Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics

Eugenia Barros[1],*, Sabine Lezar[1], Mikko J. Anttonen[2], Jeroen P. van Dijk[3], Richard M. Röhlig[4], Esther J. Kok[3], and Karl-Heinz Engel[4]

[1] Council for Scientific and Industrial Research (CSIR), Biosciences, Meiring Naude Road, Brummeria, Pretoria, 0001, South Africa,

[2] Department of Biosciences, University of Kuopio, Yliopistonranta 1E, P.O. Box 1627, 70211, Kuopio, Finland

[3] RIKILT Institute of Food Safety, Bornesesteeg 45, P.O. Box 230, 6700 AE Wageningen, The Netherlands, and

[4] Technische Universität München, Lehrstuhl für Allgemeine Lebensmitteltechnologie, Am Forum 2, D-85350 Freising-Weihenstephan, Germany

* Correspondence (Dr. Eugenia Barros; CSIR Biosciences, P.O. Box 395, Pretoria 0001, South Africa; tel 0027-12-8413221; fax 0027-12-8413651; email ebarros@csir.co.za)

KEYWORDS

GM • transcritpomics • proteomics • metabolomics • food safety • maize

SUMMARY

The aim of this study was to evaluate the use of four non-targeted analytical methodologies in the detection of unintended effects that could be derived during genetic manipulation of crops. Three profiling technologies were used to compare the transcriptome, proteome and metabolome of two transgenic maize lines with the respective control line. By comparing the profiles of the two transgenic lines grown in the same location over three growing seasons we could determine the extent of environmental variation, while the comparison with the control maize line allowed the investigation of effects caused by a difference in genotype. The effect of growing conditions as an additional environmental effect was also evaluated by comparing the Bt maize line with the control line from plants grown in three different locations in one growing season.

The environment was shown to play an important effect in the protein, gene expression and metabolite levels of the maize samples tested where 5 proteins, 65 genes and 15 metabolites were found to be differentially expressed. A distinct separation between the three growing seasons was also found for all the samples grown in one location. Together, these environmental factors caused more variation in the different transcript/protein/metabolite profiles than the different genotypes.

INTRODUCTION

Cereals are among the most important group of cultivated plants for food production worldwide. Maize (*Zea mays*) is the most widely grown cereal and according to a USDA report world maize production for 2007/2008 was 791.6 million tons (USDA, 2009).

Genetic engineering of agricultural crops has played an important role in crop improvement where it has been used to increase resistance to disease and stresses and tolerance to herbicides as well as to improve the nutritive value of crops. However, food derived from genetically modified (GM) crops have often been surrounded by controversy, particularly in Europe, despite the lack of evidence of risks associated with GM crops and the extensive safety measures taken prior to their release. One of the concerns is about unintended effects that might result from the random integration of the transgene. This may cause gene disruptions that can lead to sequence changes, production of new proteins or formation of either new metabolites or altered levels of existing metabolites that could compromise safety (Kuiper *et al*, 2001; Cellini *et al*., 2004). This includes the potential production of new allergens or toxins. Other unintended effects, related to the genetic modification, may be secondary effects of the introduced sequences. Unintended effects can also come about during conventional breeding as a result of mutagenesis, as well as hybridization and backcrossing that are integral processes of breeding programs where the genetic variation within species and between related species is used as a major source for crop improvement.

The safety assessment of GM crops is based on the principle of substantial equivalence or comparative safety analysis (OECD, 1993; FAO/WHO, 1996; Kok and

Kuiper, 2003. To this effect the GM crop is compared to its conventional counterpart at the agronomic/phenotypic level and by compositional analysis. The latter will include analysis of macro- and micronutrients as well as toxins and anti-nutrients. Because of the varied nature of GM crops the evaluation is done on a case-by-case basis (Kleter and Kuiper, 2002). The OECD has developed consensus documents for the crops of major economic interest that provide overviews of the most relevant nutrients and anti-nutrients for these crops. These documents are used as guidelines for the comparative compositional analysis. Also for maize such a consensus document has been published (OECD, 2002; OECD, 2006). Targeted analyses of key compounds have been used extensively in substantial equivalence studies of GM crops and have contributed to the establishment of databases with detailed information on the composition of some major conventionally bred crops that serve as benchmark for the assessment of the composition of the GM crops. The best example of such a database is the ILSI Crop Composition Database that has compiled data on maize, cotton and soybean (ILSI, www.cropcomposition.org). Importantly, the only data that is included has been obtained with validated methods of targeted analysis.

Profiling technologies such as transcriptomics, proteomics and metabolomics have been suggested to broaden the spectrum of detectable compounds and thus to supplement the current targeted analytical approaches (Kuiper *et al.*, 2001; Cellini *et al.*, 2004; Lehesranta *et al*, 2005; Metzdorff *et al*., 2006; Zolla *et al.,* 2008, Kok *et al.,* 2007).

In this study we used non-targeted molecular profiling to provide insight into the extent of variation in the maize transcriptome, proteome and metabolome by analyzing three maize genotypes. These included two transgenic lines modified for two different genes and the respective control line. All are commercial lines that were

grown in different locations. We report on the application of cDNA microarray for transcriptome profiling, two-dimensional gel electrophoresis for proteome profiling and ^{1}H-NMR fingerprinting and capillary gas chromatographic / mass spectrometric (GC/MS)-based metabolite profiling for analysis of the metabolome. One of the characteristics of molecular profiling is the large amount of data generated. We used multivariate data analysis for an initial exploration followed by univariate analysis to identify the genes/proteins/metabolites that were mainly responsible for the differences between the three maize lines. The data presented serve as an exploratory study into the use of ~omics techniques for safety evaluation of GM crops. Both the added value and the current challenges are discussed.

RESULTS

In this study we evaluated the effects of genotype and environmental conditions on maize kernels at the transcriptome, proteome and metabolome levels. To this effect, two GM maize varieties (GM Bt, GM RR), each containing a single insert, were subjected to comparative profiling, using the near-isogenic non-GM variety CRN3505 as the comparator. These varieties were compared in a single location, Petit, during three consecutive years. This set-up allowed the evaluation of variation caused by year of harvest and genotype as independent factors. In addition, a comparison between the non-GM and the GM Bt variety was done between three different growing locations during one growing season. This set-up allowed evaluation of effects of the factor genotype independent from differences in growing

conditions, including organic or conventional farming, as one of the locations (Potchefstroom) employed organic cultivation.

Effects of genotype and growing season: Comparison of two GM maize and one non-GM plant in one location (Petit) over 3 years

Microarray analysis

After initial data selection a total of 3541 spots were included in the data analysis. The PCA analysis of these spots allowed the characterization of samples according to growing season or genotype (Figures 1a, 2a). For both factors a separation was observed in the score plots with different combinations of components, with the separation for growing season explaining more variation in the dataset (Figure 1a).

Separate one-way ANOVAs for the factors growing season and genotype were then performed to identify differentially expressed genes. This was followed by a Tukey's HSD test to find which of the groups of samples were different from each other (P=0.01). For growing season, 69 spots were significantly different (P<0.01), of which 65 showed a significant difference for at least one of the years when Tukey's HSD post-testing was performed. The ten most significant spots showed an FDR between 20 and 24%. Within the top ten spots (Table 1), five were either not annotated or showed homology to unknown proteins or hypothetical proteins with unknown function. Of the annotated spots, only the putative ribosomal protein L26 was represented twice in the dataset with one spot for this gene not showing differential gene expression. The largest differential gene expression (non-

6

transformed) within the 65 spots was 1.6-fold, between years 2005 and 2006 for the spot with the lowest *P*-value.

For genotype, a total of 33 spots were significantly different (*P*<0.01) with two spots showing an FDR of 11 and 29%. The rest of the spots showed an FDR of 93%, but included three spots for the nonspecific lipid-transfer protein precursor (LTP), also classified as maize allergen m14 (Pastorello *et al.*, 2000) (Table 2). For this gene, expression values were significantly lower in the Bt variety compared to the non-GM variety. The largest difference in average expression was found for spot MOA21941, with a 2.2 fold difference between the Bt variety and the non-GM variety. In all cases the non-GM samples showed the highest expression of this allergen. The spot with the most significant differential gene expression was a spot for GAPDH (MOA18226). However, GAPDH was represented by 28 spots in the dataset, with the other 27 not showing significantly different gene expression between genotypes. The putative serine carboxypeptidase was represented by only one spot in the dataset.

2-D electrophoresis

When all samples collected from Petit during the three consecutive years were analysed by PCA (714 proteins), the first eight components explained 100% of the variation, 31% of the variation being allocated to the first component. The most evident result was the separation of samples collected during different growing seasons (Figure 1b). The fourth component, which explained 13% of the total variation, separated 2004 samples from the other years. In addition, the sixth component explaining 8% of the total variation separated samples collected during

2005 from those collected during 2006 and to some extent from 2004 samples. Differences between genotypes were not evident (Figure 2b).

When differences between the growing seasons were examined at individual protein levels statistically significant difference (ANOVA, $P < 0.01$) was found in 5 proteins (Table 3). Separation between growing seasons in the PCA was seen in the sixth principal component, which explained only small part of the total variation. Thus, low number of differences at the protein level was expected. Proteins 2106, 2609 and 7503 separated the 2004 samples from the other ones. The 2005 samples were separated from the 2006 samples by proteins 1426, 7501 and 7503. In general, relative quantitative differences between the highest and lowest values were modest, ranging from 1.5- to 3-fold. The highest relative difference was observed for protein 2106 which had a 2.4- fold higher expression in 2004 compared to 2006.

When differences between varieties were examined at individual protein level statistically significant difference (ANOVA, $P < 0.01$) was detected for four proteins (Table 4). Interestingly, differences were seen between the non-GM and GM varieties. Proteins 4511 and 6114 were significantly different in the RR variety compared to the non-GM one. In the RR variety the intensity of the protein spot 4511 was 1.5 times higher, whereas the intensity of spot 6114 was over 6 times lower. Similarly two proteins differed between GM Bt variety and the non-GM variety. Statistically significant difference was seen in protein 5310, being 1.3 times higher in the Bt variety. The intensity of protein 6614 was 1.5 times lower in the Bt variety but the difference was not statistically significant. Unfortunately none of the proteins could be identified.

NMR fingerprinting

PCA performed on the ^{1}H-NMR data (15 666 complex data points) showed separation among the three years of cultivation (Figure 1c). Year 2005 could be separated from the years 2004 and 2006 on the third component accounting for 12.1% of the variation. There was no visible separation among the three genotypes within the first two components which accounted for 56% of the total variation observed. This suggests that the genetic modification of the two GM plants had very little impact on their metabolic pathways indicating minimal differences between the metabolomes of GM and non-GM plants (Figure 2c). The NMR spectra were examined for significant differences among the three years of planting and 36 metabolites were identified. Separate one-way ANOVA and subsequent post hoc Tukey's HSD testing showed that 15 metabolites were significantly different ($P<0.01$) between the GM plants and the non-GM counterpart. These metabolic compounds showed some changes in the level of production even though the changes were small (Table 5). The levels of the three sugars, glucose, fructose and sucrose, were higher in 2005 when compared to the other years. Interesting changes were also observed when the fifteen statistically significant compounds were evaluated for their different levels among the three genotypes (Table 6). A 13.8 fold increase in the production of glucose and a 6.9 fold increase in the production of fructose were observed in the Bt plant compared to the non-GM and the RR plants, even though this was could not be detected at the PCA level.

GC/MS profiling

The extraction and fractionation scheme applied for the GC/MS-based metabolite profiling allowed the assessment of metabolites from different chemical classes covering a broad range of polarity. Statistical assessment of 120 compounds

by PCA of the data from the samples grown at location Petit in three consecutive growing seasons (2004-2006) showed a clear separation of year 2005 on the first principal component accounting for 23.5% of total variation (Figure 1d). Years 2004 and 2006 were only slightly differentiated on PC4 (12.0% of total variation). Among the genotypes the GM RR variety could be separated from the non-GM variety on PC3 (15.5% of total variation, whereas no differentiation was observed for the GM Bt variety (Figure 2d).

Independent one-way ANOVA for the factors genotype and year resulted in 6 compounds that were significantly different ($P < 0.05$) between genotypes and 21 compounds that were significantly different ($P < 0.05$) between growing seasons. Distribution of P-values for factors genotype and year confirm that year is the dominant impact factor. For the factor genotype, the levels of γ-tocopherol and of the phytosterol cycloartenol were significantly different (Table 7). In addition, two trace compounds, which were present in the minor lipid fraction (II) showed significant P-values by ANOVA. Among the polar compounds in fractions III and IV, the contents of inositol and asparagine were found to be significantly different between genotypes (Table 7). Post hoc Tukey's HSD testing of the compounds in Table VII revealed that only three metabolites were significant different between the GM and non-GM varieties. For the factor growing season, lower levels for fatty acids and minor lipids and higher levels for amino acids were observed in samples from 2005 compared with those of the other two years. Two exceptions were glutamic acid and pyroglutamic acid that exhibited increased levels (Table 8).

Environmental effects: Comparison of GM Bt plant with the non-GM plant in three locations in one season

During the growing season 2004 samples of the GM Bt variety and the non-GM variety were collected from 3 different growing locations. In Potchefstroom organic cultivation techniques were used whereas in Petit and Lichtenburg conventional ones were used.

Microarray analysis

For the 2004 harvest year, a separate analysis was performed on the non-GM and GM Bt maize samples. PCA revealed a strong separation according to location and genotype, while in the latter case less variation was explained (Figure 3a, 4a). A t-test revealed that a total of 15 genes showed significant differential expression ($P<0.01$) due to genotype, although the difference in (non-transformed) expression was never larger than two-fold. Interestingly, there was no overlap between the 15 spots in this dataset compared to the 50 most significantly different genes for the factor genotype in the Petit subset of samples. The top ten differentially expressed genes with $P< 0.01$ are shown in Table 9.

2-D electrophoresis

For the growing season 2004 the PCA (Figure 3b) revealed clear differences in the protein profiles between samples collected from different growing locations; however, no differences were seen between the GM Bt variety and the non-GM one.

NMR fingerprinting

PCA showed no significant differences between the Bt and the non-GM maize variety that could be attributed to effects of genetic modification (results not shown). A very small difference was observed between Potchefstroom and the other two sites.

The maize in Potchefstroom was organically grown and its profile separated along the 1st and 2nd principal components, which accounted for 28.1 and 14.8%, respectively, of all the variability in the data (Figure 3c).

GC/MS profiling

PCA of metabolite profiling data of the GM Bt and non-GM varieties from the three farming locations in season 2004 showed separation of locations on PC1 and PC2 which account for 61.6% of the total variation (Figure 3d). The samples grown at location Potchefstroom where low input farming was practiced were separated clearly on PC1. On the third principal component (18.1% of total variation) the GM Bt variety could be differentiated from the non-GM samples (Figure 4b). However, t-tests ($P < 0.05$) revealed that only one glycerol had elevated levels (+101% to +1171%) in the non-GM maize.

DISCUSSION

Maize is one of the most important agricultural crops and it is part of the staple diet of humans and livestock. It has been the subject of many crop improvement initiatives where the driving forces have been to boost maize production levels. With the developments of genetic engineering a variety of transgenic maize plants have been produced with different characteristics including insect-resistant Bt-maize and herbicide-tolerant Roundup Ready maize. Current safety assessment procedures developed for GM crops are primarily based on a targeted compositional analysis of specific safety and nutrition-related compounds (OECD, 1993;

FAO/WHO, 2000a; FAO/WHO, 2000b). Targeted analysis may, however, have its limitations in detecting unintended effects in genetically modified organisms. Consequently this prompted the further assessment of non-targeted profiling/fingerprinting technologies that are unbiased analytical approaches able to detect the potential occurrence of unintended effects.

We looked at the combined effect of genetic modification and growing season by growing the two GM plants and the non-GM one in the same location/environment over three years. A distinct separation between the three growing seasons was observed for all the samples at the proteome and metabolite levels within a single component of the PCA, and for transcriptomics in the combination of the first two components. This suggests that the environment had a strong effect in protein and gene expression and metabolite production. ANOVA revealed that sixty-five genes and five proteins were statistically differentially expressed among the three seasons and fifteen metabolites, identified by NMR, were also differentially produced among the three seasons with higher levels observed in 2005. Similarly by comparative analysis using GC/MS metabolite profiling a higher fructose level was observed for year 2005. The transcriptomics data showed some differences between the GM varieties and the non-GM maize variety in the PCA as well as the ANOVA. Probably the most interesting difference at the gene expression level was the lower level of maize allergen Zea m14 found in the GM varieties. At the metabolite level glucose and fructose were increased by 13.8 and 6.9 fold, respectively, in the Bt variety compared to the non-GM variety and γ-tocopherol and inositol were 3.7 and 1.4 times higher in the non-GM variety compared to the RR variety.

The data generated revealed that growing seasons had a stronger overall effect in the transcriptome, proteome and metabolome of the three maize genotypes than the

genetic modification. This is consistent with previous publications by Baudo *et al.*, (2006), Batista *et al.,* (2008) and Cheng *et al.,* (2008), which showed similar results for transcriptomics in wheat, soybean and rice respectively. The potential unintended effects shown could very well fall within natural variability that exists among maize lines and that was beyond the scope of this study, such as different landraces, or more diverse locations and climates.

We also evaluated the consequences of genetic modification (GM Bt vs. non-GM) in different locations, including different agricultural practices. The agricultural practice in the Potchefstroom location was organic production in contrast to the other two locations that followed high input systems. The experimental set-up only allowed for a statistical evaluation of individual variables for the factor genotype, which were minor. The multivariate PCA analyses showed that a larger portion of the total variance could be linked to environmental factors than to genotype. PCAs showed a distinct separation for the three locations using transcriptomics, proteomics and metabolic fingerprinting using NMR. Metabolic profiling using GC/MS separated the location Potchefstroom from the other two locations. Only transcriptomics separated the Bt maize variety from the non-GM one, but only at the fourth and sixth component.

The use of four "-omic' technologies allowed a holistic approach to the potential unintended effects that might have been caused by genetic modification. Although the value of these technologies for the comprehensive comparison of GM and non-GM maize is in principle large, it needs to be stressed that the amount of data generated by any of these technologies is vast. In order to simplify the interpretation, identification and presentation of the data statistical tools like PCA was used in all

four technologies, complemented with ANOVA for the determination of significant differences in transcripts, proteins or metabolites.

Large scale profiling methods described here have potential to be useful in food safety assessment. Compared to the targeted methods, -omics methods can potentially give a much wider picture of food composition. Furthermore, these non-targeted methods enable the detection of unexpected or unintended changes caused by genetic modification, traditional breeding or various external factors from environmental conditions to agricultural regimen. Our study shows that -omics approaches can be used to obtain a distinct profile for food crops grown in different environments. In addition, even small differences at individual gene, protein and metabolite levels are detectable and not lost in the mass of variables.

In order for these technologies to be used for safety evaluation there are still hurdles to be overcome. One of the obstacles in transcriptomics is the high false discovery rate, which is inherent to a dataset with a lot more variables than samples. One such example in this dataset was the identification of GAPDH as a false positive due to the lack of confirmation by other spots representing the same gene within the microarray. Furthermore, many features on microarrays are not yet linked to an annotated gene, although they frequently refer to a unigene sequence: a set of overlapping cDNA sequences that together represent the most probable gene transcript. Nevertheless, transcriptomics plays a valuable role in the assessment of potential differences between two genotypes, because of the broad coverage of the plant's metabolic routes and networks compared to the other -omics approaches. The present study showed that even low transcript activity in the mature stages led to the same groupings of samples as was found with the other -omics techniques. A potential unintended effect was discovered (lower allergen m14 expression) in the Bt lines

grown in the Petit location. The difference was not found in the other Bt samples. While present in the dataset containing the three maize varieties grown in one location in different years, it was not confirmed in the dataset containing two maize varieties grown in one year but in three different locations. This could be due to a true difference which is only apparent in one location, or it could be a false positive discovery, as the FDRs were high in both analyses. However, this does not forestall the possibility of a significant change in only one or a few genes that, while not causing a lot of variation in the whole dataset, could nevertheless have food safety implications. With this study we have shown that it is likely that such differences would be identified using transcriptomics.

The major limitation in the proteome analysis is the size of the proteome and numerous possible post-translationally modified proteins. Clearly the proteomics by the 2-DE approach, able to quantify 1000 to 2000 proteins at most, does not provide the whole proteome. Moreover limited amount of protein sequence data is available for the identification purposes. However, even with less than 1000 quantified proteins it is possible to distinguish between growing seasons and locations as seen in this study or even between different agricultural production systems (Lehesranta *et al*., 2007). In addition, protein sequence databases are constantly expanding and genome sequencing projects provide further support.

The ^{1}H NMR technique, with a detection threshold of around 5 nmol, is several orders of magnitude less sensitive than other screening techniques such as MS (10^{-12} mol), resulting in an incomplete coverage of the plant metabolome. The total number of metabolites in the plant kingdom is estimated to range from 200 000 to 1 000 000 and a single strain of *Arabidopsis thaliana* is expected to produce about 5 000 metabolites (Bino *et al.,* 2004). However, only 20 to 40 (Fan *et al.*, 1988;

Sobolev *et al.*, 2003; Le Gall *et al.*, 2004; Le Gall *et al.*, 2003) metabolites have typically been identified in metabolite profiling studies of plant samples by ^{1}H NMR. In addition, a large fraction of the metabolome may be present at very low concentrations. Overlapping signals and the dynamic range problem is a major hindrance to the identification of minor components of the metabolome (Krishnan *et al.*, 2004). In this study metabolite fingerprinting using ^{1}H NMR proved to be a fast, convenient and effective tool for discriminating between groups of related samples through the identification and quantification of fifteen significantly produced metabolites. When the three maize varieties were grown in the same location (Petit) over three growing seasons while being subjected to the same high throughput agricultural system ^{1}H-NMR data showed some separation (36.8%) among the three years of cultivation while no clear separation was observed between the non-GM and the GM maize varieties. This variation could have been the result of variation in climatic conditions which includes rainfall or any other environmental variation that occurred over the three year period. A similar observation of environmental variation due to location or to agricultural conditions was observed for the maize samples (one GM and one near-isogenic non-GM) grown in three different locations.

The suitability of GC/MS for the detection, identification and quantification of a comprehensive set of metabolites has been demonstrated for various crops, such as rice, maize and soy bean (Frank *et al.*, 2007; Hazebroek *et al.*, 2007; Frank *et al.*, 2009). The approach used in this study utilizes sub-fractionation to obtain profiles of metabolites from different chemical classes ranging from lipophilic to polar. The influence of genetic modification was assessed under different environmental conditions. Comparison of metabolite profiles revealed the effect of environment (location, year) to be more pronounced than the genetic background (GM, non-GM)

of the samples. The levels of four metabolites were significantly different in at least one of the GM lines. However, the ranges of these compounds still overlapped if all non-GM and GM lines analyzed in this study were taken into consideration. Moreover, the most pronounced difference (11-fold) detected between non-GM and GM Bt for glycerol turned out to be a one-year-effect. These observations are in agreement with the results of a recent study in which the metabolite profiles of maize cultivars differing in maturation behaviour were assessed in three consecutive years (Röhlig *et al.*, 2009). Whereas the influence of the cultivars could be clearly shown within the single years, combination of samples from all seasons revealed the environmental impact to be the most prominent impact factor. For application of GC/MS based metabolite profiling in the area of safety evaluation, the restriction of the applied approach to low molecular weight constituents has to be considered. In addition, although GC/MS metabolite profiling is an un-biased technique in principal, the range of detectable analytes may be narrowed by the choice of solvents for metabolite extraction.

In conclusion the use of the four "-omic" technologies highlighted the potential of each of these approaches in identifying the main sources of variation in transcript, protein and metabolite levels. Although the sources of variation in the dataset were the same for all the techniques used (environment being the dominant one), no functional correlations were identified between the genes, proteins and metabolites driving this variation. For an optimal application of -omics techniques, enough samples should be available for assessing interactions between environmental factors and genotypes, genes, proteins and metabolites should be fully annotated and preferably, several stages of maturation should be investigated. This particular study highlighted the possibilities and challenges for profiling/fingerprinting analysis in

food safety evaluation, be it GM-related or otherwise. The use of these technologies for risk assessment should however be considered on a case-by-case basis rather than as a routine method.

EXPERIMENTAL PROCEDURES

Plant material and plant growth

The white maize samples used in this study were derived from the transgenic Bt hybrid variety DKC78-15B (hybrid of event MON 810 from Monsanto), from the transgenic glyphosate-tolerant Roundup Ready variety DKC 78-35R (hybrid of event NK603 from Monsanto) and the near-isogenic non-GM hybrid variety CRN 3505 (maize line from Monsanto from which DKC78-15B and DKC78-35R were developed). The plants were grown in two different sites, namely, Petit and Lichtenburg (South Africa), under high-input system; the varieties were planted in Petit over three growing seasons (2004, 2005 and 2006) and in Lichtenburg over one growing season (2004). At planting the plants were fertilized with 300kg/Ha 4:3:4 (33), Topdressing 300kg/Ha KAN (28) and treated with herbicide 1.8L/Ha Guardian + 200M1/Ha Sumi Alpha. Two months after planting the plants were treated with herbicide, 2.2L/Ha A-maizing + 1L/Ha Harness + 220Ml/Ha alphacypermytrin. Three months after planting the material was treated with pesticide, 750Ml/Ha Endosulfan against stalkborer. The plant material was harvested 8 months after planting; the kernels were removed from the cobs on site by machine and packed in plastic bags. The moisture content was 11-13%. For the field trial performed at location Petit in

2005 three replicate samples were available and the results were averaged prior to further analysis for all techniques. For all other field trials one sample was analyzed.

The DKC78-15B and the control variety CRN3505 were also grown in Potchefstroom (South Africa) under low-input system which means that no fertilizer, no fungicide and no herbicide were applied throughout the growth of the plants. The plant material was also harvested after the cobs were dry around 8 months after planting.

Preparation of samples for omics analyses

The maize kernels were delivered to Technische Universität München, Germany for milling and distribution. This was done to reduce the technical variation and obtain an 'average' sample that could be used for analyses using the different technologies. The maize kernels were milled using a cyclone mill equipped with a 500-µm sieve and freeze-dried for 48 hours. Aliquots of maize powders (2g) were prepared and delivered to the different laboratories for specific analyses. Upon arrival the maize powders were kept at -20°C until use.

Microarray analysis

RNA extraction and sample preparation

RNA was isolated from 0.4 g of freeze-dried powder from maize kernels. The protocol for RNA extraction which is based on CTAB and consecutive chloroform/isoamylalcohol extractions with an overnight LiCl precipitation (Chang *et al.,* 1993) was used with the following modifications: the extraction buffer was heated to 60°C before use, the chloroform/isoamylalcohol extraction was repeated three

times before LiCl precipitation and the final precipitation of RNA in 96% ethanol was performed by cooling the tubes on ice and centrifuging at 4°C for 15 min at 14.000 g. The RNA was dissolved in 100 μl of 1mM Tris (pH 7) by heating to 65 °C for 10 minutes. RNA concentration and purity was then assessed from the absorbance measurements with the Nanodrop 1000 instrument.

Fluorescent labeling of cDNA and hybridizations

For each sample, RNA (100 μg) was labeled by incorporation of Cy3-dCTP during a cDNA synthesis reaction using 21-mer oligo-dT primers according to the method described by Boeuf *et al.,* (2001) and Franssen-van Hal *et al.,* (2002). Labeled cDNA was dissolved in MilliQ-treated water (500 μl) and 2x hybridization buffer (500 μl, Agilent, Amstelveen, the Netherlands) pre-warmed to 60°C. The DNA probes were then immobilized on maize arrays that were obtained from the Maize Oligonucleotide Array Project (U.S.A.). The microarrays consisted of 57K spots on two slides and were produced at the University of Arizona as part of the National Science Foundation Plant Genome Research Program (Gardiner *et al.*, 2005). The DNA probes were immobilized by rehydration above a 50°C water bath for 5 seconds, drying on a 45°C heating block for 5 seconds, and cooling for 1 minute at room temperature (RT). This was repeated four times after which the slides were UV cross-linked at 180 mJ. The slides were then washed in 1% SDS at RT while stirring, followed by dipping 10 times in MilliQ-treated water and 5 times in ethanol (100%). Finally the slides were incubated for three minutes in ethanol (100%) at RT and dried by centrifugation for 5 minutes at 32 g in a table centrifuge CR3i with swing–out rotor T20 (Jouan, France).The slides were prehybridized according to the protocol described by Hedge *et al.*, (2000). The hybridization mixture was equally dispersed

over the two slides. The slides were hybridized overnight at 60°C in a rotating hybridization oven inside a hybridization chamber and gasket (Agilent, the Netherlands). After hybridization the slides were washed according to the manufacturer's instructions (Agilent, Netherlands). The slides were stored in darkness at room temperature until scanning.

Scanning, image processing and data analysis

Microarrays were scanned after excitation of the Cy3 dye with 543 nm laser using the ScanArray® Express HT (Perkin Elmer. The microarrays were scanned at constant laser power (90 %) and 10 μm resolution settings. Tiff images were imported into the ArrayVision software (Imaging Research, The Netherlands) and the fluorescent intensity, background and signal-to-noise ratio (S/N) were determined for each spot. The background signal was defined as the average signal in the four corners surrounding each spot. The S/N was defined as the spot signal minus the background signal, divided by the standard deviation of the background signal. The values for the control spots on the array were then screened and no abnormalities were observed. A selection of spots, after exclusion of the control spots, was made based on the rule that a spot should have a signal higher than two times the background in at least 17 of the 18 arrays analysed, yielding 3541 spots for final analysis. Slides were median normalized after $\log_2$ transformation of Cy3 expression data. Normalisation was performed separately for the two slides. Afterwards, the data were combined for Principal Component Analysis (PCA) and Analysis of Variance (ANOVA). Prior to PCA, individual spots were also normalized for median ($\log_2$) gene expression, resulting in all spots having the same median gene expression. PCA was performed with the Genemaths software (applied-maths, Belgium). One way ANOVA was

performed with the R freeware (R-Development-Core-Team *R: A Language and Environment for Statistical Computing (manual)*; ISBN 3-900051-07-0; R Foundation for Statistical Computing: Vienna, 2005)

The microarray data obtained was deposited in NCBI's Gene Expression Omnibus (Edgar *et al.,* 2002) and are available via GEO Series accession number GSE15853 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15853).

Two-dimensional electrophoresis

Protein extraction and sample preparation

Total soluble protein was extracted from 1 g of lyophilized powder using a two step precipitation / extraction protocol described by Koistinen *et al*., (2002). Single extracts were prepared from each of the samples.

Two-dimensional electrophoresis

The extracted proteins were separated by two dimensional gel electrophoresis as described by Lehesranta *et al*., (2006). In the first dimension IEF 24 cm IPG strips with non-linear pH range 3-10 (Amersham Biosciences, Sweden) were used. Total protein, 150 µg, was loaded into strips and isoelectric focusing (IEF) was performed in Ettan IPGPhor IEF system (Amersham Biosciences). SDS-PAGE gels (12%, homogeneous) were used in the second dimension run using a Hoefer DALT system (Amersham Biosciences).

Proteins were detected in gels by SYPRO Ruby (Bio-Rad, USA) fluorescent protein staining which was done according to manufacturer's instructions using 10% methanol / 7% acetic acid solution. Gels were stained for 3 hours and scanned using

FLA-3000 (Fuji Photo Film, Japan) fluorescent image analyzer (excitation 470 nm, emission 580 nm). Scanned images were imported to the PDQuest version 7.1.1 software (Bio-Rad) for matching and quantification. Statistical analyses were then performed using SPSS version 14.0 software. For PCA the protein content data was log transformed and the analysis was executed using covariance matrix.

^{1}H-NMR fingerprinting

Sample preparation

Each maize powder extract was prepared by addition of 1 mL 70% methanol-d_4 (1% TMS as reference) / 30% buffer (100 mM K_2HPO_4/KH_2PO_4 in D_2O, pH 6.5) to 0.2 g of maize powder. The mixture was stirred for 30 min at room temperature and then centrifuged at 8 000 g for 10 min. The supernatant was filtered into an NMR tube and kept between 0^oC – 4^oC for more than 12 hours. Three technical replicates were made for each maize powder.

NMR spectroscopy

^{1}H-NMR spectra were recorded at 30^oC on a 400 MHz Varian Unity + spectrometer. A 5 mm 1H (13C/29Si/15N-31P) Indirect Detection PFG Probe was used. Methanol-d_4 was used as internal lock. Each spectrum consisted of 256 scans of 16468 complex data points with spectral width of 5000 Hz, an acquisition time of 1.64 sec and a recycle delay of 2 sec per scan. The pulse angle was 50^oC. The receiver gain was set at the same value for all samples within the series. A presaturation sequence was used to suppress the residual water signal with low power selective irradiation at the water frequency during the recycle delay. Spectra were Fourier transformed with 1 Hz line broadening phased and baseline corrected using the Varian

software. Spectra were converted to ASCII files. Spectra were further transferred to a personal computer for data analysis.

Data analysis and metabolite identification

The ^{1}H-NMR spectra saved in ASCII were read into GenStat**,** 10th Edition (VSN International Ltd., UK). Data exploration was performed by generating spectra of individual samples and mean spectrum of the experiment to check resolution and potential phase and/or frequency shift. Regions of the spectra were then removed if they contained background noise, water and methanol resonances. The reduced spectrum was normalized by sample vector-unit normalisation and was subjected to further statistical analysis. The variation in the ^{1}H-NMR data set was first determined by performing PCA on the sample variance-covariance matrix. This multivariate technique allows for the reduction of the data into a smaller number of components while still maintaining most of the variation in the data. Where differences in the quantities of metabolites were either observed or suggested ANOVA was employed. The metabolites that were significantly different (P <0.01) from one another were then identified using an NMR database (http://riodb01.ibase.aist.go.jp/sdbs/agi-bin/direct_frame_top.cgi).

GC/MS-based metabolite profiling

Metabolite extraction and sample preparation

Extraction and fractionation of freeze-dried maize flour were performed as described previously (Röhlig *et al*., 2009). Lipids and polar compounds were consecutively extracted from the flour. After transesterification the lipid extract was separated by solid phase extraction into a fraction containing fatty acid methyl esters

(FAME) and hydrocarbons (fraction I) and a fraction containing minor lipids, e.g. sterols and free fatty acids (fraction II). Selective hydrolysis of silylated derivatives was applied to separate the polar extract into a fraction containing silylated sugars and sugar alcohols (fraction III) and a fraction containing organic acids, amino acids and amines (fraction IV). The four fractions obtained were analyzed by gas chromatography coupled to mass spectrometry (GC/MS). Fractions II and IV were silylated before GC analysis. The GC conditions were as described previously (Röhlig *et al*., 2009).

Internal standards were tetracosane, 5α-cholestan-3ß-ol, phenyl-ß-D-glucopyranoside and p-chloro-L-phenylalanine) and retention time standards were hydrocarbons C11, C16, C24, C30 and C38.

Metabolite identification

Maize constituents were identified by comparing retention times and mass spectra with those for reference compounds and by comparing mass spectra with the entries of the mass spectra libraries NIST02 and the Golm metabolome database (http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.html).

Data analysis

GC/MS data were acquired and integrated using XCalibur 1.4 (Thermo Electron, Italy). Total ion current peak heights and retention times were exported to Chrom*pare* 1.1 (http://www.chrompare.com) for standardization and consolidation of the data. Mean values from triplicate analysis were subjected to further statistical analysis. PCA was performed using Systat 11 (Systat Software Inc., CA). Metabolite profiling data were autoscaled by the standard deviation of each analyte (correlation

matrix) to reduce the influence of metabolites with high abundance. ANOVA was performed by XLSTAT 7.5.2 (Addinsoft, France) using untransformed data.

Statistical Analysis

For PCA and one-way ANOVA the results from the three technical replicates were averaged for 2005. One-way ANOVA was performed in combination with Tukey's HSD (honestly significant difference) test (Tukey, 1953) to identify differences in the expression signals of a given transcript, the protein content and the metabolic compounds using ^{1}H-NMR fingerprinting and GC/MS-based metabolite profiling. Differences at the level $P < 0.01$ were considered statistically significant.

ACKNOWLEDGEMENTS

REFERENCES

Batista, R., Saibo, N., Lourenco, T. and Oliveira, M.M. (2008) Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. *Proc. Natl. Acad. Sci. USA,* **105**, 3640-3645.

Baudo, M.M., Lyons, R., Powers, S., Pastori, G.M., Edwards, K.J., Holdsworth, M.J. and Shewry, P.R. (2006) Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding. *Plant Biotechnol. J.* **4**, 369-80.

Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S. and Sumner, L.W. (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci*. **9**, 418–425.

Boeuf, S., Klingenspor, M., Van Hal, N.L., Schneider, T., Keijer, J. and Klaus, S. (2001) Differential gene expression in white and brown preadipocytes. *Physiol. Genomics* **7**, 15-25.

Cellini, F., Chesson, A., Colquhoun, I., Constable, A., Davies, H.V., Engel, K.-H., Gatehouse, A.M.R., Kärenlampi, S., Kok, E.J., Leguay, J.-J., Lehesranta, S., Noteborn, H.P.J.M., Pedersen, A.J. and Smith, M. (2004) Unintended effects and their detection in genetically modified crops. *Food Chem. Tox.* **42**, 1089-1125.

Chang, S., Puryear, J. and Cairney, J. (1993) A simple and efficient method for isolating RNA from pine trees. *Plant Mol. Biol. Rep.* **11**, 113–116.

Cheng, K.C., Beaulieu, J., Iquira, E., Belzile, F.J., Fortin, M.G. and Strömvik, M.V. (2008) Effect of transgenes on global gene expression in soybean is within the natural range of variation of conventional cultivars. *J. Agric. Food Chem.* **56**, 3057-3067.

Edgar, R., Domrachev, M. and Lash, A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* **30**, 207-210.

Fan, T.W.-M., Higashi, R.M. and Lane, A.N. (1988) An *in vivo* ^{1}H and ^{31}P NMR investigation of the effect of nitrate on hypoxic metabolism in maize roots. *Arch. Biochem. Biophys.* **266**, 592–606.

FAO/WHO (Food and Agriculture Organization/World Health Organization) (1996) Biotechnology and food safety. Report of a Joint FAO/ WHO consultation, Rome, Italy. FAO Food and Nutrition Paper 61, Food and Agriculture Organisation of the United Nations, Rome,

ftp://ftp.fao.org/es/esn/food/biotechnology.pdf.

FAO/WHO (2000a) Report of the First Session of the Codex *ad hoc* Intergovernmental Task Force on Foods Derived from Biotechnology (ALINORM

01/34). Rome: Food and Agriculture Organization of the United Nations, http://www.fao.org/es/esn/gm/biotc-e.htm.

FAO/WHO (2000b) Safety Aspects of Genetically Modified Foods of Plant Origin. Report of a Joint FAO/WHO Expert Consultation on Foods Derived from Biotechnology, Geneva, Switzerland, 29 May - 2 June 2000. Rome: Food and Agriculture Organization of the United Nations,
 http://www.fao.org/es/esn/gm/biotec-e.htm.

Frank, T., Meuleye, B.S., Miller, A., Shu, Q.-Y. and Engel, K.-H. (2007) Metabolite profiling of two low phytic acid (lpa) rice mutants. *J. Agric. Food Chem.* **55**, 11011-11019.

Frank, T., Nörenberg, S. and Engel, K.-H. (2009) Metabolite Profiling of Two Novel Low Phytic Acid (lpa) Soybean Mutants. *J. Agric. Food Chem.* **57**, 6408-6416.

Franssen-van Hal, N.L., Vorst, O., Kramer, E., Hall, R.D. and Keijer, J. (2002) Factors influencing cDNA microarray hybridization on silylated glass slides. *Anal. Biochem.* **308**, 5-17.

Gardiner, J.M., Buell, C.R., Elumalai, R., Galbraith, D.W., Henderson, D.A., Iniguez, A.L., Kaeppler, S.M., Kim, J.J., Liu, J., Zheng, L. and Chandler, V.L. (2005) Design, production, and utilization of long oligonucleotide microarrays for expression analysis in maize. *Maydica* **50**, 425-435.

Hazebroek, J., Harp, T., Shi, J. and Wang, H. (2007) Metabolomic analysis of low phytic acid maize kernels. In: *Concepts in Plant Metabolomics* (Nikolau, B.J. and Wurtele, E.S., eds), pp 221-238. Netherlands: Springer.

Hedge, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N. and Quackenbush, J. (2000) A concise guide to cDNA microarray analysis. *BioTechniques* **29**, 548-562.

Kleter, G.A. and Kuiper, H.A. (2002) Considerations for the assessment of the safety of genetically modified animals used for human food or animal feed. *Livest. Prod. Sci.* **74**, 275-285.

Kok, E.J. and Kuiper, H.A. (2003) Comparative safety assessment for biotech crops. *TIB* **21**, 439-444.

Kok, E.J., Franssen-van Hal, N.L.W., Winnubst, E.N.W., Kramer, E.H.M., Dijksma, W.T.P., Kuiper, H.A. and Keijer, J. (2007) Assessment of representational difference analysis (RDA) to construct informative cDNA microarrays for gene expression analysis of species with limited transcriptome information, using red and green tomatoes as a model. *J. Plant Physiol.* **164**, 337-349.

Koistinen, K.M., Hassinen, V.H., Gynther, P.A.M., Lehesranta, S.J., Keinänen, S.I., Kokko, H.I., Oksanen, E.J., Tervahuta, A.I., Auriola, S. and Kärenlmapi, S.O. (2002) Birch PR-10c is induced by factors causing oxidative stress but appears not to confer tolerance to these agents. *New Phytol.* **155**, 381-391.

Krishnan, P., Kruger, N.J. and Ratcliffe, R.G. (2004) Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **56**, 255-265.

Kuiper, H.A., Kleter, G.A., Notenorn, H.O.J.M. and Kok, E.J. (2001) Assessment of the food safety issues related to genetically modified foods. *Plant J.* **27**, 503-528.

Le Gall, G., Colquhoun, I.J., Davis, A.L., Collins, G.J.and Verhoeyen, M.E. (2003) Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J. Agric. Food Chem.* **51**, 2447-2456.

Le Gall, G., Colquhoun, I.J. and Defernez, M. (2004) Metabolite profiling using ^{1}H NMR spectroscopy for quality assessment of green tea, *Camellia sinensis* (L.). *J. Agric. Food Chem.* **52**, 692-700.

Lehesranta, S.J., Davies, H.V., Sheperd, L.V.T., Nunan, N., McNicol, J., Auriola, S., Koistinen, K.M., Suomalainen, S., Kokko, H.I. and Kärenlampi, S. (2005) Comparison of tuber proteomes of potato varieties, landraces, and genetically modified lines. *Plant Physiol.* **138**, 1690-1699.

Lehesranta, S.J., Davies, H.V., Shepherd, L.V.T., Koistinen, K.M., Massat, N., Nunan, N., McNicol, J.W. and Kärenlampi, S.O. (2006) Proteomic analysis of the potato tuber life cycle. *Proteomics* **6**, 6042-6052.

Lehesranta, S.J., Koistinen, K.M., Massat, N., Davies, H.V., Sheperd, L.V.T., McNicol, J., Cakmak, I., Cooper, J., Lück, L., Kärenlampi, S. and Leifert, C. (2007) Effects of agricultural production systems and their components on protein profiles of potato tubers. *Proteomics* **7**, 507-604.

Metzdorff, S.B., Kok, E.J., Knuthsen, O. and Pedersen, J. (2006) Evaluation of a non-targeted 'omic' approach in the safety assessment of genetically modified plants. *Plant Biol.* **8**, 662-672.

OECD (Organisation for Economic Cooperation and Development) (1993) Safety evaluation of foods derived by modern biotechnology: concept and principles,
http://www.oecd.org/dataoecd/57/3/1946129.pdf.

OECD (2002) Consensus Document on Compositional Considerations for New Varieties of Maize (*Zea mays*): Key Food and Feed Nutrients, Anti-Nutrients and Secondary Plant Metabolites,
http://www.dis.oecd.org/ollis.2002doc.nsf.

OECD (2006) An introduction to the food/feed safety consensus documents of the Task Force. Series on the Safety of Novel Foods and Feeds
http://appli1.oecd.org/olis/2006doc.nsf/43bb6130e5e86e5fc12569fa005d004c/f6714ef
c9beea840c1257156005638c2/$FILE/JT03207766.PDF.

Pastorello, E.A., Farioli, L., Pravettoni, V., Ispano, M., Scibola, E., Trambaioli, C., Giuffrida, M.G., Ansaloni, R., Godovac-Zimmermann, J., Conti, A.,

Fortunato, D. and Ortolani, C. (2000) The maize major allergen, which is responsible for food-induced allergic reactions, is a lipid transfer protein. *J. Allergy Clin. Immunol.* **106**, 744-751.

Röhlig, R.M., Eder, J. and Engel, K.-H. (2009) Metabolite profiling of maize grain: differentiation due to genetics and environment. *Metabolomics* DOI 10.1007/s11306-009-0171-5.

Sobolev, A.P., Segre, A. and Lamanna, R. (2003) Proton high-field NMR study of tomato juice. *Magn. Reson. Chem.* **41**, 237–245.

Tukey, J.W. (1953) The problem of multiple comparisons. Unpublished manuscript. In: *The Collected Works of John Tukey VIII. Multiple comparisons,* pp 1-300. New York: Chapman and Hall.

USDA (2009) The foreign agricultural service report for current world production, market and trade,
http://www.fas.usda.gov/psdonline/psdreport.

Zolla, L., Rinalducci, S., Antonioli, P. and Righetti, P.G. (2008) Proteomics as a complementary tool for identifying unintended side effects occurring in transgenic maize seeds as a result of genetic modifications. *J. Proteome Res.* **7**, 1850-1861.

Zörb, C., Langenkämper, G., Betsche, T., Niehaus, K. and Barsch, A. (2006) Metabolite profiling of wheat grains (*Triticum aestivum* L.) from organic and conventional agriculture. *J. Agric. Food Chem*. **54**, 8301-8306

TABLES

Table 1 Mean levels of gene expression obtained by microarray profiling present in maize grown in Petit and *P* values from one-way ANOVA for the factor year

Spot ID	Closest homology {species}	Log$_2$ expression values[a]			P value
		2004	2005	2006	
MOA16866	Unknown protein {*Arabidopsis thaliana*;}	-0.27 a	0.01 b	-0.67 c	<0.01
MOB01558	Unknown protein {*Oryza sativa* (japonica cultivar-group);}	-0.42 a	-0.25 a	-0.70 b	<0.01
MOA11457	Ran binding protein-1 {Lycopersicon esculentum;}	0.12 a	0.01 a	-0.53 b	<0.01
MOA27326	Putative cytochrome c oxidase subunit VIa precursor {*Oryza sativa* (japonica cultivar-group);}	-0.31 a	-0.50 a	-1.02 b	<0.01
MOA27426	NA[b]	-0.07 a	-0.43 b	-0.69 b	<0.01
MOB05509	Unknown protein {Arabidopsis thaliana;}	0.81 a	0.51 b	0.41 b	<0.01
MOA19956	Putative ribosomal protein L26 {*Oryza sativa* (japonica cultivar-group);}	-0.15 a	0.30 b	-0.54 a	<0.01
MOA20829	Acidic ribosomal protein P2a-2 {Zea mays;}	1.87 a	1.33 b	0.89 b	<0.01
MOB17522	OSJNBa0032F06.20 {*Oryza sativa* (japonica cultivar-group);}	-0.12 a	-0.41 b	-0.02 a	<0.01
MOA06880	Unknown protein {*Oryza sativa* (japonica cultivar-group);}	-0.30 a	-0.24 a	-0.84 b	<0.01

[a] Different letters on rows indicate statistically significant difference (Tukey HSD, *P* < 0.01).

[b] NA, Not annotated.

Table 2 Mean levels (three years) of gene expression obtained by microarray profiling, present in maize grown in Petit and *P* values from one-way ANOVA for the factor genotype

| Spot ID | Closest homology {Species} | Log$_2$ expression values[a] | | | *P* value |
		non-GM	GM Bt	GM RR	
MOA18226	*Glyceraldehyde 3-phosphate dehydrogenase*, cytosolic 1(GAPDH) (EC 1.2.1.12). {*Zea mays*;}	1.69 c	1.32 b	0.99 a	<0.01
MOB09726	Hypothetical protein {*Arabidopsis thaliana*;}	-0.95 a	-0.76 a	-0.43 b	<0.01
MOA21941	Nonspecific lipid-transfer protein precursor (LTP) (Phospholipidtransfer protein) (PLTP) (Allergen Zea m 14). {Zea mays;}	3.45 b	2.33 a	3.01 ab	<0.01
MOA16533	Nonspecific lipid-transfer protein precursor (LTP) (Phospholipidtransfer protein) (PLTP) (Allergen Zea m 14). {*Zea mays*;}	1.30 b	0.46 a	0.98 ab	<0.01
MOB22216	NA[b]	-0.92 a	-0.99 a	-0.57 b	<0.01
MOA11125	Nonspecific lipid-transfer protein precursor (LTP) (Phospholipidtransfer protein) (PLTP) (Allergen Zea m 14). {*Zea mays*;}	3.28 b	2.25 a	3.04 b	<0.01
MOB24171	Putative *serine carboxypeptidase*, PF00450 {*Oryza sativa* (japonica cultivar-group);}	-0.20 a	-0.21 a	-0.91 b	<0.01
MOA13700	Unknown protein {*Oryza sativa* (japonica cultivar-group);}	2.06 b	1.90 ab	1.69 a	<0.01
MOB25949	NA	0.92 ab	1.11 b	0.52 a	<0.01
MOB18146	hypothetical protein At2g34690 [imported] - *Arabidopsis thaliana* {*Arabidopsis thaliana*;}	-0.52 ab	-0.85 a	-0.36 b	<0.01

[a] Different letters on rows indicate statistically significant difference (Tukey HSD, *P* < 0.01).

[b] NA, Not annotated.

Table 3 Mean levels of protein expression obtained by protein 2D gel profiling, present in maize grown in Petit and *P* values from one-way ANOVA for the factor year

Spot ID	Spot intensity			*P* value
	2004[a]	2005	2006	
1426	49 ± 18 a	148 ± 20 b	71 ± 25 a	<0.01
2106	873 ± 67 b	522 ± 132 a	364 ± 36 a	<0.01
2609	403 ± 69 a	597 ± 49 ab	617 ± 34 b	<0.01
7501	153 ± 8 ab	195 ± 21 b	125 ± 12 a	<0.01
7503	375 ± 40 b	249 ± 18 a	292 ± 27 ab	<0.01

[a],Average ± standard deviation, different letters on rows indicate statistically significant difference (Tukey HSD, $P < 0.01$).

Table 4 Mean levels (three years) of protein expression obtained by protein 2D gel profiling, present in maize grown in Petit and *P* values from one-way ANOVA for the factor genotype

Spot ID	Spot intensity			*P* value
	non-GM[a]	GM Bt	GM RR	
4511	175 ± 9 a	193 ± 25 ab	258 ± 19 b	<0.01
5310	868 ± 21 a	1157 ± 109 b	1010 ± 51ab	<0.01
6114	2042 ± 225 b	1837 ± 464 b	296 ± 172 a	<0.01
6614	6450 ± 907 ab	4282 ± 69 a	6660 ± 515 b	<0.01

[a] Average $\pm$ standard deviation, different letters on rows indicate statistically significant difference (Tukey HSD, $P < 0.01$).

Table 5 Mean levels of metabolites obtained by ^{1}H-NMR-based metabolite fingerprinting present in maize grown in Petit and *P* values from one-way ANOVA for the factor year

NMR spectral region (ppm)	Assignment	Relative level[a]			*P* value
		2004	2005	2006	
3.20, 3.46, 3.68-3.90	Glucose	2.391 ab	2.624 b	2.036 a	<0.01
3.68, 3.71, 3.81, 3.93, 4.04	Fructose	1.087 ab	1.328 b	0.813 a	<0.01
3.51, 3.70-3.87, 3.68, 4.17, 5.42	Sucrose	4.312 ab	5.214 b	4.504 a	<0.01
3.54, 3.62, 3.72	Glycerol	0.923 ab	1.126 b	0.716 a	<0.01
3.23, 3.44, 3.64, 4.03	Inositol	0.482 a	0.733 b	0.532 ab	<0.01
3.67, 3.79	Adonitol	1.547 ab	1.666 b	1.451 a	<0.01
2.05, 2.18, 2.44, 3.66	L-glutamine	0.326 b	0.693 a	0.331 b	<0.01
3.58, 3.72, 4.66, 5.24, 8.12	Adenosine	0.830 a	1.414 b	1.069 ab	<0.01
3.64, 3.19, 4.13, 4.45, 7.41	Guanosine	1.700 ab	2.535 b	1.585 a	<0.01
7.06, 7.2, 7.12, 7.16	Tyrosine	0.112 b	0.060 ab	0.041 a	<0.01
3.02, 3.35, 3.53, 7.07, 7.25	L-tryptophan	0.014 b	nd[b] a	nd a	<0.01
1.39, 1.71, 2.05, 3.65, 4.53	Tricycol (3.3.1.1(3,7)) decan-2-ol	0.516 b	0.572 a	0.517 b	<0.05
0.84, 1.11-1.30, 1.67, 1.7	Decylcyclohexane	0.763 b	1.313 a	0.787 b	<0.01
3.13, 3.16, 3.91, 4.80, 5.20	D-(+)-raffinose	1.922 b	2.232 a	1.873 b	<0.01
1.01, 2.25, 3.47	Valine	0.165 b	0.086 a	0.136 ab	<0.01

[a] Peak heights relative to that of internal standard, different letters on rows indicate statistically significant difference (Tukey HSD, *P* < 0.01).

[b] nd, Not detected, i.e. level below limit of detection.

Table 6 Mean levels (three years) of metabolites obtained by ^{1}H-NMR-based metabolite fingerprinting, present in maize grown in Petit and P values from one-way ANOVA for the factor genotype

Compound	Relative level[a]			P value
	non-GM	GM Bt	GM RR	
Glucose	1.837 a	25.443 b	1.672 a	<0.01
Fructose	2.810 a	19.477 b	2.274 a	<0.01
Sucrose	0.723 a	1.282 b	1.047 ab	<0.01
Glycerol	0.864 a	1.134 b	0.918 ab	<0.01
Inositol	1.044 ab	0.975 a	1.100 b	<0.05
Adonitol	0.793 b	0.693 a	0.737 ab	<0.01
L-glutamine	0.732 b	0.937 a	0.788 b	<0.01
Adenosine	0.273 a	0.363 ab	0.430 b	<0.01
Guanosine	0.831 b	0.697 a	0.732 ab	<0.01
Methionine	0.076 a	0.078 a	0.092 b	<0.01
Tyrosine	0.279 b	0.283 b	0.180 a	<0.01
L-tryptophan	0.007 a	0.007 a	0.012 b	<0.01
Tricyclo (3.3.1.1(3.7)) decan-2-ol	0.032 a	0.053 b	0.049 ab	<0.01
D-(+)-raffinose	0.831 ab	0.975 b	0.732 a	<0.01
Valine	0 b	0.004 a	0 b	<0.01

[a] Peak heights relative to that of internal standard, different letters on rows indicate statistically significant difference (Tukey HSD, $P < 0.01$).

Table 7 Mean levels of metabolites obtained by GC/MS-based metabolite profiling present in maize grown in Petit and *P* values from one-way ANOVA for the factor year

Compound	Relative level[a]			*P* value
	2004	**2005**	**2006**	
15:1 ME[b]	0.025 b	0.012 a	0.023 ab	< 0.05
16:1 ME	0.138 b	0.130 a	0.133 ab	< 0.05
16:0 ME	23.460 b	20.218 a	20.903 ab	< 0.05
20:2 ME	0.064 b	0.054 a	0.060 ab	< 0.05
Unknown[c]	0.029 ab	0.025 a	0.031 b	< 0.05
24:0 TMS[d]	0.026 a	0.036 a	0.038 a	< 0.05
Cholesterol	0.009 ab	0.010 b	0.002 a	< 0.05
Gramisterol	0.016 a	0.022 ab	0.029 b	< 0.05
24-methylen-cycloartanol	nd[e] a	0.011 a	0.016 a	< 0.05
Erytritol	0.042 a	0.099 b	0.042 a	< 0.01
Fructose	0.408 a	1.084 a	0.385 a	< 0.05
Unknown[f]	nd a	0.063 b	n.d. a	< 0.01
Unknown	0.094 b	0.083 b	n.d. a	< 0.01
Alanine	0.194 a	0.428 b	0.197 a	< 0.01
Ethanolamine	0.006 a	0.052 b	nd a	< 0.01
Glycine	0.091 a	0.133 b	0.083 a	< 0.05
Serine	0.069 ab	0.091 b	0.059 a	< 0.05
Threonine	0.021 ab	0.043 b	0.032 a	< 0.05
Pyroglutamic acid	0.220 b	0.160 a	0.268 b	< 0.01
GABA	0.055 a	0.463 b	0.031 a	< 0.01
Glutamic acid	0.300 b	0.104 a	0.292 b	< 0.01

[a] Peak heights relative to that of internal standard; different letters on rows indicate statistically significant difference (Tukey HSD, *P* < 0.05).

[b] ME, Metabolites detected as fatty acid methyl esters.

[c] Detected in fraction II (minor lipids).

[d] TMS, Metabolites as persilylated derivatives.

[e] nd, Not detected, i.e. level below limit of detection.

[f] Detected in fraction III (sugars, sugar alcohols).

Table 8 Mean levels (three years) of metabolites obtained by GC/MS-based metabolite profiling, present in maize grown in Petit and *P* values from one-way ANOVA for the factor genotype

Compound	Relative level[a]			*P* value
	non-GM	GM Bt	GM RR	
γ-tocopherol	0.179 b	0.148 ab	0.048 a	< 0.05
Cycloartenol	0.028 b	0.013 a	0.014 ab	< 0.05
Inositol	0.242 b	0.219 ab	0.163 a	< 0.05
Asparagine	0.575 ab	0.476 a	0.626 b	< 0.05

[a] Peak heights relative to that of internal standard; different letters on rows indicate statistically significant difference (Tukey HSD, *P* < 0.05).

Table 9 Mean levels of gene expression obtained by microarray profiling, present in maize harvested in 2004 and *P* values from a student's T-test for the factor genotype

| ID | Putative_Annotation | Log$_2$ expression values | | *P* value |
		non-GM	GM Bt	
MZ00043132	Putative P18 {Oryza sativa (japonica cultivar-group)}	0.07	-0.08	<0.01
MZ00045186	NA[a]	-0.26	0.21	<0.01
MZ00052429	S-adenosylmethionine synthetase {Oryza sativa;}	0.63	-0.04	<0.01
MZ00027213	Unnamed protein product; dbj|BAA96220.1 gene_id:MSJ1.2 similar to unknown protein {Arabidopsis thaliana}	0.36	-0.01	<0.01
MZ00039893	Isovaleryl-CoA dehydrogenase (EC 1.3.99.10) precursor, mitochondrial [imported] - Arabidopsis thaliana {Arabidopsis thaliana}	-0.84	-0.03	<0.01
MZ00030261	DNA directed RNA polymerase II polypeptide K {Arabidopsis thaliana}	0.16	-0.12	<0.01
MZ00015623	Molybdopterin synthase (CNX2) {Arabidopsis thaliana}	-0.09	0.17	<0.01
MZ00013993	Putative eukaryotic translation initiation factor 6 {Oryza sativa (japonica cultivar-group)}	0.03	-0.26	<0.01
MZ00044574	ADP-ribosylation factor {Oryza sativa (japonica cultivar-group)}	-0.26	0.10	<0.01
MZ00024053	Spermidine synthase 1 (EC 2.5.1.16) (Putrescine aminopropyltransferase1) (SPDSY 1). {Oryza sativa}	-0.26	0.22	<0.01

[a] NA, Not annotated.

FIGURE LEGENDS

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

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

Figure 3 PCA score plots of maize harvested from Petit, Potchefstroom and Lichtenburg in 2004. Separation between the locations obtained for **(a)** microarray data, **(b)** proteomics data, **(c)** ^{1}H-NMR spectra, **(d)** GC/MS metabolite profiles.

Figure 4 PCA score plots of maize harvested from Petit, Potchefstroom and Lichtenburg in 2004. Separation between the non-GM and GM varieties obtained for **(a)** microarray data, **(b)** GC/MS metabolite profiles.
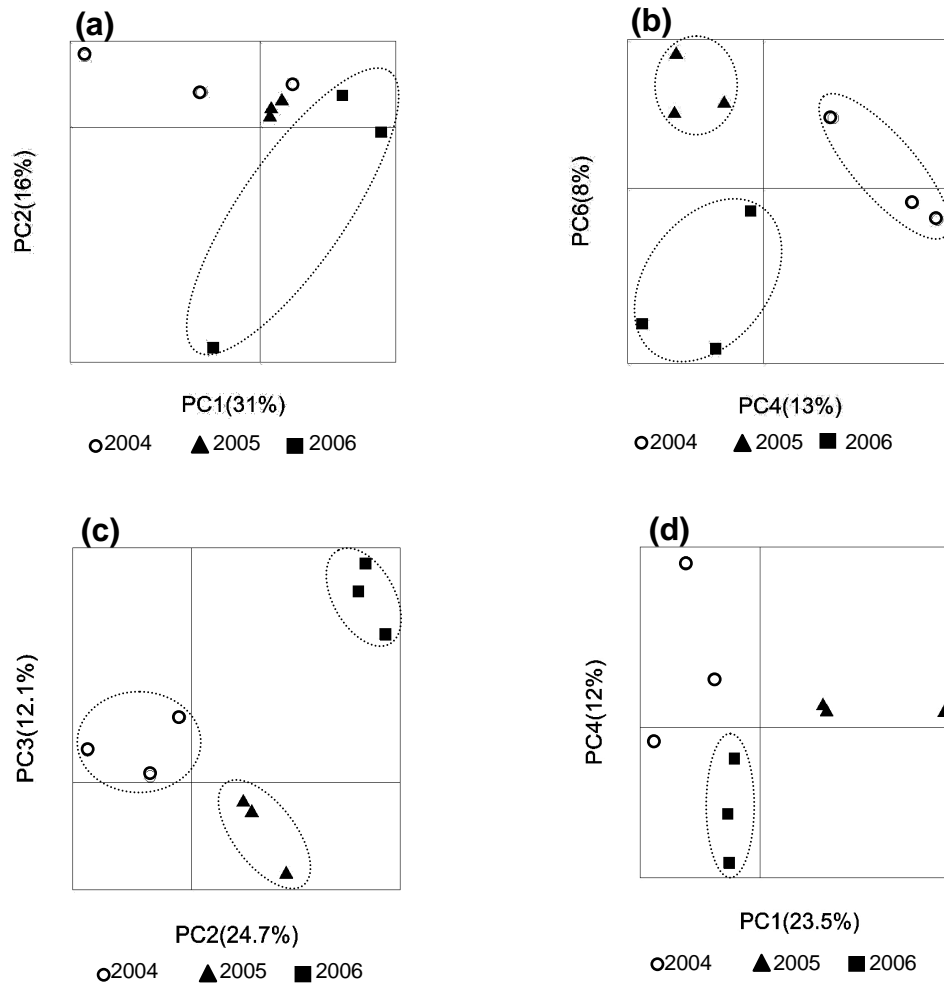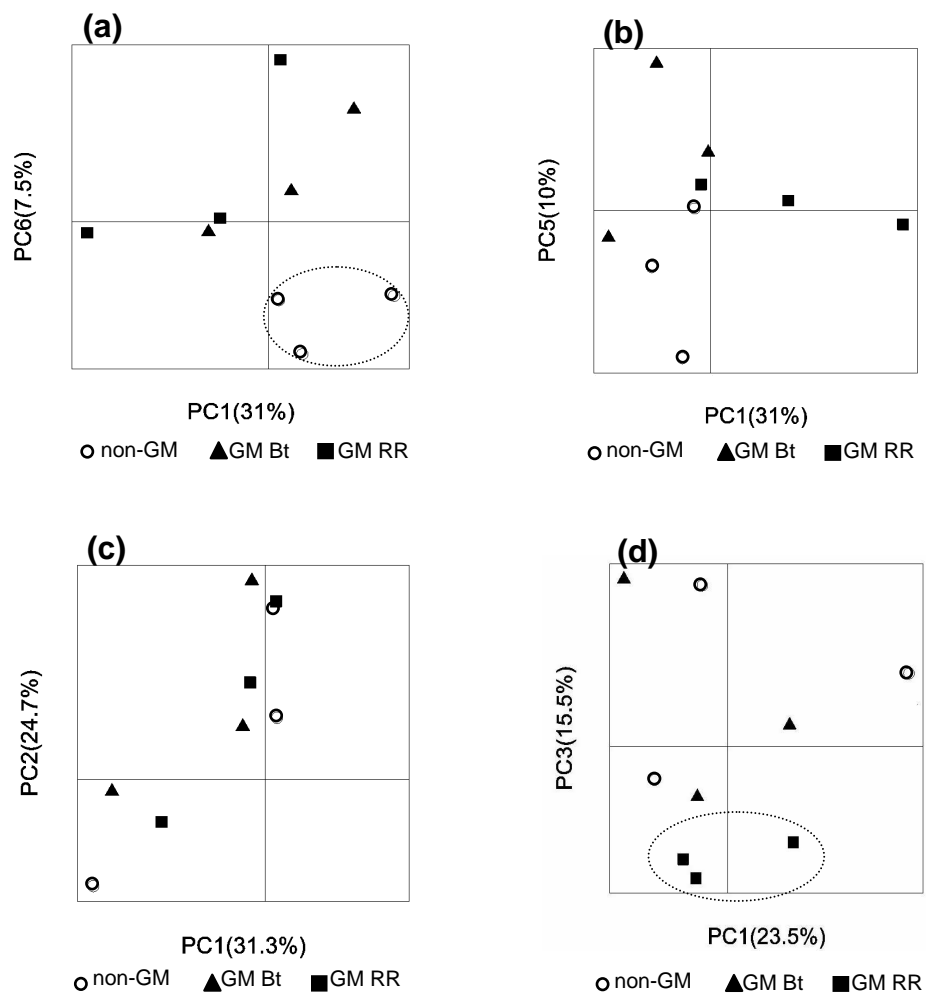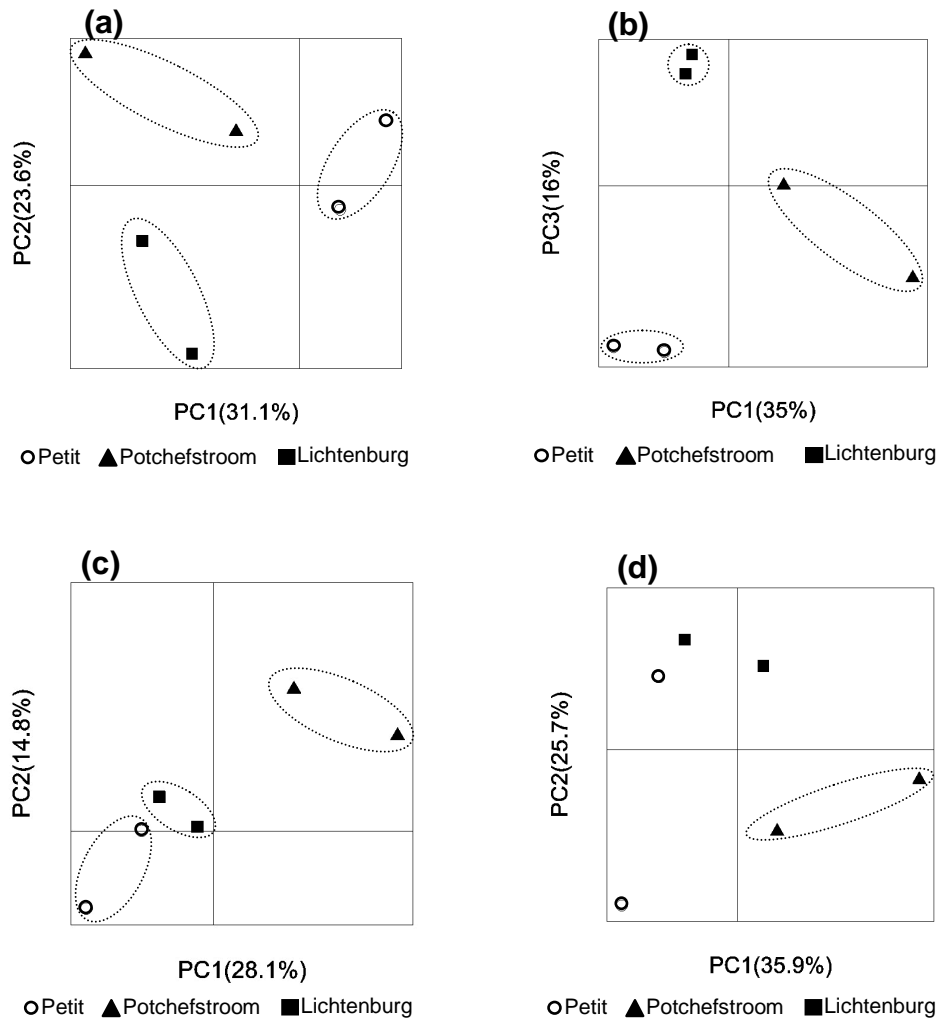
(a)

PC2(16%)

PC1(31%)

○2004 ▲2005 ■2006

(b)

PC6(8%)

PC4(13%)

○2004 ▲2005 ■2006

(c)

PC3(12.1%)

PC2(24.7%)

○2004 ▲2005 ■2006

(d)

PC4(12%)

PC1(23.5%)

○2004 ▲2005 ■2006

Figure 1

Figure 2

Figure 3

(a)

PC6(8.3%)

PC4(12.5%)

○non-GM　　▲GM Bt

(b)

PC3(18.1%)

PC1(35.9%)

○non-GM　　▲GM Bt

Figure 4