Use of omics analytical methods in the study of genetically modified maize varieties tested in 90 days feeding trials

Maria CORUJO[1], Maria PLA[2*], Jeroen van DIJK[3], Marleen VOORHUIJZEN[3], Martijn STAATS[3], Martijn SLOT[3], Arjen LOMMEN[3], Eugenia BARROS[4], Anna NADAL[2], Pere PUIGDOMÈNECH[1], José Luís LA PAZ[1], Hilko van der VOET[5], Esther KOK[3]

[1]Centre for Research in Agricultural Genomics (CRAG), Edifici CRAG, Campus UAB, 08193 Cerdanyola, Barcelona, Spain

[2]Universitat de Girona, Institute for Food and Agricultural Technology (INTEA), Campus Montilivi (EPS-1), 17003 Girona, Spain

[3]RIKILT Wageningen University & Research, Akkermaalsbos 2, Wageningen, The Netherlands

[4]Council for Scientific and Industrial Research (CSIR), Biosciences, Brummeria, Pretoria, South Africa

[5]Wageningen University & Research, Biometris, Wageningen, the Netherlands

[*]Corresponding author: M. Pla (maria.pla@udg.edu, phone: +34 972 419 852)

Maria Corujo: mariacorujobesga@gmail.com

Maria Pla: maria.pla@udg.edu

Jeroen van Dijk: jeroen.vandijk@wur.nl

Marleen Voorhuijzen: marleen.voorhuijzen@wur.nl

Martijn Staats: martijn.staats@wur.nl

Martijn Slot: martijn.m.slot@gmail.com

Arjen Lommen: arjen.lommen@wur.nl

Eugenia Barros: ebarros437@gmail.com

Anna Nadal: anna.nadal@udg.edu

Pere Puigdomènech: pere.puigdomenech@cragenomica.es

José Luís La Paz: joseluis.lapaz@cragenomica.es

Hilko van der Voet: hilko.vandervoet@wur.nl

Esther Kok: esther.Kok@wur.nl

Highlights

- The omics profiles of GM maize and several commercial non-GM varieties are compared
- GM and non-GM differences do not exceed those between non-GM commercial varieties
- Like rat feeding trials, omics results do not identify any new hazards
- The plant omics analysis approach identified grain samples of inferior quality
- Omics profiling can simplify the risk assessment procedure of new/GM plant varieties

Abstract

Genetically modified (GM) maize and their non-modified counterparts were compared using MON810 varieties, the only GMO event cultivated in Europe. The differences in grain samples were analysed by omics profiles, including transcriptomics, proteomics and metabolomics. Other cultivated maize varieties were analysed as a reference for the variability that will exist between cultivated varieties. The observed differences between modified and non-modified maize varieties do not exceed typical differences between non-modified varieties. The use of these advanced analytical approaches to analyse novel plant materials as compared to the results from animal feeding trials with whole foods is assessed. No indications were observed for changes in the GM varieties that warrant further investigations. Furthermore, it was shown that such indications will be obtained if maize samples of inferior quality are analysed similarly. Omics data provide detailed analytical information of the plant material, which facilitates a risk assessment procedure of new (GM) plant varieties.

Keywords: GMO (genetically modified organism); risk assessment; transcriptomics; proteomics; metabolomics; one-class model

1. Introduction

According to the existing legislation, before any new GM plant variety is allowed to enter the European market, it should be assessed for possible risks related to their safety for human and animal consumption and for their impact on the environment. This risk assessment focuses primarily on the intended effect of the genetic modification, i.e. on any new characteristic that has been incorporated into the GM plant variety. The new attributes have until now essentially been tolerance to a specific herbicide or resistance to one or more insects or their larvae, but it could also be an improved nutritional characteristic or the absence of an intrinsic allergenic compound that is present in the unmodified conventional counterpart. The assessment of the intended effect is usually focussed on the genetic modification produced, and can best be performed on a case-by-case basis. In practice, the assessment is largely globally harmonized and performed using the most appropriate internationally recognized and well-established guidelines (Codex Alimentarius, 2008; EFSA, 2011; Implementing Regulation (EU), 2013; OECD, 1993).

In addition to the assessment of these intended effects, an assessment of potential unintended effects is required. The procedure for testing unintended effects in new plant varieties normally includes the assessment of i) the molecular biological analysis of the locus of insertion of the construct, as well as of the flanking regions, ii) the phenotypic and agronomic aspects of the new GM variety compared to a genetically close conventional counterpart, and iii) the composition of constituents produced from the new GM variety compared to the non-GM conventional counterparts. Within the EU, the assessment for the absence of potential unintended effects derived from the genetic modification has been supplemented with the obligatory performance of a 90-day feeding study in rats with the whole food derived from the GM plant variety, including the non-GM comparator as a control group. The regulatory procedures in many other countries do not include feeding trials with whole foods without significant differences having been found in earlier analyses, as it is argued that this type of study is not sensitive enough to identify any potentially adverse effects derived from plant breeding procedures that include genetic modification that would not also show up in earlier experiments. The perceived lack of sensitivity is directly related to the fact that the whole food can only be incorporated into the animal's diet to a certain level, above which it would lead to unbalanced diets that may result in physiological effects in the animal that are unrelated to possible alterations in the plant derived from the modification and

breeding process. The situation would have been different if significant changes had been observed and reported in the GM crop plant compared to its nearest comparator, and if moreover the observed changes were considered to be of any toxicological concern. However, this situation has not yet been encountered in GMO risk assessments in Europe.

Already in 1996 it was proposed that advanced analytical methodologies might be more informative to assess potential unintended effects from plants resulting from plant breeding strategies, including genetic modification (FAO-WHO, 2000). Strategies based on advanced massive analysis of molecular data have been developed and applied to screen new plant varieties for aberrant transcriptomic, proteomic or metabolomic profiles (Ricroch, Bergé, & Kuntz, 2011). These non-targeted molecular profiling technologies were successfully used to demonstrate the sources of variation in transcript, protein and metabolite levels of two GM maize varieties compared to their non-GM counterparts that were attributed to environmental factors and to natural variation between the two different genotypes used and not to the transgenes (Balsamo, Cangahuala-Inocente, Bertoldo, Terenzi, & Arisi, 2011; Barros et al., 2010; Coll et al., 2008, 2010, 2011; Frank, Röhlig, Davies, Barros, & Engel, 2012; Ricroch, 2013). Genetic modification did not produce new proteins in addition to those related to the intended effects and did not alter the levels of endogenous metabolites or formed new metabolites and therefore no unintended effects were detected that could affect the safety of the plant materials. Coll et al. (2008, 2010) compared the transcriptomes of two GM maize varieties to those of the corresponding near-isogenic varieties and concluded that the differences could be attributed to the natural variability of the maize plants and environmental factors. Frank, Röhlig, Davies, Barros and Engel (2012) compared the metabolic profiles of two transgenic maize varieties modified with two different genes to the profiles of their respective control varieties and showed that the differences in the profiles also did not exceed those that were due to natural variability where the dominant factor driving the variability were of environmental origin. However, what has not been reported so far is the complementary evaluation using extensive omics technologies of the same GM plant materials that were used in animal feeding trials with whole foods designed to detect potential unintended effects that have their basis in the GM plant variety. In the European Union-funded project GRACE (GMO Risk Assessment and Communication of Evidence) transcriptomics, proteomics and metabolomics technologies were used for the systematic characterization of both GM and conventional maize samples, which were analysed in

parallel in animal feeding trials with whole foods following the currently-established approaches developed by the European Food Safety Authority (Zeljenková et al., 2014).

In the present article we report the results of the omics analyses of maize materials from two insect-resistant MON810 GM maize varieties that are authorised for cultivation in Europe as well as the corresponding non-GM maize counterparts. The maize varieties have specifically been grown in Spain for the GRACE project. In the present study the outcome of the omics analytical approaches were compared to the outcome of the 90-day feeding trials that have used the same maize materials in order to assess the extent to which omics profiling approaches and animal feeding trials with whole foods/feeds can be of added value to the risk assessment of GM crops beyond targeted compositional analysis. The experimental results were analysed in two ways: (i) by the direct comparison of the GM versus the non-GM materials, in line with the targeted compositional analysis that is currently part of the standard comparative compositional analysis and (ii) using the Soft Independent Modelling of Class Analogy (SIMCA) one-class model approach. In the latter approach the omics profiles of the GM maize varieties, of the conventional counterparts and of the other maize varieties considered as safe, were analysed in order to diagnose for aberrant profiles, if any, rather than focusing on individual components (van Dijk et al., 2014).

2. Materials and Methods

2.1. Maize Samples

Maize materials were the same that were used in two 90-day and a 1-year feeding trials with whole foods carried out in the frame of the European Union-funded project GRACE (GMO Risk Assessment and Communication of Evidence) and were described in (Zeljenková et al., 2014, 2016). This included two GM MON810 varieties produced by different seed companies, their corresponding non-GM near-isogenic counterparts and five additional conventional varieties (Table 1). Seeds were purchased at the Spanish local market and cultured in Foixà (Catalonia, Spain, 42º05'N, 3ºE) in the 2012 and also (except 2 conventional varieties) in the 2013 growing seasons, according to conventional agricultural practices, with no application of insecticide. Climatic data showed differences in the pluviometry. Agronomic and health parameters were as usual in the region, with below 0.4% infestation with *Sesamia nonagrioides*

and *Ostrinia nubilalis* and no relevant fungal or viral infection in 2012 while up to 13% corn borer infestation was reached in some non-GM varieties in 2013, with fungal infection observed in up to 10% stalks. Grains were dried down to <14% humidity and batches of 35-90 kg (2012) or 500 kg (2013) were transported to Mucedola srl (Milan, Italy), coded and milled. Both after coding maize grains and after milling, 1-kg samples were taken according to the ISO24333.2009 guidelines for cereals and cereal products, distributed to GRACE partners and used to obtain RNA, protein and metabolite extracts.

2.2. Transcriptomic analysis

2.2.1. RNA extraction and Illumina sequencing – CRAG-UDG

Maize grains were frozen in liquid nitrogen and embryos were manually excised and used for RNA extraction with the Maxwell 16 LEV simplyRNA Tissue Kit (Promega, Madison, WI, USA), according to the instructions by the manufacturer. 1.5 g of embryos were ground in liquid nitrogen in a mortar with pestle and then suspended in 5 mL of homogenisation solution. After centrifugation (13,000 rpm, 5 min, 4ºC) 200 μL were treated with 10 μL of DNAse I solution and used for RNA extraction.

The concentration of RNA samples were measured through absorbance at 260 nm using a spectrophotometer NanoDrop 1000 (Thermo Fisher Scientific). Quality control was based on RNA Integrity Number (RIN) and ratio of ribosomal (rRNA) peaks 28s/18s, using the Agilent RNA 6000 Plant Nano Kit Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) according to manufacturer's instructions.

RNA samples with RIN values above 8 and rRNA ratios above 2 were used for RNA-Seq at Beijing Genomics Institute (BGI, Hong Kong, China) using the HiSeq 2000 Illumina platform. Two RNA extractions and two Illumina runs were performed per variety and season. Fifty-bp single-ended reads were generated with a 40M reads/run depth. Sequences will be available at the CADIMA database.

2.2.2. RNA extraction and Illumina sequencing – RIKILT-WUR

RNA from whole kernels was isolated according to van Dijk et al. (2009) (see further details in Material S1). RNA samples were measured using a Nanodrop 1000 and absorbance measurements were used to assess the purity and concentration. For integrity evaluation, 1 μg RNA was migrated on a denaturing

agarose gel (1% agarose, 1% formamide, 1x TBE) for 60 min at 80 V and stained with ethidium bromide. Gels were visualized using a GelDoc XR+sytem (Bio-Rad) and analysed using the Quantity One 1-D software (Bio-Rad).

After quality assessment samples were sent for RNA-Seq to BGI. Samples were sent meeting the manufacturer's demands and sequenced using Illumina HiSeq/TruSeq. One RNA extraction and two Illumina run was performed per variety and season.

2.2.3. Bioinformatic analysis

Row files were analysed with FASTQC software for quality control. Row data cleaning was performed with Trim Galore! to trim reads containing adaptor- or vector-derived sequences and rRNA was filtered with SortMeRNA. Cleaned reads were mapped to the *Zea mays* reference genome assembly (Zea_mays AGPv3.31) using HISAT2_v2.0.4 (Table 2a) and the number of reads mapping every gene on the different analysed samples were calculated using the HTSeq_v0.6.1 software. After annotation quality control and data normalisation (Material S1) differential expression analysis was performed using Limma; and values were sorted by B-value. This statistic is the log-odds that that gene is differentially expressed. A threshold was established at B = 1 (probability > 73%). A false discovery rate (adjusted p-value) was also calculated. Differentially expressed genes were subjected to enrichment analysis to determine the associated functions and interpret biological processes, using the AgriGO tool (Du, Zhou, Ling, Zhang, & Su, 2010).

For detection of transgene expression, the samtools, bamtofastq, and fastq_to_fasta software were consecutively used to extract the unaligned reads to the reference genome of *Zea mays*. Then, unaligned reads were blastn-ed against the *CryIA(b)* transgene sequence as a single-sequence database.

2.2.4. Statistical analysis using a one-class model

The one-class classification tool and its use to identify aberrant compositional profiles of a large set of potato varieties in a risk assessment procedure are detailed in (Kok et al. accompanying article). Briefly, multivariate analysis is used to calculate for each sample a statistical value representing the distance to the centre of the one class model depicting the safe varieties. A 95% confidence level is used to classify

a profile as being inside or outside the single class. The multivariate model needs to be calibrated by deciding on the dimensionality (number of principal components). Cross-validation is a common way to do this, where in this case all samples of a single variety are left out of the multivariate model, and the lowest number of components is chosen such that the left-out samples are all classified within the model. Further, the prediction quality of the calibrated model cannot be taken for granted and should be evaluated using external non-GM (safe) samples. For this another layer of cross-validation is used, again leaving out all samples from a single non-GM variety. Thus, the classifier is built using a set of samples considered as safe; then refined with a second set of 'safe' samples and finally tested with a different 'safe' test sample. For the GRACE studies all varieties used to build and test the one-class model were commercial varieties that were on the market and thus considered as safe. They are listed in Table 1. In this study there were 3 GM and 17 non-GM classes. All conventional profiles were repeatedly divided into the three described sets of samples; for every combination a classifier (submodel) was defined and tested. Every submodel was subsequently used to classify the different GM samples and, for every GM variety, the results were integrated. The outcome of the classification for each variety was either inside or outside the class of commercial varieties that are considered as safe.

2.2.5. cDNA synthesis and RT-qPCR analysis

Complementary DNA (cDNA) was synthesized starting from 100 ng of total RNA, using 50 pmol Oligo-$d(T)_{20}$ primer and 200 U SuperScript-IV Reverse Transcriptase (Themofisher Scientific, Wilmington, DE, USA), according to the instructions of the manufacturer. Specific qPCR reactions were carried out in a 50-µL final volume using SYBR Premix Ex Taq (Takara Bio Inc, Shiga, Japan) with 200 nM specific primers (Table S2) and 1 µL of cDNA. PCR parameters were 10 min at 95ºC for enzyme activation; 45 cycles of 10 s at 95ºC, 30 s at 60ºC; 30 s at 72ºC; and a melting curve program (2 s at 95ºC, 15 s at 65ºC and a 19-s ramp to 95ºC). Maize ubiquitin was used as endogenous control. Non-template and RT-negative controls were systematically included to test for DNA contamination. All reactions were run in duplicate. Quantification of target mRNA was performed using the ΔΔCt method. The efficiency and linearity of the reactions were $E>0.9$ and $R^2>0.99$, as determined using serial dilutions of the corresponding amplicons.

2.3. Proteomic analysis

Protein extracts were prepared from milled grain samples following a protocol based on trichloroacetic acid (TCA) / acetone precipitation (Material S2), with two replicates per variety and season. Every protein extract was analysed in two 2-D gels (that is, 4 gels per variety and season).

2.3.1 Two dimensional electrophoresis (2D IEF SDS-PAGE)

Protein isoelectric focusing (IEF) was performed using the IPGphore system (Amersham Biosciences, Uppsala, Sweden). In a first dimension, 150 μg of protein extract were loaded onto 18 cm strips (Immobiline DryStrip pH 4-7, GE-Healthcare, Little Chalfont, UK) at room temperature. After active rehydration (50 V for 10 h) proteins were focused (500 V for 90 min, 1000 V for 90 min, 2000V for 90 min, 4000 V for 90 min, 8000V to a total of 60,000 KVh) and the strips were kept at -20ºC for >1 h. Prior to SDS-PAGE they were successively incubated for 15 min in equilibration buffer (EB: 50 mM Tris-HCl pH 8, 6 M urea, 30% glycerol; 2% SDS and 0.002% Bromophenol Blue) supplemented with 10 mg/ml dithiothreitol (DTT) and EB supplemented with 25 mg/ml iodoacetamide. They were loaded onto 12% polyacrylamide gels and run at 16ºC at 2.5 W per gel for 30 min, and then at 15 W per gel until the dye reached the end of the gel. Gels were fixed overnight in 40% ethanol/10% acetic acid and silver stained.

2.3.2. Image analysis and statistics

2D gels were scanned using an UMAX Image Scanner (Amersham Biosciences, Uppsala, Sweden) and spots were analysed using the Ludesi Redfin_3 software (Maldö, Sweden, https://ludesi.wordpress.com/). After automatic spot detection and matching, manual edition allowed correcting unmatched and mismatched spots. Spot volumes were normalized and used to compare the different samples with One-way ANOVA and Tukey test (with 0.01 significance). The profiles of every GM near-isogenic variety pair were specifically compared using t-test. Statistical analyses and graphic design were performed with the R software (R Core team, 2016).

2.3.3. Liquid chromatography-mass spectrometry (LC-MSMS)

Relevant spots were individually cut out of the gels for LC-MSMS-based protein identification at the Barcelona Parc Científic (Spain). Briefly, excised spots were trypsin-digested, washed, reduced and alkylated, extracted from the gel matrix with 10% formic acid and acetonitrile and finally analysed in a nanoAcquity liquid chromatographer (Waters) coupled to a LTQ-Orbitrap Velos (Thermo Scientific)

mass spectrometer. The Thermo Proteome Discover software and the Mascot search engine were used to search for peptide identity against a plant Uniprot SwissProt-TrEMBL. Proteins showing at least 2 high-confidence peptides (FDR≤ 0.01) were included in a candidate list; and those identified in maize with maximum score and coverage were considered the best candidates.

Information on the properties of the maize candidate proteins was retrieved from the Uniprot database (Apweiler et al., 2017). The AgriGO tool (Du et al., 2010) was used to assess enrichment of GO terms, with the Fisher statistical test and the Yekutieli multi-test adjustment method (with α, 0.05). Functional classification of differentially expressed transcripts and proteins was based on GO terminology, using GORetriever and GOSlimViewer (McCarthy et al., 2006).

2.4. Metabolomic analysis

2.4.1. UHPLC-MS metabolomic analysis - CSIR

Metabolite extracts were prepared from milled grain samples according an optimized method based on the protocol described by de Vos et al. (2007) (Material S3), with one technical replicate per sample. Each sample (5 µL) was analysed on a Waters Acquity UPLC high definition MS instrument equipped with an Acquity BEH C8 column (150 mm x 2.1 mm with a particle size of 1.7µm, Waters Corporation, Milford, MA, USA). The details of the chromatographic method used are indicated in Table S3. The runtime was 44 min and the column temperature maintained at 60°C. The samples were measured in a randomized setup and after each series of 10 samples a standard sample was analysed to check the stability of the system.

Chromatographic data analysis was done using MassLynx software (Version SCN704). Statistical data analysis was done with MarkerLynx XS^{TM} software (Version SCN704, Umetrics_v2.0.0.0). The noise rejection threshold of the software was set to 100 counts to remove the excessive noise. The cut-off value was specific for the LC-MS method and was influenced by the extraction method, solvent purity, sample complexity and instrument method used.

The identification of the five metabolites that showed differential expression was based on the monoisotopic mass value using ChemSpider database (Pence & Williams, 2010) from the Royal Society of Chemistry available at http://www.chemspider.com/PropertiesSearch.aspx.

2.4.2 LC-MS metabolomic analysis – RIKILT-WUR

Extraction was performed using 75% methanol and 0.1% formic acid (Material S3), with two replicates per variety and season. For analysis, 250 µL sample was combined with 250 µL methanol 0.125% FA/water=75/25 in a filter vial (Whatman Mini-uniprep). Injection was only performed one time out of each vial. Analyses are performed using Exactive LCMS (Orbitrap), measurements are performed in a positive mode. An Acquity UPLC BEH C8 1.7 µm 2.1 x x150 mm; 186003377 (Waters) column was used at 40°C. The injection volume was 2 µL. The composition of eluents and the gradient used are depicted in Table S3.

Exactive LC-MS datasets were preprocessed and aligned using metAlign software (Lommen, 2009; Lommen & Kools, 2012). The aligned data are output as an excel-compatible spreadsheet for further statistical analysis.

3. Results

3.1. Transcriptomics

The transcriptomes of maize embryos of a total of 11 grain samples were sequenced using mRNA-seq. These included one genetically modified MON810 variety and near-isogenic variety pair grown in two seasons, 2012 and 2013, another MON810 variety from a different seed company and its near-isogenic variety pair grown in 2012, and four additional conventional varieties one of which was grown in both seasons, and three were cultivated only in 2012 (Table 1).

Table 2a summarizes the results of RNA sequencing and mapping to the maize reference genome. There were on average 44,050,457 reads of *ca.* 49 nt per experimental replicate. Quality control for raw reads showed no specific issues regarding low quality reads or GC content. There were significant numbers of overrepresented sequences in all data files, which proved to be either adapters, poly(A) tails and sequences from cloning vectors that were removed for subsequent analyses. Ribosomal RNA represented *ca.* 10% of every set of clean and trimmed reads, and it was filtered. On aligning clean reads to the *Zea mays* reference genome, the average percentage of mapped reads was found to be

78%, with values ranging from 69% to 86%. Also, on average 82% of all mapped reads aligned exactly once to the genome (25 E6 reads). This indicates good overall sequencing accuracy and low presence of contaminating DNA. Using the gene annotation of the reference genome we calculated the number of reads of the different analysed samples that were mapping every gene The average percentage of detected genes was found to be *ca.* 91% of those estimated to be expressed in the transcriptome, with values ranging from 88% to 93%, indicating an adequate sequencing depth. Alignment and count data quality control analyses are shown in Table S1. Although *ca.* 75% of genes were mapped by at least one read per million (counts per million, CPM), only *ca.* 30% genes in any sample were mapped by more than 10 CPM and *ca.* 40% genes were mapped by more than 5 CPM (Fig. S1). This may possibly reflect the nature of the analysed tissue, corresponding to mature and dry embryos. For differential expression analysis, low-count genes were filtered using a gene expression threshold that was computed basing on a comparison of the distribution of read counts in annotated gene regions to read counts observed in intergenic regions; and included genes in the 40% highest expression in at least one sample. Clustering of the completely processed data showed no separation of GM from conventional varieties (Fig. 1a), suggesting overall similarity between the GM and their corresponding near-isogenic varieties. A score plot on the first two axes of a Principal Component Analysis, PC1 and PC2, explaining 28 and 19% variability, respectively, gave the same results.

GM and near-isogenic varieties were compared, without distinguishing company or season, in the linear modelling software package Limma (Ritchie et al., 2015). Values were sorted by the log-odds that that gene is differentially expressed (B statistic). A B-statistic of zero corresponds to a 50-50 chance that the gene is differentially expressed. A filter was set at B-values above one, i.e. roughly 75% probability of differential expression. This was considered as a non-restrictive value and facilitates visualisation of differences. There were four genes with B values above one, GRMZM2G152436, GRMZM2G047097, GRMZM2G456487 and GRMZM2G098679. The two former ones had fold-changes lower than 1.5-fold while the two last ones were 4.0 and 3.5-fold down-regulated in the GM crop, respectively. GRMZM2G456487 and GRMZM2G098679 correspond to a putative WAK receptor-like protein kinase and a sugar transporter. Pairwise comparisons were then carried out to evaluate differential gene expression between every pair of GM and near-isogenic varieties grown in every season. There were 12 genes differentially expressed in DKC6667YG and DKC6666 grown in the 2012 season (Table S4).

PR33D48 and PR32T16 showed no differences in the 2012 or the 2013 seasons. Gene ontology analysis showed no statistically overrepresented terms in regulated transcripts.

As a complementary approach, classification of the maize transcriptomics profiles was performed on the basis of the SIMCA one-class model (van Dijk et al., 2014). Mature kernels of fourteen conventional commercial maize varieties, some of which cropped in two different seasons and including the conventional counterparts of the GM varieties included in the study (Table 1), with 39,787 variables per profile, were used to construct the one-class SIMCA classification model. The variables resulted from mapping the RNA-seq data of all individual samples to a maize reference genome. Figure S2 integrates variability of gene expression in the transcriptomes of these maize kernel samples. For the SIMCA classification tool, a total of 182 (= 14 x 13) submodels were constructed from fourteen cross validation samples with thirteen test samples each for the transcriptomics data from the maize kernels (see also Kok et al. accompanying article). The SIMCA one-class model was also applied to the maize embryo transcriptomics data. In this case 30 (= 6 x 5) submodels were constructed based on 6 conventional varieties, including the two parent lines (Table 2b). GM maize variety PR33D48, separately for the two cropping seasons, was classified for the maize kernel transcriptomics; and both GM maize varieties, PR33D48 and DKC6667YG, were classified for the maize embryo transcriptomics (Table 2c). For each sample, the percentage of the submodels was calculated for which the sample was classified as inside the model (score lower than the threshold). This percentage was assessed in two ways: i) as the majority classification of the submodels, i.e. if more than 50% of the submodels classifies the sample as inside the model, the sample is overall classified as inside of the model, and ii) the GM variety is classified as inside the reference class if the GM variety is classified within the model more often than the commercial varieties that were used in the combined test set (91.8% for the maize kernel transcriptomics and 85.0% for the maize embryo transcriptomics, Table 2b). The latter approach is clearly more stringent in terms of 'in' classification compared to the 'majority' classification. The repeats of the GM maize samples of PR33D48, and for both GM varieties, were all classified as inside the model regardless of the threshold applied, for the SIMCA one class model for the maize kernel transcriptomics, and the maize embryo transcriptomics, respectively (Table 2c).

3.2. Proteomics

Two-dimensional IEF and SDS-PAGE proteome profiles of a total of 16 maize grain samples were obtained. These included two GM and near-isogenic variety pairs grown in two seasons, and five additional conventional varieties from which three were grown in 2012 and 2013 and two were cultured only once (Table 1). An average of 1400 spots were clearly detected in each variety with pI values in the 4 to 7 range and Mw from 10 KDa to 245 KDa, representing the most abundant proteins in maize mature kernel, mainly seed storage proteins (an illustrating example is shown in Fig.S3).

The overall similarity between the proteomes of the different grain genotypes and growing seasons was assessed using Principal Component Analysis (PCA), taking the normalized spot volumes as variables. About 30% variability between the samples was explained within the two first components (Fig. 1b). Grain samples tended to show different PC1 values (explaining 20% variability) as a function of the corresponding growing season; and there was no visible separation of GM and near-isogenic varieties in the PCA plot. This suggested that the environmental conditions and normal non-GM genetic background had a higher impact on maize grain proteome than transgene insertion. The 2013 growing season was characterized by an unusually strong hailstorm at the onset of flowering.

Further pairwise comparisons of MON810 and near-isogenic non-GM samples were performed by direct comparison of the normalized spot volumes using the t-test, with thresholds established at 2-fold change, p value < 0.01. A total of 15 spots had different volumes in at least one GM and near-isogenic variety pair and season. Their fold-changes in all GM and near-isogenic pairwise comparisons, together with their LC-MSMS based identification, are summarized in Table 3. There was no conservation in the differential proteome pattern. DKC6667YG and DKC6666 had no differential spot in 2012 and 9 in 2013, which corresponded to the LEA (late embryogenesis abundant) group 6 D-34 protein and the storage protein Globulin-1 S allele. PR33D48 and PR32T16 had 3 and 4 differential spots in 2012 and 2013, respectively. Lactoylglutathione lyase (or glyoxalase I, EC 4.4.1.5) was commonly up-regulated in the GM variety in 2012 and 2013; whereas two nutrient reservoir proteins (Globulin-2 and Globulin-1) were regulated in 2012, and the 22.0 kDa class IV small heat shock protein (sHSP), rRNA N-glycosylase and the LEA D-34 in 2013. Proteins with storage and nutrient reservoir function are well known to accumulate to very high levels in mature seeds. The rest of differentially expressed proteins also accumulate in seeds during the last stage of maturation, when desiccation occurs (Wu et al., 2015), and have been related to the response to drought and other abiotic stress conditions (Gong, Yang, Tai, Hu, & Wang, 2014). They participate in adaptive response to dehydration and component protection mechanisms

389 (Battaglia, Olvera-Carrillo, Garciarrubio, Campos, & Covarrubias, 2008); detoxification of methylglyoxal
390 (MG, which natural levels increase significantly under drought and other abiotic stress conditions,
391 (Yadav, Singla-Pareek, Ray, Reddy, & Sopory, 2005); or defence-related functions in these stress
392 conditions (Bass et al., 2004).

393 The volumes of all 15 spots were within the range of conventional varieties analysed in this study; and
394 most often the near-isogenic variety grown in the same season had the closest confidence interval (one-
395 way ANOVA and Tukey post-test, 95% confidence interval). In addition, no spot had differential volumes
396 in the two variety pairs and seasons.

397 *3.3. Metabolomics*

398 Metabolite profiles were generated for 8 maize varieties by a non-targeted (untargeted) approach using
399 UHPLC-MS technology. The samples included two MON810 GM varieties and their respective near-
400 isogenic lines, and four additional commercial maize varieties (Table 1). Evaluation of the metabolite
401 data set (392 variables across all samples) was done using multivariate analysis carried out in
402 MarkerLynx XS^{TM} software (Umetrics Version 2.0.0.0). It included principal component analysis (PCA),
403 which is an unsupervised multivariate linear model, followed by the orthogonal projection to latent
404 structures-discriminant analysis (OPLS-DA) that is a supervised model. Principal Component Analysis
405 (PCA) shows the similar groupings of the two GM varieties with their respective near-isogenic lines and
406 with the other four maize varieties (Fig. 2a). The Hotelling's T2, a generalisation of the Student's *t*-
407 distribution applied to multivariate situations, confirmed that no samples were detected outside the 95%
408 confidence interval of the modelled variation (Fig.S4a).

409 A comprehensive evaluation of the metabolite data of the two GM varieties and their respective near
410 isogenic lines was performed using orthogonal partial least squares discriminant analysis (OPLS-DA) in
411 order to maximize the differences between these two groups. The difference between the two GM
412 varieties and their corresponding near isogenic lines shows variation between the two groups, seen in
413 the first component, t1P; variation within the groups is seen in the second orthogonal component, t2o of
414 the OPLS-DA score plot (Fig. S4b-c). OPLS-DA loadings generated an S-plot, based on retention time
415 and metabolite mass data that allows the visualisation of the metabolites responsible for the differences
416 between the samples (Fig. 2b and Fig. 2c). The metabolites distributed in the lower and upper outer

417 regions of the S distribution plot represent those metabolites that are differentially produced i.e. down
418 regulated or up regulated metabolites and are therefore responsible for the group separation. Although
419 a definite identification of metabolites from databases of metabolite masses is not possible with
420 untargeted metabolomics studies, from the OPLS-DA scores the factor of change in metabolite
421 concentrations between the GM and near-isogenic lines is shown in Table 4a. The concentrations of
422 three metabolites changed by a factor of 1.3 and 1.6 in each of the sets whereas the concentrations of
423 two metabolites (mass 496.3357 and 518.3156) changed in both sets of GM and near-isogenic lines by
424 a factor change of 1.3, 2.9 and 3.1; no new metabolites were found in any of the comparisons between
425 the two GM and near-isogenic lines. The higher changes in concentrations found in one of the genotypes
426 highlights the normal variation expected between different genetic backgrounds (exemplified by
427 backgrounds derived from the two different seed companies, Table 1). Attempts to classify the five
428 metabolites that showed differential expression from databases using only the monoisotopic mass
429 generated a list of possible compounds; however using the ChemSpider database and narrowing the
430 mass interval range to 0.001 and 0.0001 resulted in the identification of fewer candidate metabolites
431 (Table 4a).

432 Comparison of the six maize varieties with exception of the two GM varieties showed the effect of natural
433 variation that exists among them. The PCA plot showed the patterns of the metabolites spread among
434 the four quadrants representing the diversity among maize varieties (Fig. S4d).The procedure to assess
435 the metabolomics profiles of the GM maize variety in the light of similar profiles obtained from the near-
436 isogenic comparator as well as from other conventional varieties that are commercially available and
437 considered as safe, was similar to the one as described for the transcriptomics profiles (i.e. by applying
438 the SIMCA model). For the classification of the metabolomics profiles, seven conventional varieties,
439 including the conventional counterpart of the GM variety as well as biological repeats (Table 4b), were
440 used to construct the one-class SIMCA model, with 128,873 variables (metabolites) measured for each
441 individual metabolomics profile as obtained in the procedure by RIKILT Wageningen University &
442 Research (46 profiles) and 392 variables in the profiles as obtained by CSIR (36 profiles). In this way a
443 total of 42 (= 7 x 6) and 24 (=6 x 5, with 6 submodels failing as not all profiles of the inner cross-validation
444 set were classified as ‘in’) submodels, respectively, were constructed. The conventional maize varieties
445 that were used as test varieties were classified by each submodel, resulting in an overall test set
446 threshold of 64.1% and 70.8%, respectively, of (commercial) test samples that were classified as inside

of the one-class model. The assessment of the model performance is based on classification of the conventional counterpart, if available, compared to the test set classifications (Table 4b). For the model to be acceptable, the percentage for the conventional varieties should be higher than the combined percentage for the test set sample, indicating that the conventional counterpart variety is positioned in the centre of the natural variation included in the classification model. Here it was observed, however, that the near-isogenic conventional comparator DKC6666 showed a much lower percentage 'inside the model' compared to the other conventional varieties combined in the test set. This means that for this conventional comparator the model is of insufficient discriminatory power, as the parent line is insufficiently central in the resulting model, which may lead to similar profiles being too easily classified as outside of the models. Therefore this model was not further included in the assessment. Also for the other conventional comparator PR32T16 only the CSIR model met the quality criteria, the RIKILT model also being slightly below the set criteria of a higher percentage inside for the conventional comparator compared to the average of the combined test set.

Taking these limitations into account, the SIMCA model was subsequently used to classify the GM maize variety PR33D48 that has been used in the GRACE 90-days animal feeding trials (Steinberg, 2015; Zeljenková et al., 2014). In both cases this GM variety was classified as inside of the model, whether based on the majority or the test set threshold. In addition to these samples two additional maize samples were assessed that were fungal infected and considered to be of inferior feeding quality as a result of this (Table 4c). The fungal infection was assessed visually. The fungal infected samples were both classified as outside of the model (Table 4c).

3.4. Analysis of the intended effect

The transcriptome data of the GM variety PR33D48 and its conventional counterpart [PR32T16] were additionally assessed for the presence of newly expressed RNAs that were not present in the non-transgenic maize transcriptome. This assessment was performed in two steps: in the first step the transcriptomes of the GM maize variety PR33D48 and of the conventional counterpart PR32T16 were compared to the maize reference genome. For the sequences that were not recognized in this way, a de-novo assembly was performed, i.e. longer sequences were built based on similarity. On the resulting sequences a BLAST analysis was performed to find their identity and the transcript of the *cryIA(b)* gene

clearly appeared. This approach allowed for identification and confirmation of the anticipated transcripts corresponding to the transgene *cryIA(b)* sequence in the MON810 samples. Following this analysis, 1,169 and 989 transcripts were identified that could not be aligned to the reference genome in two biological repeats of the PR33D48 MON810 GM maize variety, respectively, against 1,745 in the parent line. When comparing the unaligned transcripts in the GM lines versus the parent line, 44 transcripts were identified that were present in the GM lines but not in the parent line. Most of these transcripts are short and not informative: when transcripts were selected that were > 1 kb long, and thus possibly biologically meaningful, only 5 transcripts remained. Four of these transcripts related to hypothetical proteins, one of them to a maize mRNA sequence, and the additional transcript corresponded to the *cry1A(b)* gene. In an alternative approach, embryo RNA-seq dataset was additionally assessed for the presence of the transgene sequence. The sets of unmapped reads were extracted and blasted against the *cryIA(b)* sequence. The number of reads that were found to map the transgene were on average 335, 345 and 495 for the GM varieties and seasons PR33D48_2012, PR33D48_2013. These results produce a confirmation of the presence of the transgene transcripts even in the mature maize embryo.

4. Discussion and Conclusions

An important part of the risk assessment of GM crops is generally based on the comparison of the GM plant with the nearest conventional counterpart and additional comparators that have a history of safe use (Implementing Regulation (EU), 2013). The comparison focuses on phenotypic and agronomic aspects as well as on the compositional analysis of the new variety versus the conventional counterpart. Other elements of the risk assessment procedure that relate to the identification of potential unintended effects are a detailed molecular characterization of the genetic modification, as well as, in Europe and a few other countries, on the performance of toxicological studies with the whole food derived from the GM plant variety. It has been advocated in the past that omics approaches could be more informative and more cost-efficient compared to the current targeted approach, and may in practice provide more information on potential perturbations in the physiology of plants and their possible contribution to harmful effects, compared to the obligatory animal feeding trials with whole foods.

In the present study, the insect-resistant GM maize event MON810, the only GM event presently cultivated in Europe was used as a proof of concept. The GM maize material, as well as the near-

isogenic variety were the same materials that have also been used in the corresponding 90-day animal feeding trials with whole foods in the European GRACE project (Steinberg, 2015; Zeljenková et al., 2014). In the GRACE studies, both the animal feeding trials as well as the present omics studies, two MON810 and related near isogenic genotypes were included: the GM varieties PR33D48 and DKC6667YG were pairwise compared to their conventional counterparts PR32T16 and DKC6666, respectively. Additional conventional varieties were included in the comparison, and crops were grown in two different seasons that had different meteorological conditions in order to be able to interpret observed differences in the light of the natural variation in plant composition. In this respect the omics comparison has been performed in a way that is directly comparable to the current approach for targeted analyses according to both the valid European procedure as well as procedures as proposed in widely accepted international guidelines (Codex Alimentarius, 2008; EFSA, 2011; Implementing Regulation (EU), 2013; OECD, 1993). The additional conventional varieties included in the comparisons were all commercially available and thus considered as safe. All maize varieties were agriculturally cultivated in the same zone in Spain, thus removing the effect of location and environmental influences as a source of variability.

Transcriptomic, proteomic and metabolomic analyses were carried out in parallel in two independent laboratories, using grain samples prepared from the same batches as were shipped for the preparation of the animal feeds for the animal feeding trials with whole foods (Zeljenková et al., 2014). The additional conventional maize samples were processed and assessed in the same way. Maize kernels were selected as the tissue of choice, as they were also included in the animal feeding trials and as they directly relate to food and feed products. For now, all three omics strategies were applied, whereas in future times the combination of metabolomics and proteomics may be most informative. At the moment transcriptomics still has the largest relative coverage. Direct comparison of the GM variety versus the near-isogenic variety showed limited differences, below 1%, in both cases. Differences can already be expected based on the large number of analyses and the statistical approach that considers a 95% confidence interval, resulting in 5% observed differences when a normal distribution is assumed under a H_0- hypothesis. In the present study, when the observed differences were considered in the frame of the natural variation as seen in the additional non-GM varieties, the levels of all transcripts, proteins and secondary metabolites analysed were within the range of the levels found in the conventional varieties and no indications were found for any unintended effect of the genetic modification on the physiology of

the GM maize materials. Similar conclusions were reported upon targeted nutritional and compositional assessment of MON810 kernels (BCH, 2002) according to the OECD recommendations (OECD, 2002). These include key food and feed nutrients, anti-nutrients and secondary plant metabolites, in particular proximates (protein, fat, total dietary fibre, ash, carbohydrates), amino acids, fatty acids, minerals, vitamins and phytic acid, raffinose, furfural, ferulic acid and p-coumaric acid. Complementary to the OECD recommendations, deeper analytical methods such as omics can provide untargeted data on thousands of gene expression, protein and metabolite parameters, giving access to far more information than existing requirements which increases confidence of no unintended impacts of GMOs.

To further assess the transcriptomic and metabolomic profiles, the SIMCA one class classification model was applied (van Dijk et al., 2014). This allowed the screening of new profiles for differences when comparing the new GM variety profiles to a set of profiles from conventional varieties that are considered as safe, including the profiles of the conventional comparator. In all cases it was found that the GM varieties were classified as inside the one class model. As positive controls, i.e. to represent samples of inferior quality, two samples were included that were fungus-infected. Based on the same one class model, these two maize samples were both classified outside of the model. These observations confirm that the one class model does perform in the way it has been developed, i.e. with a focus on the classification of profiles as 'out' at least in those cases where the underlying plant materials are of inferior quality. The model may lead to false positives; while this is considered acceptable in those cases additional investigations will need to be performed to understand the underlying differences that may not be related to the genetic modification as such, which is at the basis of the classification. In practice it will be relatively easy to further assess observed differences based on the available omics profiles to evaluate whether additional research may be required in those cases where a profile of a new variety is classified outside of the one class of varieties that are considered as safe.

It has furthermore been shown that transcriptomic data are useful to confirm anticipated changes in the physiology of plants related to the intended effect. The same approach, comparing the transcript dataset to the reference genome of the species of interest, may even be informative to identify any possible unintended effects of a plant breeding program resulting in newly expressed transcripts, but this will need to be further investigated. Here it has been shown that in practice, the number of transcripts that may differ between a GM plant variety and its conventional counterpart may be manageable and sufficient to provide a basis to screen transcripts for characteristics of toxicological concern.

On the basis of these combined findings it can be argued that these analytical data provide insight into relevant differences in the profiles of the GM varieties when compared to similar data from the near-isogenic comparator as well as a range of conventional commercial varieties. Furthermore the results also show that in the case of the fungus infested maize materials this approach does indicate when relevant differences are observed that warrant further investigations. Similarly, the small differences in the MON810 GM plant variety versus its nearest control, that are all within the ranges of natural variation, are not identified as differences that require a toxicological follow-up. Together these results do seem to indicate that analytical approaches are more informative compared to animal feeding trials with whole foods, where the limitations in terms of sensitivity have well been documented (Kok, Keijer, Kleter, & Kuiper, 2008; Kuiper, Kok, & Davies, 2013). In the GRACE study, the direct comparison between omics approaches and animal feeding trials with whole foods has for the first time been made based on exactly the same plant materials. This direct comparison has shown that the analytical approach allows a much broader comparison with additional conventional varieties compared to the animal feeding trials with whole foods, against a fraction of the costs of the trial. Also, it has been shown that the analytical data can provide insight into the actual changes in the plant's physiology due to the added genetic characteristic, as well as an appropriate assessment of the presence or absence of unintended changes in the metabolism of the plant in the light of the natural variation within the same species. Based on the analyses included in the omics profiles, no indications have been observed for changes in the physiology of the MON810 GM plant varieties that warrant further investigations. At the same it was shown that such indications will be obtained if maize samples of inferior quality are also included in the assessment. From the present results obtained in the GRACE project, it can be stated that omics data provide detailed analytical information of the plant material which facilitates a risk assessment procedure of new (GM) plant varieties. In particular cases, when deviations of specific parameters indicating a safety concern are observed, they may provide arguments for the need to carry out focused feeding trials with the plant-derived whole food based on clear-cut questions.

Acknowledgements

This study was carried out as part of the GRACE project ("GMO Risk Assessment and Communication of Evidence"), financially supported by the 7th Framework Programme of the European Community for

Research, Technological Development and Demonstration Activities (FP7), Grant Agreement No. 311957. We thank Dr. Paul Steenkamp (CSIR) for performing the LC-MS/MS analysis; and Dr. Martí Bernardo and Dr. Víctor M. González (CRAG) for informatics support.

Declarations of interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Material S1. Transcriptomic analysis

Material S2. Proteomic analysis

Material S3. Metabolomic analysis

Table S1. Embryo RNA-seq: alignment and count data quality control summary.

Table S2. List of primers used in RT-qPCR based validation of RNA-Seq results.

Table S3. UHPLC gradient conditions for the chromatographic analysis.

Table S4. Differentially expressed genes in MON810 and near-isogenic varieties, without distinguishing company or season (MON810 / near-isogenics) and only considering DKC6667YG and DKC6666 grown in 2012.

Fig. S1. Distribution of genes as a function of their count number.

Fig. S2. Analysis of integrated variability of gene expression in the transcriptomes of maize kernel samples**.**

Fig. S3. Example of a typical 2D IEF SDS-PAGE analysis of maize seeds.

Fig. S4. Maize metabolomics: graphical representation of multivariate data analysis of the metabolite profiles of maize varieties.

References

Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., … Yeh, L.-S. L. (2017). UniProt: the universal protein knowledgebase. Nucleic Acids Research, 45(D1), D158–D169.

Balsamo, G. M., Cangahuala-Inocente, G. C., Bertoldo, J. B., Terenzi, H., & Arisi, A. C. M. (2011). Proteomic analysis of four Brazilian MON810 maize varieties and their four non-genetically-modified isogenic varieties. *Journal of Agricultural and Food Chemistry*, *59*, 11553–11559.

Barros, E., Lezar, S., Anttonen, M. J., van Dijk, J. P., Rohlig, R. M., Kok, E. J., & Engel, K. H. (2010). Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnology Journal*, *8*, 436–451.

Bass, H. W., Krawetz, J. E., Obrian, G. R., Zinselmeier, C., Habben, J. E., & Boston, R. S. (2004). Maize ribosome-inactivating proteins (RIPs) with distinct expression patterns have similar requirements for proenzyme activation. *Journal of Experimental Botany*, *55*, 2219–2233.

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., & Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiology*, *148*, 6–24.

Biosafety Clearing House. (2002). Safety assessment of YieldGard insect-protected corn event MON810. URL http://bch.cbd.int/database/record.shtml?documentid=14750. Accessed 16.03.2018

Codex Alimentarius. (2008). Guideline for the Conduct of Food Safety Assessment of Foods Derived from Recombinant-DNA Plants. Codex alimentarius, Joint FAO/WHO Food Standards Program. URL http://www.fao.org/input/download/standards/10021/CXG_045e.pdf. Accessed 20.09.2017

Coll, A., Nadal, A., Collado, R., Capellades, G., Kubista, M., Messeguer, J., & Pla, M. (2010). Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices. *Plant Molecular Biology*, *73*, 349–362.

Coll, A., Nadal, A., Palaudelmàs, M., Messeguer, J., Melé, E., Puigdomènech, P., & Pla, M. (2008). Lack of repeatable differential expression patterns between MON810 and comparable commercial

varieties of maize. *Plant Molecular Biology*, *68*, 105–117.

Coll, A., Nadal, A., Rossignol, M., Collado, R., Puigdomènech, P., & Pla, M. (2011). Proteomic analysis of MON810 and comparable non-GM maize varieties grown in agricultural fields. *Transgenic Research*, *20*, 939–949.

De Vos, R. C., Moco, S., Lommen, A., Keurentjes, J. J., Bino, R. J., & Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nature Protocols*, *2*, 778–791.

Du, Z., Zhou, X., Ling, Y., Zhang, Z., & Su, Z. (2010). AgriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research*, *38* (Web Server), W64–W70.

EFSA. (2011). Guidance for risk assessment of food and feed from genetically modified plants. *EFSA Journal*, *9*(5):2150, 1-37.

FAO-WHO. (2000). Safety aspects of genetically modifief foods of plan origin. Report of a Joint FAO/WHO Expert Consultation on food derived from biotechnology, Geneva, Switzerland, 29 May to 2 June 2000. URL Ftp://ftp.fao.org/es/esn/food/gmreport.pdf. Accessed 20.09.2017

Frank, T., Röhlig, R. M., Davies, H. V., Barros, E., & Engel, K. H. (2012). Metabolite profiling of maize kernels-genetic modification versus environmental influence. *Journal of Agricultural and Food Chemistry*, *60*, 3005–3012.

Gong, F. P., Yang, L., Tai, F. J., Hu, X. L., & Wang, W. (2014). “Omics” of Maize Stress Response for Sustainable Food Production: Opportunities and Challenges. *OMICS: A Journal of Integrative Biology*, *18*, 714–732.

Implementing Regulation (EU). (2013). Commission Implementing Regulation (EU) No. 503/2013 on applications for authorisation of genetically modified food and feed in accordance with Regulation (EC) No. 1829/2003 of the European Parliament and of the Council and amending Commission Regulations. *Official Journal of the European Union*, *L157*, 1–48.

Kok, E. J., Keijer, J., Kleter, G. A., & Kuiper, H. A. (2008). Comparative safety assessment of plant-derived foods. *RegulatoryToxicology and Pharmacology*, *50*, 98–113.

669 Kuiper, H. A., Kok, E. J., & Davies, H. V. (2013). New EU legislation for risk assessment of GM food:
670 No scientific justification for mandatory animal feeding trials. *Plant Biotechnology Journal*, *11*, 781-
671 784.

672 Lommen, A. (2009). Metalign: Interface-driven, versatile metabolomics tool for hyphenated full-scan
673 mass spectrometry data preprocessing. *Analytical Chemistry*, *81*, 3079–3086.

674 Lommen, A., & Kools, H. J. (2012). MetAlign 3.0: Performance enhancement by efficient use of
675 advances in computer hardware. Metabolomics, 8, 719–726.

676 McCarthy, F. M., Wang, N., Magee, G. B., Nanduri, B., Lawrence, M. L., Camon, E. B., … Burgess, S.
677 C. (2006). AgBase: A functional genomics resource for agriculture. BMC Genomics, 7.

678 OECD. (1993). Safety Evaluation of Foods Derived by Modern Biotechnology. URL
679 http://www.oecd.org/dataoecd/37/18/41036698.pdf. Accessed 25.09.2017.

680 OECD. (2002). Consensus document on compositional considerations for new varieties of maize (Zea
681 mays): key food and feed nutrients, anti-nutrients and secondary plant metabolites. Series on the
682 Safety of Novel Foods and Feeds, No. 6.

683 Pence, H. E., & Williams, A. (2010). Chemspider: An online chemical information resource. *Journal of*
684 *Chemical Education,* 87, 1123–1124.

685 R Core Team. (2016). R: A Language and Environment for Statistical Computing. R Foundation for
686 Statistical Computing, Vienna, Austria. URL http://www.R-Project.org/. Accessed 05.10.2017

687 Ricroch, A. E. (2013). Assessment of GE food safety using “-omics” techniques and long-term animal
688 feeding studies. *New Biotechnology*, *30*, 349–354.

689 Ricroch, A. E., Bergé, J. B., & Kuntz, M. (2011). Evaluation of genetically engineered crops using
690 transcriptomic, proteomic, and metabolomic profiling techniques. *Plant Physiology*, *155*, 1752–
691 1761.

692 Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers
693 differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids*
694 *Research*, *43*, e47.

695 Steinberg, P. (2015). Response to a report and press release by Bauer-Panskus and Then (2014)
696 criticizing the presentation and interpretation of the results of recently published 90-day feeding
697 studies with diets containing genetically modified MON810-maize varieties and their comparators
698 (Zeljenková et al. 2014). *Archives of Toxicology*, *89*, 137-139.

699 van Dijk, J. P., Cankar, K., Scheffer, S. J., Beenen, H. G., Shepherd, L. V, Stewart, D., … Kok, E. J.
700 (2009). Transcriptome analysis of potato tubers-effects of different agricultural practices. *Journal*
701 *of Agricultural and Food Chemistry*, *57*, 1612–1623.

702 van Dijk, J. P., de Mello, C. S., Voorhuijzen, M. M., Hutten, R. C. B., Arisi, A. C. M., Jansen, J. J., …
703 Kok, E. J. (2014). Safety assessment of plant varieties using transcriptomics profiling and a one-
704 class classifier. *Regulatory Toxicology and Pharmacology*, *70*, 297–303.

705 Wu, X., Gong, F., Yang, L., Hu, X., Tai, F., & Wang, W. (2015). Proteomic analysis reveals differential
706 accumulation of small heat shock proteins and late embryogenesis abundant proteins between
707 ABA-deficient mutant vp5 seeds and wild-type Vp5 seeds in maize. *Frontiers in Plant Science*, *5*,
708 1–18.

709 Yadav, S. K., Singla-Pareek, S. L., Ray, M., Reddy, M. K., & Sopory, S. K. (2005). Methylglyoxal levels
710 in plants under salinity stress are dependent on glyoxalase I and glutathione. *Biochemical and*
711 *Biophysical Research Communications*, *337*, 61–67.

712 Zeljenková, D., Aláčová, R., Ondrejková, J., Ambrušová, K., Bartušová, M., Kebis, A., … Steinberg, P.
713 (2016). One-year oral toxicity study on a genetically modified maize MON810 variety in Wistar Han
714 RCC rats (EU 7th Framework Programme project GRACE). *Archives of Toxicology*, *90*, 2531–
715 2562.

716 Zeljenková, D., Ambrušová, K., Bartušová, M., Kebis, A., Kovrižnych, J., Krivošíková, Z., … Steinberg,
717 P. (2014). Ninety-day oral toxicity studies on two genetically modified maize MON810 varieties in
718 Wistar Han RCC rats (EU 7th Framework Programme project GRACE). *Archives of Toxicology*,
719 *88*, 2289–314.

720

Figure Captions

Figure 1. Maize transcriptomics and proteomics. (A) Clustering of transcriptome data of maize embryos. Cluster dendrogram of the RNA-Seq completely processed data. Hierarchical cluster obtained from Euclidean distance matrix data using the complete-linkage cluster method in the R 'dendextend' package ('hclust' function). Every pair of GM and the corresponding near-isogenic are shown in a different colour (blue, PR33D48/PR32T16; green, DKC6667YG/DKC6666), the GM varieties labelled with an asterisk. Other conventional varieties are shown in black. Codes indicate the commercial identification of every variety (7 digits), growing season (2 digits) and experimental repeat (a and b). **(B) Maize grain proteomics.** Analysis of integrated variability of protein spots in the proteomes of 16 grain maize samples from 8 maize varieties and grown in two different seasons. Principal Component Analysis (PCA) of normalized spot volumes resulted in two principal components (PC1 and PC2) with Eigenvalues above 1, which explained 20.01% and 11.05% of the overall variability, respectively. Every pair of GM and the corresponding near-isogenic are shown in a different colour (blue, PR33D48/PR32T16; green, DKC6667YG/DKC6666), the GM varieties are labelled with an asterisk. Other conventional varieties are shown in black. Codes indicate the commercial identification of every variety (7 digits) and growing season (2 digits).

Figure 2. Maize metabolomics. (A) Graphical representation of the metabolite profiles of the 8 maize varieties grown in the 2012 season. PCA score plot showing the groupings of the two MON810 varieties (red and pink, labelled with an asterisk), their respective near-isogenic varieties (black and blue) and the other four maize varieties. **(B)** OPLS-DA S-plot identifying the 5 metabolites (circled in blue) that best represent the group separation between the GM variety DKC6667 and near-isogenic line (DKC6666). The possible identities of the 5 metabolites are indicated in Table 4a and have the following monoisotopic mass values: 258.1064; 441.1987; 496.3357; 518.3156 and 520.3388. **(C)** OPLS-DA S-plot identifying the 3 metabolites (circled in blue) that best represent the group separation between the GM variety PR33D48 and the near-isogenic line (PR32T16). The identities of the 3 metabolites are indicated in Table 4a and have the following monoisotopic mass values: 496.3357; 518.3156 and 438.2361.

Tables

Table 1. Maize samples and omics approaches. Maize samples used in 90-day feeding trials by Zeljenková et al. (2014) are shown in blue; and those used in 1-year trial by Zeljenková et al. (2016) are shaded in orange. e, embryo; k, kernel.

Maize type	Variety	Company*	Season	Transcriptomics	Proteomics	Metabolomics
GM, MON810	DKC6667YG	M	2012	e	k	k
non-GM, near-isogenic (of DKC6667YG)	DKC6666	M	2012	e+k	k	k
GM, MON810	PR33D48	P	2012	e+k	k	k
non-GM, near-isogenic (of PR33D48)	PR32T16	P	2012	e+k	k	k
non-GM	DKC6815	M	2012	e+k	k	k
non-GM	PR33W82	P	2012	e+k	k	k
non-GM	SYNEPAL	K	2012	e+k	k	k
non-GM	PR32T83	P	2012	e	k	k
non-GM	DKC6717	M	2012	k	k	
GM, MON810	DKC6667YG	M	2013		k	
non-GM, near-isogenic (of DKC6667YG)	DKC6666	M	2013		k	
GM, MON810	PR33D48	P	2013	e+k	k	
non-GM, near-isogenic (of PR33D48)	PR32T16	P	2013	e+k	k	
non-GM	DKC6815	M	2013	e+k	k	
non-GM	PR33W82	P	2013		k	
non-GM	SYNEPAL	K	2013		k	
non-GM	Alinea	MS	2013	k		
non-GM	Calcio	MS	2013	k		
non-GM	Helen	A	2013	k		
non-GM	Laricio	MS	2013	k		
non-GM	MAS37V	MS	2013	k		
non-GM	MAS70F	MS	2013	k		
non-GM	MAS74G	MS	2013	k		
non-GM	Tietar	M	2013	k		

* M, Monsanto; P, Pioneer Hi-Bred; K, Koipesol Semillas; A, Advanta; MS, Maisadour Semences

Table 2. Maize embryo transcriptomics. (A) Results of the sequencing and mapping RNA-seq. Samples are identified with the variety, type and season. All values correspond to the mean of two biological replicates. Alignment metrics include the number of reads either not aligned to the maize reference genome, aligned to one chromosomal locus (1 time) or aligned to multiple loci (> 1 time); and the corresponding percentages. **(B) Test set prediction per variety**. **(C) Classification of GM varieties.**

A

	Maize samples	DKC6667YG	DKC6666	PR33D48	PR32T16	DKC6815	PR33W82	SYNEPAL	PR32T83	PR33D48	PR32T16	DKC6815	
	type	GM	near-isogenic	GM	near-isogenic	conventional	conventional	conventional	conventional	GM	near-isogenic	conventional	
	season	2012	2012	2012	2012	2012	2012	2012	2012	2013	2013	2013	mean values
	Cleaning metrics												
	Raw data (nucleotide count)	2,122,648,101	2,121,420,627	2,181,259,500	2,182,735,723	2,183,181,256	2,178,711,574	2,170,238,763	2,177,544,394	2,181,084,276	2,121,843,668	2,122,528,345	2,158,472,384
Read number	**Raw data**	43,319,349	43,294,299	44,515,500	44,545,627	44,554,720	44,463,502	44,290,587	44,439,682	44,511,924	43,302,932	43,316,905	44,050,457
	Clean	42,021,000	41,808,604	43,471,907	43,369,119	43,490,414	42,567,005	42,194,612	41,858,153	43,032,206	42,233,753	42,092,465	42,558,112
	Trimmed	42,020,166	41,807,770	43,471,025	43,368,312	43,489,601	42,566,081	42,193,855	41,857,233	43,031,311	42,232,958	42,091,690	42,557,273
	Non-RNA	37,459,017	37,502,163	38,732,823	39,901,493	39,113,021	36,377,222	36,598,773	35,258,889	38,567,809	37,707,471	38,488,883	37,791,597
	Clean data percentage	86%	87%	87%	90%	88%	82%	83%	79%	87%	87%	89%	86%
	Alignment metrics												
	Unaligned	7,695,164	7,624,537	6,664,061	5,966,947	5,964,836	9,669,537	8,919,362	9,702,918	6,569,382	7,835,637	7,415,911	7,638,935
		21%	20%	17%	15%	15%	27%	24%	28%	17%	21%	19%	20%
	Aligned 1 time	24,858,887	24,671,382	26,442,883	28,327,016	27,507,044	21,717,889	22,614,257	20,721,023	26,471,074	24,563,265	25,744,698	24,876,311
		66%	66%	68%	71%	70%	60%	62%	59%	69%	65%	67%	66%
	Aligned > 1 time	4,904,967	5,206,244	5,625,880	5,607,530	5,641,141	4,989,797	5,065,154	5,182,527	5,527,354	5,308,570	5,328,275	5,307,949
		13%	14%	15%	14%	14%	14%	14%	15%	14%	14%	14%	14%

Table 2. Maize embryo transcriptomics.

B

	Total classifications	Classified 'in'	
		number	percentage
Grain transcriptome			
Total test set	182	167	91.8
Alinea	13	13	100.0
Calcio	13	13	100.0
DKC6717	13	13	100.0
DKC6666	13	13	100.0
DKC6815	13	13	100.0
Helen	13	13	100.0
Laricio	13	0	0.0
MAS37V	13	12	92.3
Mas70F	13	12	92.3
Mas74G	13	13	100.0
PR32T16	13	13	100.0
PR33W82	13	13	100.0
SY.NEPAL	13	13	100.0
Tietar	13	13	100.0
Embryo transcriptome			
Total test set	60	51	85.0
PR32T83	10	10	100.0
DKC6666	10	10	100.0
DKC6815	7	10	70.0
PR32T16	10	10	100.0
PR33W82	10	10	100.0
SY.NEPAL	4	10	40.0

Table 2. Maize embryo transcriptomics.

C

	Replicates	Total classifications	Classified 'in'	Percentage 'in'	By majority vote	By test set threshold
Grain transcriptome						
Total test set		182	167	91.8		
PR33D48 (1) (GM)	1	182	182	100	in	in
PR33D48 (2) (GM)	1	182	182	100	in	in
Embryo transcriptome						
Total test set		60	51	85		
PR33D48 (GM)	2	60	60	100	in	in
DKC6667YG (GM)	2	60	60	100	in	in

Table 3. Maize grain proteomics. Protein spots showing differential accumulation on pairwise comparison (t-test) of the proteomes of two MON810 and near-isogenic variety pairs grown in two different seasons. Proteome profiles were obtained using 2D IEF SDS-PAGE and spot identification was performed by LC-MSMS.

			Factor of change (log2)						
			DKC6667YG / DKC6666		PR33D48 / PR32T16				
Spot ID	Mass (kDa)	pI	2012	2013	2012	2013	Accession	Description	Function
6	71.09	6.73		1.26			K7W272	Vicilin-like antimicrobial peptides 2-2	Storage proteins
23	27.16	5.57		3.43			B6UH67	Late embryogenesis abundant protein D-34	Stress and defense response
28	71.09	6.73		1.55			K7W272	Vicilin-like antimicrobial peptides 2-2	Storage proteins
69	27.16	5.57		1.60			B6UH67	Late embryogenesis abundant protein D-34	Stress and defense response
70	18.72	4.67		-2.82			B4FL17	Translationally-controlled tumor protein homolog	Protein folding and assembly
331	21.17	5.44		1.31			B6SNS4	Late embryogenesis abundant protein D-34	Stress and defense response
482	27.16	5.57		1.48			B6UH67	Late embryogenesis abundant protein D-34	Stress and defense response
669	21.17	5.44		1.24			B6SNS4	Late embryogenesis abundant protein D-34	Stress and defense response
501	21.17	5.44		1.62			B6SNS4	Late embryogenesis abundant protein D-34	Stress and defense response
85	21.07	5.63			13.18		B6T8D8	Lactoylglutathione lyase	Metal ion binding
202	64.86	6.86			1.43		C0PGM3	Uncharacterized protein	Nutrient reservoir activity
185	49.89	6.61			-1.63		Q7M1Z8	Globulin-2 OS=Zea mays	Nutrient reservoir activity
169	32.32	6.19				1.07	C0PK05	Lactoylglutathione lyase	Metal ion binding
179	22.87	6.43				3.98	B6TXB5	22.0 kDa class IV heat shock protein	Stress response
363	33.25	6.43				1.55	B4FLJ4	rRNA N-glycosidase	Defense response

Table 4. Maize metabolomics. (A) Metabolites produced in MON810 and near-isogenic lines at different concentrations. All these metabolites were extracted from the OPLS-DA output and were identified by ChemSpider database (Pence & Williams, 2010) using the monoisotopic mass approach. The number of metabolite hits within the mass interval range of 0.001 or 0.0001 is indicated for each metabolite. (**B) Test set prediction per variety. (C) Classification of GM varieties.**

A

		Factor of change		Metabolite identified using ChemSpider database			
Mono-isotopic Mass (DA)	**Retention time (min)**	**DKC6667YG / DKC6666 2012**	**PR33D48 / PR32T16 2012**	**Mono-isotopic Mass (DA)**	**Molecular formula**	**ChemSpider ID**	**Metabolite and number of hits with similar monoisotopic mass**
258.1064	0.95	1.3		257.098663	$C_{14}H_{15}N_3S$	114422	6-Methyl-5,6,6a,7-tetrahydro-4H-benzo[de][1,3]thiazolo[4,5-g]quinolin-9-amine *(1st of 1967 hits at ±0.0001)*
441.1987	12.31	1.3		440.190948	$C_{20}H_{24}N_8O_4$	8340044	1,4-Bis(4,6-diacetyl-1,3,5-triazin-2-yl)-1,4-diazoniabicyclo[2.2.2]octane *(only hit at ± 0.0001)*
496.3357	25.48	1.3	2.9	495.328156	$C_{22}H_{41}N_9O_4$	150667	N^5-(Diaminomethylene)-L-ornithyl-L-prolyl-L-lysyl-L-prolinamide *(1st of 6 hits at ± 0.001)*
518.3156	25.48	1.3	3.1	517.308655	$C_{27}H_{43}N_5O_3S$	2423355	3-[3-(Diethylamino)propyl]-1-[(6-ethoxy-2-oxo-1,2-dihydro-3-quinolinyl)methyl]-1-[3-(4-morpholinyl)propyl]thiourea *(one of 19 hits at ± 0.001)*
520.3388	25.11	1.3		519.330811	$C_{28}H_{45}N_3O_6$	16537387	Methyl N-{(2,4-dimethylphenyl)[(2-methyl-2-propanyl)(N-{[(2-methyl-2-propanyl)oxy]carbonyl}leucyl)amino]acetyl}glycinate *(1st of 221 hits at ± 0.001)*
438.2361	8.67		1.6	437.228333	$C_{26}{}^{13}CH_{32}O_5$	48059603	(4aR,4bS,6aS,7S,9aS,9bS)-3-[(1-^{13}C)Ethanoyl]-4a,6a-dimethyl-2-oxo-2,3,4,4a,4b,5,6,6a,7,8,9,9a,9b,10-tetradecahydroindeno[5,4-f]chromen-7-yl benzoate *(only hit at ± 0.0001)*

Table 4. Maize metabolomics.

B

	Total classifications	Classified 'in'	
		number	percentage
RIKILT			
Total test set	198	127	64.1
DKC6717	12	12	100.0
DKC6815	36	25	69.4
PR32T83	24	20	83.3
PR33W82	36	27	75.0
SY-NEPAL	36	19	52.8
DKC6666	24	7	29.2
PR32T16	30	17	56.7
CSIR			
Total test set	120	97	80.8
DKC6815	24	20	83.3
PR32T83	24	23	95.8
PR33W82	24	12	50.0
SYNEPAL	24	24	100.0
DKC6666	12	7	58.3
PR32T16	12	11	91.7

Table 4. Maize metabolomics.

C

	Replicates	Total classifications	Classified 'in'	Percentage 'in'	By majority vote	By test set threshold
RIKILT						
Total test set		198	127	64.1		
DKC6667YG (GM)	2	84	50	59.5	in	out
PR33D48 (GM)	3	126	107	84.9	in	in
Fungal infected sample 1	2	84	0	0.0	out	out
Fungal infected sample 2	2	84	0	0.0	out	out
CSIR						
Total test set		120	97	80.8		
DKC6667YG (GM)	MA4 a	24	24	100.0	in	in
	MA4 b	24	24	100.0	in	in
	MA4 c	24	24	100.0	in	in
PR33D48 (GM)	MB8 a	24	24	100.0	in	in
	MB8 b	24	24	100.0	in	in
	MB8 c	24	24	100.0	in	in

Supplementary Materials

Material S1. Transcriptomic analysis

RNA extraction and Illumina sequencing – RIKILT-WUR

In short, 5 ml to 60° C pre-warmed RNA extraction buffer (2% CTAB, 2% PVP k30, 100 mM TRIS pH 8.0, 2.0 M NaCl and 2% β-mercaptoethanol added just before use) was added to 0.5 g lyophilized sample and vortexed vigorously. An equal amount of CIA (chloroform:isoamylalcohol (24:1)) was added and the mixture was vortexed vigorously for 15 sec. Subsequently, the mixture was centrifuged for 10 to 15 min at ~8014 g, 15°C after which the aqueous phase was transferred to a new tube using a pipette. CIA extraction was repeated twice. Lithium chloride was added (1/4 volume, 10 M) to the remaining aqueous phase, mixed well and stored O/N at 4°C.

The second day, the tubes were centrifuged for 30 min at ~8014 g, 4°C before pouring off the supernatant. 500 µl to 60° C pre-warmed SSTE buffer (1.0 M NaCl, 0.5% SDS, 10 mM TRIS pH 8.0 and 1 mM EDTA pH 8.0) was added to dissolve the pellet and the solution was transferred to an Eppendorf tube. 500 µL of CIA was added, the mixture was vortexed vigorously for 15 sec and centrifuged for 10 min at 14.000 g, RT. After transferring the aqueous phase to a new tube, 2 volumes of 96% ethanol were added. The tubes were cooled on ice for 5 min and kept on ice as much as possible for the remainder of the protocol. The tubes were centrifuged for 30 min at 14.000 g, 4°C and were drained afterwards. 250 µL 75% ethanol was added and the solution was mixed well. After centrifugation for 10 min at 14.000 g, 4°C the supernatant was pipetted off carefully and pellets were dried at room temperature for 10 min. To dissolve the RNA 100 - 150 µL of 10mM TRIS, pH >7 was added, and tubes were placed at ~65°C for 10 min. Afterwards the solution was pipetted up and down. RNA samples were stored at -80° C until further use.

Bioinformatic analysis

The gene annotation of the reference genome was used to calculate the number of reads mapping every gene on the different analysed samples using the HTSeq_v0.6.1 software. The following parameters were used by default: Multi-mapping reads mode, union; feature type: exon; id attribute, gene_id. Stranded (-s) was set to “no”. Prior to normalization, a set of annotation quality control parameters were

checked. First, the R package UndetectedGenes (http://computational.biology.langebio.cinvestav.mx/DOWNLOAD/UndetectedGenes/) was used to estimate the number of undetected or missing genes that are likely to be expressed in an RNA-seq library but were missed in a particular sample of the library. Second, a quality of the alignment data and of the metrics and bias estimations was conducted using Qualimap v2.2 (Online Resource 1). Finally, a GC bias analysis and correction was conducted with the R package NOISeq v2.16. The GC-normalized count data set was subjected to the standard TMM normalization method using the R package NOISeq (v2.16.0), using its function tmm with parameters long=1000 and lc=0.

Material S2. Proteomic analysis

2-D proteomic analysis: protein extraction

To prepare protein extracts, 150 mg of milled maize was solubilized in extraction buffer (50 mM Tris-HCl pH 7.5,0.5 M NaCl and 0.3% Triton X-100) containing a protease inhibitor cocktail (10 µg/L aprotinin, 0.5 µg/L leupeptin, 1 µg/Lpepstatin, 1 µg/LE-64 and 0.1 mM de PMSF), 1 U DNase I and 1 U RNase A. After centrifugation at 15,000 rpm for 15 min at 4 ºC, TCA was added to the supernatant up to 15% and the solution was incubated for 30 min on ice. The protein extract was centrifuged at 15,000 rpm for 15 min at 4ºC and the pellet was then washed with acetone at 4ºC. The samples were sonicated in an ultrasonic cleaner Bransonic 2510 (Branson, Connecticut, EEUU) device and re-centrifuged in the same conditions. After drying in the open air for 30 min, protein pellets were resuspended in lysis buffer (50 mM Tris-HCl pH 8, 7 M urea, 8 M thiourea, 4% chaps).

Material S3. Metabolomic analysis

UHPLC-MS metabolomic analysis - CSIR

Milled grain (1 g) was homogenized in 10 mL extraction solution (75% Methanol absolute (HPLC supra-gradient grade) and 0.1% Formic acid) using a vortex followed by sonication in a water bath for 5 min at room temperature. Milli-Q water (3.3 mL) was added to each sample, vortexed twice for 15 sec each time followed by sonication for 15 min. The samples were extracted for 1 hour in a rotation mixer followed by 15 min sonication. The homogenates were the centrifuged at 5000 rpm for 10 min at room

temperature and the resulting supernatants were transferred by pipetting into 12 mL tube with a cap. The extracted metabolite samples were diluted 5x in methanol: water (60:40), filtered using Pall Acrodisc syringe filters (GHP membrane, 0.2 µm) and transferred to LCMS vials and stored at -80ºC.

LC-MS metabolomic analysis – RIKILT-WUR

One gram of ground, dried plant material was weighed into a 40 mL amber glass tube (Grace Alltech). 10.0 mL methanol/FA (0.125% v/v formic acid) was added to inactivate the enzymes and the mixture was immediately vortexed until homogenous, followed by for 5 min. at 35 kHz in a sonication bath (VWR) at room temperature (20° C). When the water was warmed up too much during sonication ice was added to the water. 3.3 ml Milli-Q water (Millipore) was added to each sample in order to have a final concentration of 75% methanol and the mixture was immediately vortexed. After adding water to all samples, 15 additional seconds of vortexing were carried out. Samples were sonicated for 15 min at room temperature, again ice was added when the water was warmed up during the process. Extractions were performed head-over-head for 1 h. Samples were sonicated for 15 min. at room temperature and subsequently centrifuged for 10 min, 5,000 rpm at room temperature. Samples were pipetted into a 12 mL tube (Greiner). As a control sample 250 µL of each sample was combined in an extra 12 mL tube. Samples were stored at -80° C until further analysis. On the day of analysis, samples were taken out of the -80˚C freezer and thawed.1.5 mL supernatant was transferred into a 2 mL tube, put on ice for 10 min and centrifuged at 13,000 rpm for 10 min at 4°C.

SUPPLEMENTARY TABLES

Table S1. Embryo RNA-seq: alignment and count data quality control summary.

									Reads genomic origin (%)			Transcript coverage profile			Junction analysis (%) sample		
Maize variety	**Type**	**Season**	**Sample code**	**% Clipped reads**	**% GC**	**Coverage**	**Mean map. quality**	**% General error rate**	**Exonic**	**Intronic**	**Intergenic**	**5' bias**	**3' bias**	**5'-3' bias**	**Known**	**Partly known**	**Novel**
DKC6815	non-GM	2012	**DKC6815_12a**	9.30	51.1	2.7 ± 26.8	23.7	0.07	84.9	5.9	9.1	0.59	0.31	1.67	80	4.5	15.4
			DKC6815_12b	9.00	53.9	3.3 ± 34.9	24.7	0.07	83.7	7.4	8.9	0.58	0.36	0.15	79	4.4	15.8
		2013	**DKC6815_13a**	8.90	52.5	2.4 ± 32.0	24.5	0.08	82.9	8.5	8.7	0.59	0.34	1.59	80	4.4	15.4
			DKC6815_13b	9.80	50.4	2.0 ± 36.9	23.9	0.09	80.0	8.0	12.0	0.56	0.37	0.15	78	4.5	16.9
PR32T16	non-GM near-isogenic	2012	**PR32T16_12a**	9.30	51.3	2.6 ± 33.3	24.0	0.07	83.6	7.0	9.4	0.53	0.32	1.75	80	4.4	15.5
			PR32T16_12b	9.60	52.7	3.5 ± 37.8	23.8	0.07	83.7	6.8	9.5	0.54	0.38	1.53	79	4.4	15.6
		2013	**PR32T16_13a**	9.00	52.1	2.2 ± 33.8	24.1	0.09	84.2	6.8	9.1	0.53	0.32	1.62	80	4.4	15.2
			PR32T16_13b	10.20	51.3	2.3 ± 30.9	24.0	0.09	79.4	9.7	10.9	0.5	0.33	1.55	78	4.5	16.6
PR33D48	GM	2012	**PR33D48_12a**	9.10	51.6	2.5 ± 33.3	23.7	0.07	84.4	5.8	9.8	0.46	0.31	1.71	79	4.6	15.6
			PR33D48_12b	9.30	48.3	1.4 ± 64.2	25.2	0.12	71.6	9.1	19.2	0.26	0.34	1.25	77	4.5	18.4
		2013	**PR33D48_13a**	9.00	51.6	2.1 ± 28.0	23.7	0.09	83.2	6.9	9.9	0.51	0.31	1.68	80	4.8	15.2
			PR33D48_13b	9.40	50.7	2.2 ± 35.3	24.2	0.09	78.3	8.8	12.9	0.45	0.37	1.61	78	4.7	16.6
DKC6666	non-GM near-isogenic	2012	**DKC6666_12a**	9.60	51.2	2.1 ± 30.7	23.6	0.09	81.8	7.9	10.3	0.53	0.35	1.52	80	4.4	15.4
			DKC6666_12b	9.70	53.1	2.8 ± 37.6	24.0	0.08	83.2	6.9	9.9	0.65	0.42	1.42	80	4.3	15.3
DKC6667YG	GM	2012	**DKC6667_12a**	9.00	50.9	2.2 ± 27.2	23.4	0.08	83.3	7.6	9.1	0.59	0.35	1.64	80	4.5	15.1
			DKC6667_12b	9.80	50.5	2.6 ± 34.7	23.7	0.08	81.1	7.8	111.2	0.55	0.37	1.6	79	4.3	16.1
PR33W82	non-GM	2012	**PR33W82_12a**	11.50	48.2	1.7 ± 57.1	24.3	0.11	76.9	7.3	15.7	0.35	0.38	1.3	78	4.1	17.4
			PR33W82_12b	11.20	47.7	3.4 ± 77.6	24.0	0.06	72.3	8.9	18.8	0.37	0.37	1.44	76	4.0	19.1
PR32T83	non-GM	2012	**PR32T83_12a**	9.50	53.6	2.1 ± 39.3	25.7	0.11	80.8	9.3	9.9	0.59	0.42	1.37	80	4.3	15.5
			PR32T83_12b	10.30	51.0	2.1 ± 35.2	25.6	0.11	75.9	12.3	11.8	0.55	0.45	1.41	78	4.5	16.7
SYNEPAL	non-GM	2012	**SYNEPAL_12a**	11.30	50.7	2.8 ± 63.3	23.9	0.07	80.2	7.0	12.7	0.47	0.39	1.46	79	4.0	16.6
			SYNEPAL_12b	10.90	48.6	1.5 ± 77.0	24.4	0.12	74.9	9.1	16.0	0.32	0.33	1.45	79	4.3	16.1

Table S2. List of primers used in RT-qPCR based validation of RNA-Seq results.

Name	Sense	Sequence(5'→3')	Isoform ID
Maize1	forward	ACCTCATCACAGATCAGGATTTCA	GRMZM2G047274_T01
	reverse	GTTCACGGGACTGGGTCTTA	
Maize2	forward	TCATTATCGGCACAAAAGGACA	GRMZM2G098999_T01
	reverse	GCTACCGATAGCTGCTTTGGA	
Maize3	forward	AGTTGAGGCTGGTGTCGATT	GRMZM2G098875_T02
	reverse	TTGAGGTGCTTTGACTCTCTG	
Maize4	forward	TGTCGCTAGCTGTCAGTGTC	GRMZM2G316362_T01
	reverse	TCCAATCTGGGTTCCAAATCGT	
Maize5	forward	TCCGCACGAAAAACATCACC	GRMZM2G002805_T01
	reverse	TGAGTCATCTCTTCGCGGTC	
Maize6	forward	GGAGCTCTGAACAGTCAAACG	GRMZM2G346839_T01
	reverse	ACGTCGTGGAACATACAGAACA	
Maize7	forward	GCGTGACGATGCATTCAGAC	GRMZM2G375517_T01
	reverse	GCAACGGGATCAATACGCAC	
Maize8	forward	AAGCGATTGCAATAGGGACCTC	GRMZM2G456487_T01
	reverse	GGACTCCATGTCAGTGCTACC	
Maize9	forward	TGTCAAGAGAAGGTGGGACGA	GRMZM2G154278_T02
	reverse	CGGTGGAAGTCACTCCTGAT	
Maize10	forward	ATCGGGAGGAAGGCAACAAG	GRMZM2G000236_T01
	reverse	TAGAAGGTGGAGCGGTCGTA	
Ubi	forward	TAAGCTGCCGATGTGCCTGCGTCG	GRMZM2G409726_T01
	reverse	CTGAAAGACAGAACATAATGAGCACAGGC	

Primers were designed and *in silico* tested for specificity using PrimerBlast Software.

Table S3. UHPLC gradient conditions for the chromatographic analysis.

A= 0.2% formic acid (CSIR: 99% UPLC-MS Sigma-Aldrich; RIKILT-WUR: 99% ULC-MS Biosolve) in deionized water (CSIR: Milli-Q Advantage A10, Millipore-Merck; RIKILT-WUR: ULC-Biosolve) and B= 0.2% formic acid (99% UPLC-MS Sigma-Aldrich) in acetonitrile (CSIR: UPLC Burdick & Jackson, Honeywell; RIKILT-WUR: ULC-Biosolve).

Gradient		
Time (min)	**Solvent A (%)**	**Solvent B (%)**
0	100	0
2	100	0
13	70	30
18	70	30
24	0	100
38	0	100
38	100	0
44	100	0

Table S4. Differentially expressed genes in MON810 and near-isogenic varieties, without distinguishing company or season (MON810 / near-isogenics) and only considering DKC6667YG and DKC6666 grown in 2012.

	MON810 / near-isogenics			DKC6667YG/DKC6666 (2012)			
gene ID	**log FC**	**adj. P val**	**B val**	**log FC**	**adj. P val**	**B val**	**gene description**
GRMZM2G152436	-0.67	1.2E-01	1.61				uncharacterized protein
GRMZM2G098679	-1.81	7.4E-02	1.20				sugar transporter
GRMZM2G047097	0.53	2.0E-01	1.20				hypothetical protein
GRMZM2G456487	-2.00	8.9E-03	3.72	-1.73	1.9E-01	1.34	putative WAK receptor-like protein kinase
GRMZM5G845024				2.61	1.3E-01	2.51	hypothetical protein
GRMZM2G098999				-2.20	1.8E-01	2.11	F-Box protein
AC234185.1_FG004				-2.11	1.8E-01	1.89	GATase1 CTP Synthase
GRMZM2G481605				1.75	1.8E-01	1.73	heat shock protein
GRMZM2G049767				1.68	1.8E-01	1.71	heat shock protein
GRMZM2G123842				-2.85	1.8E-01	1.64	hypothetical protein
GRMZM2G173090				1.53	1.9E-01	1.57	heat shock protein
GRMZM2G369182				-2.25	1.8E-01	1.25	hypothetical protein
GRMZM5G831224				1.62	1.9E-01	1.24	hypothetical protein
GRMZM2G122028				1.60	2.1E-01	1.10	hypothetical protein
GRMZM2G003489				1.28	2.1E-01	1.04	hypothetical protein

SUPPLEMENTARY FIGURES

Fig S1. Distribution of genes as a function of their count number, in each sample. For each sample the percentages of genes with counts per million (CPM) above 10, 5, 2, 1 and 0 are represented. Maize varieties and seasons are indicated below, with two columns corresponding to two replicates. The horizontal lines are the percentage of features (genes with CPM above 10, 5, 2, 1 or 0) with those CPM in at least one of the samples. 100% sensitivity corresponds to the 39,625 genes in the maize reference genome.

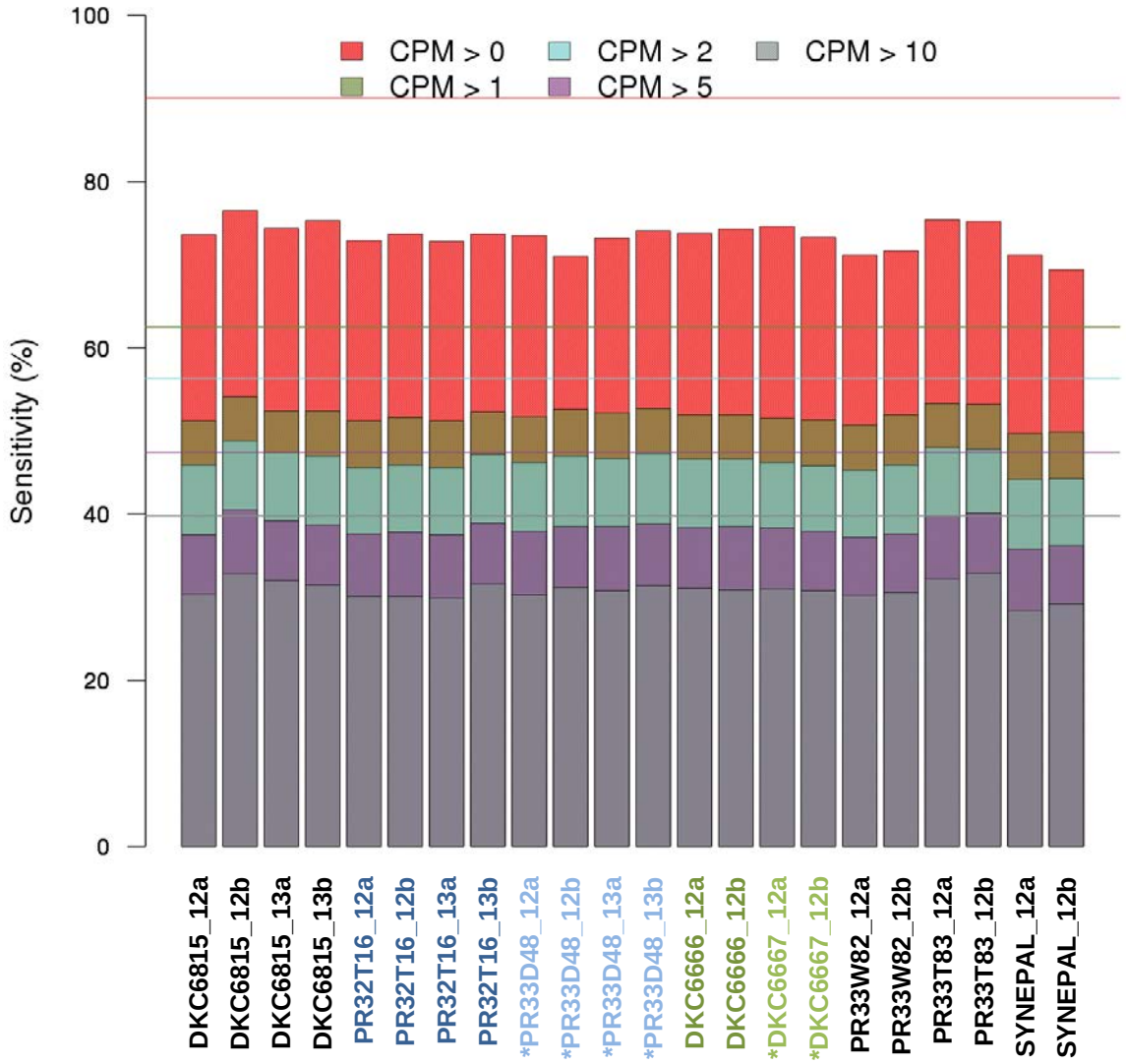


Fig. S2. Analysis of integrated variability of gene expression in the transcriptomes of maize kernel samples. Principal Component Analysis (PCA) score plots of the transcriptomic profiles of kernel maize profiles with grouping per variety. Scores for principal components 1 and 2 (PC1 and PC2) are plotted on X- and Y-axis, along with the percentage of the total variation they explain. Codes for the fourteen distinct varieties are indicated. (Infometrix (2015). Pirouette 4.5. (https://infometrix.com/pirouette/).

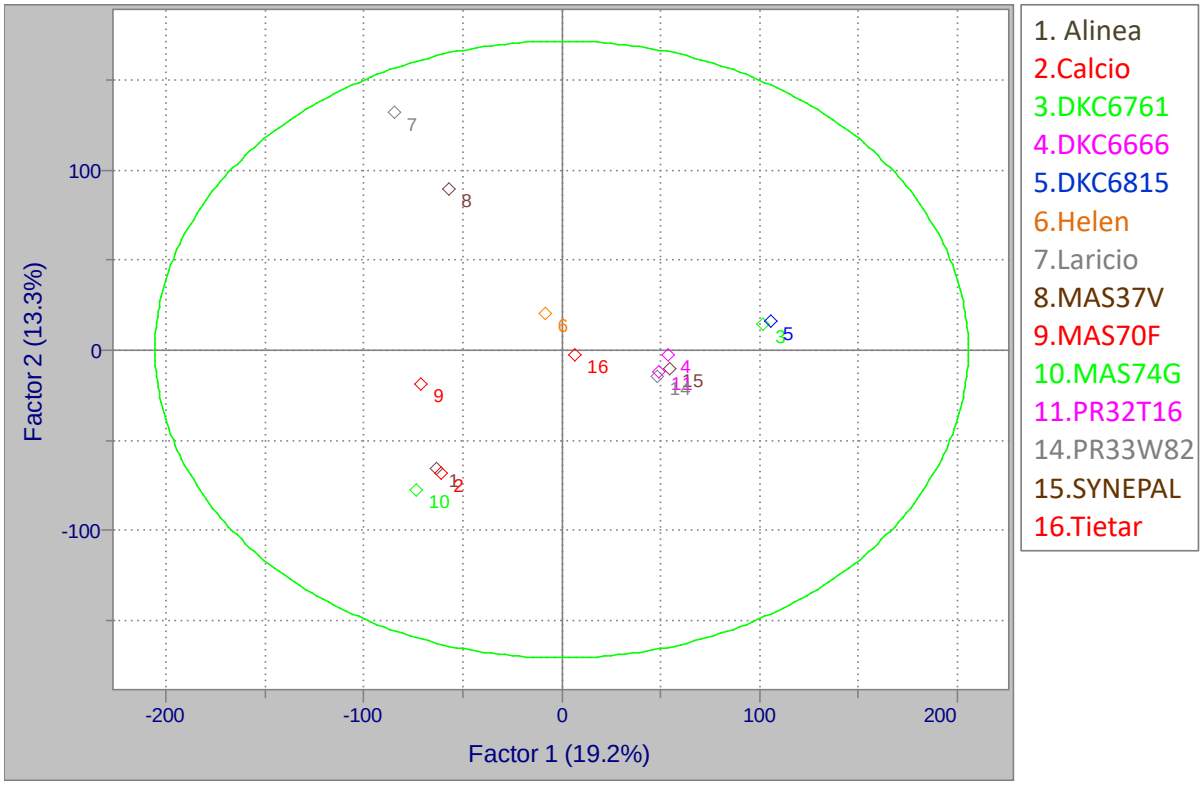


Figure S3. Example of a typical 2D IEF SDS-PAGE analysis of maize seeds. Differential spots identified in Table 4 are labelled in red.

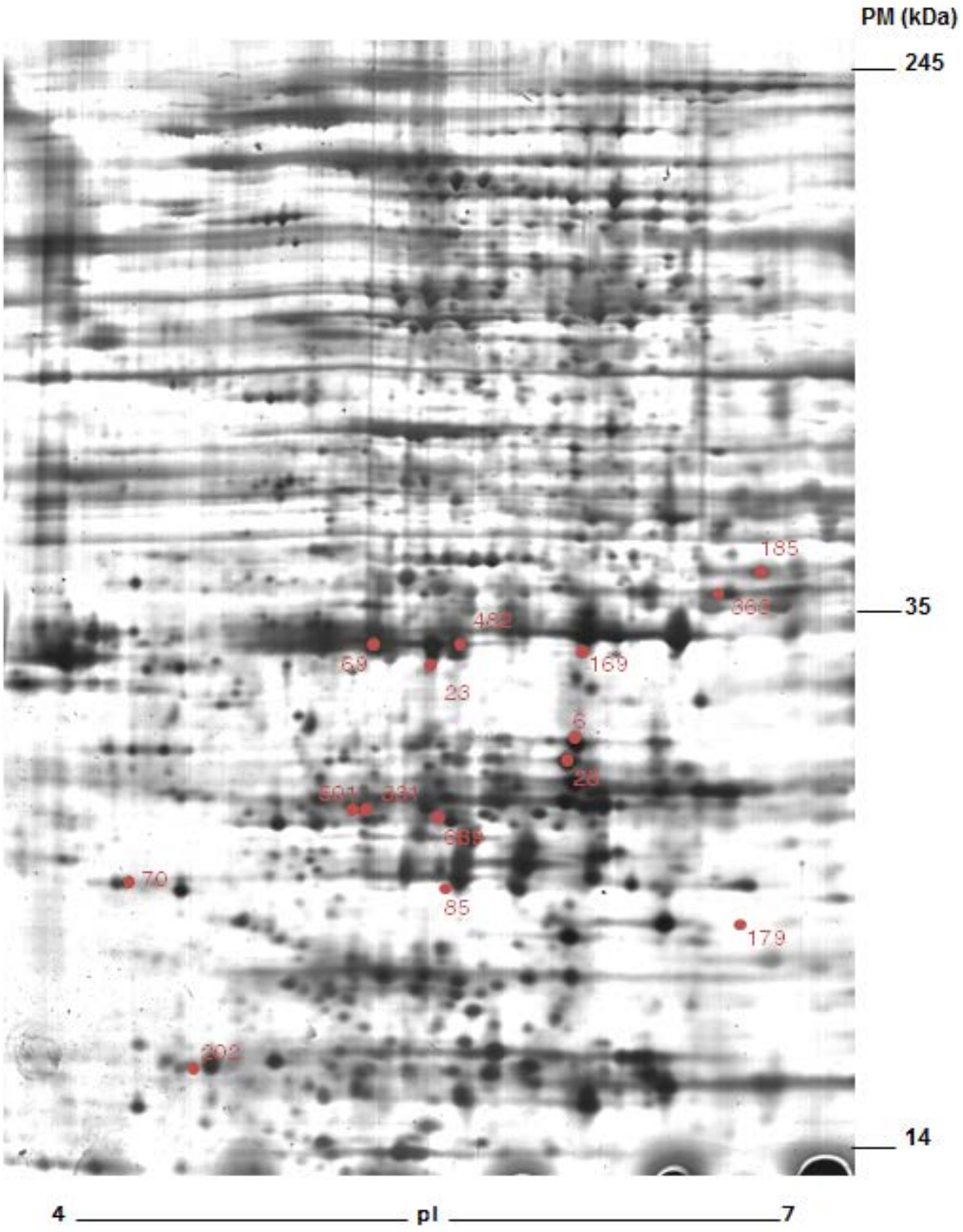
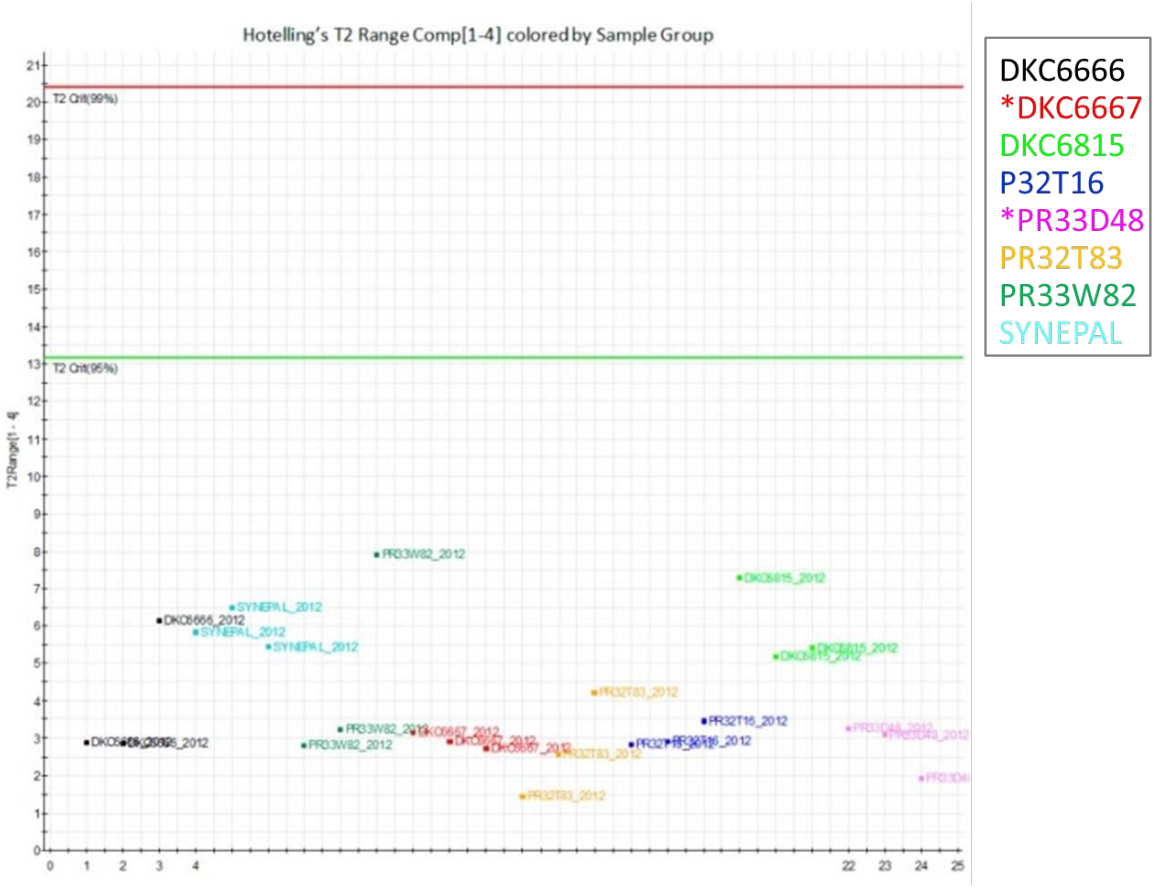
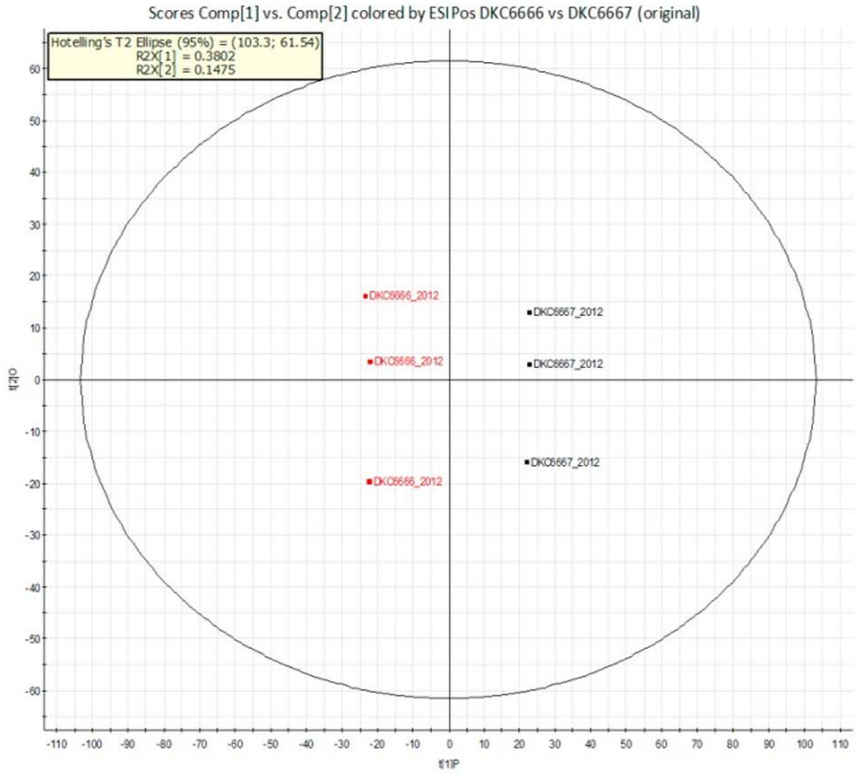


Figure S4. Maize metabolomics: graphical representation of multivariate data analysis of the metabolite profiles of maize varieties. (**A**) Hotelling's T^2 range plot (95% confidence interval) corresponding to the analysis of 8 varieties. (**B**) OPLS-DA score plot indicating the differences in the metabolite composition between the GM variety DKC6667_2012 and the near-isogenic line (DKC6666_2012). (**C**) OPLS-DA score plot indicating the differences in the metabolite composition between the GM variety PR33D48_2012 and the near-isogenic line (PR32T16_2012). (**D**) PCA score plot showing the patterns of metabolites among the six non-GM maize varieties as evidence of natural variation. * Transgenic varieties.

A

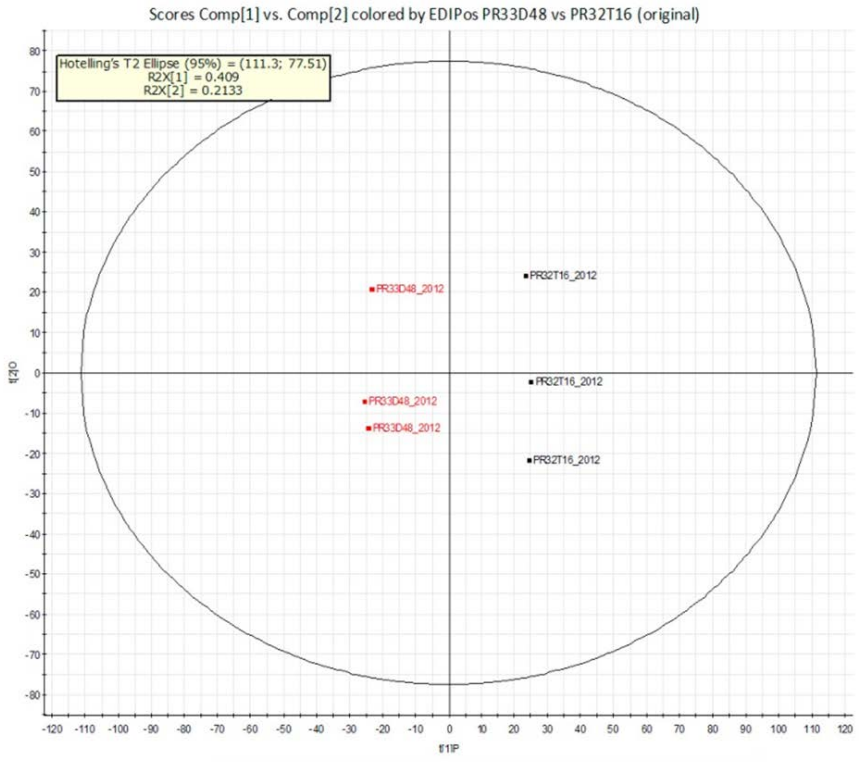


B



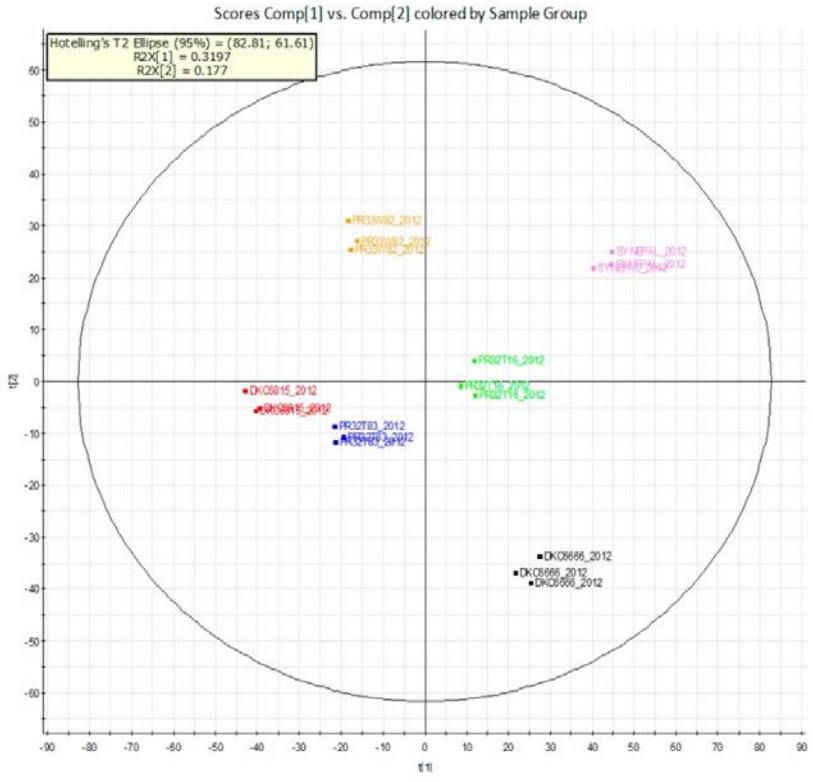
*DKC6667YG
DKC6666

C



PR32T16
*PR33D48

D



DKC6666
DKC6815
PR32T16
P32T83
PR33W82
SYNEPAL