and natural toxins as well as any unforeseen, unintended effects of genetic engineering. Thus non-targeted profiling technologies have been developed that allow screening of potential changes in the physiology of transgenic plants at different cellular integration levels, such as the genome level during gene expression and protein translation, and at the metabolic pathway level.<sup>3</sup> These molecular profiling technologies are also known as “omics” technologies and refer to the comprehensive analysis of biological systems. In a recent paper published in *Plant Biotechnology Journal*,<sup>4</sup> we compared two transgenic white maize lines with the non-transgenic counterpart to investigate two possible sources of variation: genetic engineering and environmental variation, using three profiling technologies.



**Figure 1. PCA score plots of maize grown at Petit over 3 consecutive years.**<sup>4</sup> Separation between the 3 growing seasons for (a) microarray data, (b) proteomics data, (c) <sup>1</sup>H-NMR spectra, (d) GC/MS metabolite profiles.

### Molecular profiling of two GE maize varieties with a near-isogenic non-GE maize variety

In this study the extent of variation caused by genetic engineering and by environmental factors was evaluated by growing two transgenic maize varieties and their non-GE counterpart in a single location, Petit, during three consecutive years. The two GE maize lines were developed from the same near-isogenic non-GE hybrid maize variety; this line was included in this study as the non-GE counterpart. The two GE lines consisted of one GE *Bt* variety and one GE RR (glyphosate-tolerant Roundup Ready®) variety. Three profiling technologies were used, transcriptomics, proteomics, and metabolomics. These technologies generate a large amount of data, even when a limited number of samples are used. So, to obtain meaningful information from the profiles generated by each of the technologies, a two step statistical data analysis process was followed. The first step aimed to reduce the multidimensional data sets to smaller numbers of new variables that could give an indication of where the major sources of variation could be identified. This was done using Principal Component Analysis (PCA). The second step aimed at examining each component individually using Analysis of Variance (ANOVA), while taking into account all relevant features of the experimental design.



**Figure 2. PCA score plots of maize grown at Petit over 3 consecutive years.**<sup>4</sup> Separation between the non-GM and GM varieties for (a) microarray data, (b) proteomics data, (c) <sup>1</sup>H-NMR spectra, (d) GC/MS metabolite profiles.

A maize oligonucleotide microarray was used to evaluate gene expression. In total, 3,541 spots were included in the data analysis, and PCA results showed a separation of the samples according to growing season and genotype (**Fig. 1a** and **2a**).<sup>4</sup> When variation sources were investigated using ANOVA ( $P < 0.01$ ), the largest variation was also due to year—65 genes were statistically differentially expressed among the three seasons.<sup>4</sup> A much lower variation was due to genotype, where the most interesting difference was the lower level of the maize allergen *Zea m14* found in the GE varieties.<sup>4</sup> However, one obstacle in transcriptomics is a high false discovery rate, which can be attributed to a dataset that has more variables than samples.

Protein profiles were generated by two-dimensional (2-D) gel electrophoresis, and 714 proteins were analyzed by PCA. The most evident separation was from samples obtained during the three growing seasons (**Fig. 1b** and **2b**).<sup>4</sup> The ANOVA test on individual protein levels found five proteins that showed statistically significant differences ( $P < 0.01$ ) between growing seasons. When differences between genotypes were examined at the individual protein level, four proteins were detected ( $P < 0.01$ ). Unfortunately none of the proteins could be identified. Although this technology allows the comparative analysis

of protein patterns, changes in protein concentrations, and post-translational modifications, there are currently two major shortcomings. One is that only highly expressed proteins can be detected in a complex protein mixture, and the other is that there are not sufficient protein sequence data available for identification purposes.

The analysis of plant metabolites is generally complicated due to their highly complex nature and vast chemical diversity. There is, however, a range of technologies that can identify individual compounds with alterations, such as sugars, fats, acids, and other metabolites. Two such technologies, <sup>1</sup>H-NMR fingerprinting and GC/MS, were used to generate metabolite profiles of the three maize genotypes.

<sup>1</sup>H-NMR fingerprinting can be used to dissect the relationship between sequence and biological function. Although coverage of the plant metabolome is incomplete, 15,666 complex data points were produced by <sup>1</sup>H-NMR and analyzed by PCA. A distinct separation existed among the three years of cultivation, but there was no visible separation among the three genotypes (**Fig. 1c** and **2c**).<sup>4</sup> While 36 metabolites were significantly different ( $P < 0.01$ ) among the three years of planting, 15 metabolites were significantly different ( $P < 0.01$ ) between the GE plants and the non-GE counterpart, even though the changes were small.

GC/MS metabolite profiling identified 120 compounds that were included in data analysis, and a separation was observed for seasons (stronger in 2005) and for genotypes (**Fig. 1d** and **2d**).<sup>4</sup> Even though this technology provides valuable information on the structural identity of compounds, its limitation lies in the range of detectable analytes, which is dependent on the choice of solvents used in metabolite extraction as well as the restriction to low molecular weight constituents.

### Concluding remarks

Profiling techniques offer a multidisciplinary approach to the safety assessment of GE plants, and provide a rigorous scientific basis for the identification of any possible unintended health effect that could arise from genetic engineering. In this study non-targeted molecular profiling technologies were used to provide insight into the extent of variation in the maize transcriptome, proteome, and metabolome by analyzing three maize genotypes, two of them transgenic, grown in the same location over three years/growing seasons. The observed variation was caused mainly by growing season, with the associated environmental factors, and not due to

## REGULATORY

genotype. Even though the environment was the dominant source of variation, no common drivers of variation were identified.

This study also highlights the possibilities as well as the challenges of profiling analysis for food safety evaluation. A big challenge of the 'omics' technologies is the vast amount of data generated, making it extremely complex to evaluate individual GE lines and difficult to form a meaningful interpretation. Other challenges include the many gaps related to the number of genes for which a function has been identified and the limited coverage of the proteome and metabolome. Furthermore these technologies should not be limited to the safety assessment of GE-derived crops, which are already subjected to extensive

and costly pre-market analysis, but should be extended to plant varieties improved by mutagenesis. Mutagenised rice plants are reported to have more changes in their transcript profiles than genetic engineered rice plants, even though changes in the transcriptome do not necessarily correlate with risk; proteomic studies need to be performed to provide information on the nature of the proteins.<sup>5</sup> These technologies still need to be validated before they can be used on a case-by-case basis to confirm or supplement the current targeted analytical approaches. They are not intended to replace existing analyses, but they may trigger the need for a more detailed and targeted analysis of specific groups of genes, proteins, and metabolites.

### References

1. Kuiper HA, Kleter GA, Noteborn PJM and Kok EJ (2001) Assessment of the food safety issues related to genetically modified foods. *The Plant J.* 27, 503-528
2. Kuiper HA and Kleter GA (2003) The scientific basis for risk assessment and regulation of genetically modified foods. *Trends Food Sci. Tech.* 14, 277-293
3. Rischer H and Oksman-Caldentey K-M (2006) Unintended effects in genetically modified crops: revealed by metabolomics? *Trends Biotechnol.* 24, 102-104
4. Barros E, Lezar S, Anttonen MJ, van Dijk JP, Röhlig RM, Kok EJ and Engel K-H (2010) Comparison of two GM maize varieties with a near isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotech J.* 8
5. Batista R, Saibo N, Lourenço T and Oliveira MM (2008) Microarray analyses reveal that plant mutagenesis may include more transcriptomic changes than transgene insertion. *PNAS* 105, 3640-3645

Eugenia Barros  
Principal Researcher  
Council for Scientific and Industrial Research (CSIR) Biosciences  
ebarros@csir.co.za