Journal of Experimental Botany Advance Access published July 10, 2006

Journal of Experimental Botany, Page 1 of 13
doi:10.1093/jxb/erl025

Journal of Experimental Botany

RESEARCH PAPER

A metabonomic study of transgenic maize (*Zea mays*) seeds revealed variations in osmolytes and branched amino acids

Cesare Manetti*, Cristiano Bianchetti, Lorena Casciani, Cecilia Castro, Maria Enrica Di Cocco, Alfredo Miccheli, Mario Motto and Filippo Conti

Dipartimento di Chimica, Università degli studi di Roma 'La Sapienza', Piazzale Aldo Moro 5, I-00185 Roma, Italy and Istituto Sperimentale per la Cerealicoltura, Sez. Bergamo, Via Stezzano 24, I-24126 Bergamo, Italy

Received 27 October 2005; Accepted 12 April 2006

Abstract

The aim of the research was to investigate metabolic variations associated with genetic modifications in the grains of *Zea mays* using metabonomic techniques. With this in mind, the non-targeted characteristic of the technique is useful to identify metabolites peculiar to the genetic modification and initially undefined. The results obtained showed that the genetic modification, introducing *Cry1Ab* gene expression, induces metabolic variations involving the primary nitrogen pathway. Concerning the methodological aspects, the experimental protocol used has been applied in this field for the first time. It consists of a combination of partial least square-discriminant analysis and principal component analysis. The most important metabolites for discrimination were selected and the metabolic correlations linking them are identified. Principal component analysis on selected signals confirms metabolic variations, highlighting important details about the changes induced on the metabolic network by the presence of a Bt transgene in the maize genome.

Key words: GMO, metabolomics, metabonomics, multivariate analysis, NMR, *Zea mays*.

Introduction

There is more maize (*Zea mays*) produced in the world than any other crop (http://apps.fao.org). It is used as a model plant to evaluate chemical, physical, and environmental effects, and also in genetic studies.

The use of commercial transgenic crops expressing *Bacillus thuringiensis* (Bt) toxins has escalated in recent years because of their advantages over traditional chemical insecticides (James, 2004). Bt is an aerobic, Gram-positive, spore-forming bacterium commonly found in the environment. It produces a number of insect toxins, the most distinctive of which are protein crystals formed during sporulation (reviewed in de Maagd *et al*., 1999). The Cry proteins are selectively active against a narrow range of insects and, as a class of protein, are effective against a wide range of insect pests. They act by binding to specific sites of the midgut cells of susceptible insects and forming ion-selective channels in the cell membrane (Knowles and Dow, 1993). The *Cry1Ab* gene in the Bt strain HD-1 (Geiser *et al*., 1986) is insecticidal only to lepidopteran insects and is the prototype for the gene currently expressed in most commercial Bt maize events for the control of the European corn borer (*Ostrinia nubilalis*), a major insect pest of maize in North American and European agriculture (Betz *et al*., 2000).

The usefulness of this genetically modified crop will be cut short if the insects evolve resistance to Bt toxins. So far, field-evolved resistance to Bt crops has not been reported (Tabashnik *et al*., 2003). Accordingly, the 'high-dose-refuge' strategy is widely recommended by the biotechnology industry and regulatory authorities to delay pest adaptation to transgenic crops that produce Bt toxins. This involves cultivating non-transgenic plants (refuges) in close proximity to crops producing a high dose of Bt toxin (Chilcutt and Tabashnik, 2004). Gene flow in plants occurs when pollen from one plant fertilizes another plant, carrying with it the genes from the

* To whom correspondence should be addressed. E-mail: c.manetti@caspur.it

first plant; the transgenes may be integrated into the genome of the offspring, altering the traits of the offspring.

This study investigates the gene flow impact on the metabolism of non-modified plants once they get in contact with the modified one and produce hybrids containing the MON810 insertion event. To achieve this, a metabonomic strategy was chosen. This is an emerging holistic approach, complementary to genomics and proteomics, for studying the complex biological system response to chemical and physical input, and also to genetic variations (Lindon *et al*., 2001; Sumner *et al*., 2003). Because of their characteristics, these technologies can afford global insight into the active processes of the cell, without any loss of intrinsic complexity. The main purpose of the '-omics' technologies is the non-targeted characterization of all the genetic products (transcripts, proteins, and metabolites) present in a specific biological system (Fiehn, 2002; Weckwerth, 2003). This is useful when the introduction of new traits in a plant, perturbating system characteristics, changes the metabolite levels and the transformation routes in an unpredictable manner.

Metabolite profiling has been made possible via a number of instrumental techniques, including nuclear magnetic resonance (NMR) and mass spectrometry, which are the complementary techniques used in this field. Although less sensitive, NMR gives a direct fingerprint of the system, and the spectra obtained contain metabolic information that can be distilled by multivariate data analysis.

Most studies only address the classification aspect, using discriminant analysis to identify the metabolites that differentiate between groups of samples: wild-type versus mutant and transgenic plant, healthy versus infected plant. These studies can also characterize the maturation steps and be used to assess the impact of stress conditions on the plant metabolome (Krishnan *et al*., 2004). In particular, tomato (Le Gall *et al*., 2003), *Arabidopsis* (Ward *et al*., 2003), *Silene cucubalus* (Bailey *et al*., 2003), *Catharantus roseus* (Choi *et al*., 2004), and maize (Manetti *et al*., 2004) are examples of systems studied using this approach.

Accordingly, in this study, the effect of the introduction of an insect-resistance characteristic, i.e. the *Cry1Ab* gene, has been investigated in the genome of a maize inbred line in terms of metabolite concentrations, and the analysis extended to provide hypotheses on the pathways involved. In particular, NMR spectra are explored in terms of discriminant signals, using partial least square–discriminant analysis (PLS-DA) to obtain the first goal, and, in terms of correlation structure, using principal component analysis (PCA) for the exploration of the metabolic network. As a final step, assigned signal analysis (ASA) (Giuliani *et al*., 2004) was performed on the data to identify the invariant features of the two systems. The classification approach is used as a guideline to address the successive analysis, highlighting the key metabolites to the response of the system to the perturbation.

The aim of this article is to provide, from a new perspective, a more in-depth interpretation of the correlation between metabolites, inspecting recently published experimental data which integrate more theoretical approaches to the same matter (Steuer *et al*., 2003; Camacho *et al*., 2005).

Materials and methods

Plant material

The samples of maize seed used in this study were derived from the inbred line La73 as a control, and its transgenic version, i.e. La73-Bt, containing the *cry1Ab* gene (MON810) from *Bt* conferring resistance to the European corn borer. MON810 was developed and kindly provided by Mosanto Co. (St Louis, MO, USA). The MON810 insertion event was introduced to the La73 background by backcrossing six times, followed by two self-pollinations to obtain homozygous plants. The transgenic plants were selected following RT-PCR analysis Plants derived from the seeds of the two inbreds were grown under greenhouse conditions at 25/18 °C (day/night) with a 16/8 h (light/dark) cycle. At flowering, plants of each inbred line were self-pollinated; the ears were harvested after physiological maturity, dried at 30 °C, and stored in sealed plastic bags at 4 °C. For each sample a seed derived from the central portion of a single ear was used for the analyses. Eight samples for each maize genotype were analysed.

NMR methods

NMR sample preparation: For each sample, single maize seeds were weighted (~200 mg) and then frozen in a stainless-steel mortar by liquid N_2, before being pulverized to a fine powder with a pestle chilled in liquid N_2, and maintained in liquid N_2 bath during the pulverization procedure.

Three millilitres of methanol:chloroform mixture (2:1 v:v) were added to the powder. The powder was stirred and 1 ml of chloroform and 1.2 ml of water were added (Bligh-Dyer modified) (Miccheli *et al*., 1988; Ricciolini *et al*., 1994). The sample was stored at 4 °C for 1 h and then centrifuged at 10 000 *g* for 20 min at 4 °C. The resulting upper hydro-alcoholic and lower chloroformic phases were separated. The extraction procedure was performed twice on the pellet in order to obtain a quantitative extraction. After the second extraction, the two hydro-alcoholic phases obtained were recollected, dried under N_2 flux, and stored at −80 °C prior to analysis.

NMR data collection: For the NMR spectra, the dried sample was dissolved in 1 ml of 0.5 mM TSP solution in D_2O PBS buffer (pH=7.4) to avoid chemical-shift changes due to pH variation (Defernez and Colquhoun, 2003). The dissolved extracts were transferred to a 5 mm NMR tube.

NMR spectra were recorded on a Bruker (Bruker GmbH, Rheinstetten, Germany) DRX 500 spectrometer, operating at ^{1}H frequency of 500.13 MHz. ^{1}H NMR spectra were obtained at T=300 K, 256 scans were acquired, with data collected into 64 k data points, and a spectral width of 12 ppm, using a 20 s delay for a full relaxation condition. The water resonance was suppressed by irradiation during a 5 s relaxation delay at a power level 70 dB below maximal transmitter power setting (Rahman, 1989). Prior to Fourier transformation, an exponential multiplication was performed, using a line broadening equal to 0.09 Hz; this value represents an optimum balance between the noise reduction and the line-broadening effects, considering digital resolution. Spectra were referenced to TSP [sodium salt of 3-(trimethylsilyl)propionic-2,2,3,3-d_4 acid] at a final concentration of 0.5 mM. TSP was used as a reference both for chemical shift (0.00 ppm) and quantitation of the signals.

The processing of the spectra was carried out using ACD/SpecManager 7.00 software (Advanced Chemistry Development

Inc., 90 Adelaide Street West, Toronto, Ontario, Canada M5H 3V9). The spectra were phased, baseline corrected using the usual ACD routine selecting only two points located at the extremes of the spectra in the part that contains only noise. The spectra were scaled fixing the area of TSP signal to a value of 10.

NMR data pre-processing treatment: One-dimensional 500 MHz ^{1}H spectra were reduced to 499 discrete chemical shift regions by digitization to produce a matrix of sequentially integrated regions of 0.02 ppm in width between −0.5 ppm and 9.5 ppm, using ACD; column 1 corresponds to the bucket −0.5 ppm to −0.48 ppm. No region was excluded during the digitization to standardize the procedure. This choice makes this step unsupervised and avoids the necessity of changing the considered regions at the occurrence of new signals in unknown samples. Integrals corresponding to the water region randomly vary because of the presat effect, so they result in uncorrelated noise that the multivariate approach (PCA, PLS-DA) filtered. This characteristic is very useful when the purpose is the construction of a database.

Multivariate data analysis

Principal components analysis (PCA): This is a well-known multivariate technique, originally developed early last century (Spearman, 1904). This technique has had an almost universal application, ranging from hydrodynamics (Craddock, 1965; Preisendorfer, 1988; Ghil and Vautard, 1991) to sociological (Aitkin, 1974) and biological research (Gage *et al*., 1989; Giuliani *et al*., 1991).

The main purpose of PCA is to define the real dimensionality of the data field under study. When measuring N variables (NMR signals in the present case) on K units (samples in the present case), a situation is delineated which appears to be N-dimensional. However, these variables may be correlated in various ways among themselves, and so an equally satisfactory description could be obtained with a $P(P < N)$ number of axes, which are called factors or components and represent the degree of freedom of the system.

From a geometrical point of view, these dimensions (factors or components or latent variables) represent the directions in the data field along which the variability of the data clouds is maximal (Labart *et al*., 1984). From a mathematical point of view, components are eigenvectors of the correlation matrix among the original variables; they are orthogonal to each other and extracted by the algorithm in the order of percentage of explained variability. Thus, the first factor will be the one explaining the highest proportion of variation embedded in the original data matrix. Factors are constructed so as to have a mean value of zero and a unitary standard deviation over the entire data set.

Different variations of PCA can be performed by varying the nature of the data in X. X can be mean-centred or standardized (mean-centred and columns scaled to unit variance). One advantage of the first method is that the eigenvectors (or loadings) retain the scale of the original data, and will often resemble spectra. By contrast, the loadings obtained by standardized data are usually very unfamiliar in appearance. However, one advantage of this approach is that the PCA is influenced by all spectral features equally, whereas in the other approach, larger resonances tend to dominate. Consequently, the second data pre-processing method can be useful when minor constituents, with small spectral contributions are of primary interest (Belton *et al*., 1998). In the present case, this second data pre-processing method was applied to be sure that all the spectral regions make their contribution to the discrimination between the different samples.

Partial least square–discriminant analysis (PLS-DA): The PLS technique was originated by Wold (1966) for the modelling of complicated data sets in terms of chains of matrices, the so-called path models. After this first use, PLS was applied to spectrometric calibration (Haaland and Thomas, 1988) to monitor and control industrial processes (Wang *et al*., 2003) and, in recent years, to metabonomics (Lindon *et al*., 2001; Brindle *et al*., 2002).

PLS is a method for constructing predictive models when the factors are many and highly collinear. It maximizes the covariance between the predictor space (matrix of NMR data, X) and the response space (matrix of the information on maize lines to which the seeds belong, Y). The overall goal is to use the factors to predict the responses in the population. This is achieved indirectly by extracting latent variables T (X-scores) and U (Y-scores) from sampled factors and responses, respectively. The extracted factors T are used to predict the U, and then the predicted Y-scores are used to construct predictions for the responses (Randall, 1995). In other words, the data set is interpreted in terms of X- and Y-scores (T, U), X-loading (P), X-and Y-weights (w, c), and PLS regression coefficients (B) (Wold *et al*., 2001). Once a PLS model has been calculated and validated, it can be used for the prediction of class membership of unknown samples.

A particular version of this method (reported as PLS-DA) is done by a regression of the data (X) against a 'dummy matrix' (Y), which describes variation according to class. In the training set, the input Y 'dummy' matrix has a row, for each sample, containing 1 for the y variable corresponding to the right group and 0 for all the others.

To explore further the results obtained by PLS-DA, another statistic summarizing the contribution a variable makes to the model: the variable importance for projection (VIP) of Wold (1994), has been used. The VIP represents the value of each predictor in fitting the PLS model for both predictors and response. Wold considers a value <0.8 to be 'small' for the VIP.

VAST scaling procedure: As an alternative to the traditional scaling procedure, VAST (VAriable STability) was used (Keun *et al*., 2003). It can be considered as the sequential application of mean-centring and unitary variance scaling, initially to put each variable on the same level, and then scaling by the coefficient of variation, to incorporate stability. This stability parameter is equal to the ratio of the standard deviation and the mean of each variable calculated on the uncentred data set. The VAST procedure can also be applied in a supervised manner, in that the coefficient of variation within each prior class can be calculated separately, and then the mean of the class coefficients of variation used as the stability scale weight. This last version of the VAST procedure was the one used in this work. This method, however, has been used only in the PLS-DA procedure, so that no modifications are introduced to the PCA standard procedure. This could involve changes in the space where data are represented, changes that are not simply rationable.

Analysis of variance (ANOVA): This is a technique for analysing experimental data in which one or more response (or dependent) variables are measured under various conditions identified by one or more classification variables. The variation in the response is separated into variation attributable to differences between the classification variables and variation attributable to random errors. An analysis of variance constructs tests to determine the significance of the classification effects. A typical goal is to compare means of the response variables for various combinations of the classification variables. ANOVA was applied, considering the type of seed (control or transgenic) as the classification variable, and the concentrations of the metabolites as a dependent variable.

SAS (Statistical Advanced Software) software, version 8 (SAS Institute Inc., www.sas.com) was used for all the statistical analysis.

Results

^{1}H-NMR assignment of extracts of Zea mays *seeds*

The assignment of maize extract spectra was obtained using a combination of two-dimensional NMR experiments,

comparison with chemical shift values reported in literature [Fan, 1996; Le Gall *et al*., 2003; Spectral Database for Organic Compounds SDBSWeb, http://www.aist.go.jp/RIODB/SDBS/ (National Institute of Advanced Industrial Science and Technology)], and spiking of extracts with test compounds. Chemical shifts of compounds that were identified are listed in Table 1.

Classification

NMR spectra analysis of control (C) and modified (M) kernel samples has been made reducing the spectra in two different ways. In the first, the data-set variables were the integrals of 499 discrete regions (buckets) of the unassigned spectra; these data sets were identified by the name ent-data set (i.e. ent-C, ent-M, ent-C+M). In the second kind of analysis, the data sets contained integrals of selected assigned signals. These were named ASA-data set (i.e. ASA-C, ASA-M, ASA-C+M).

The choice of using selected signal integrals as variables corresponds to obtaining, for the single statistical unit (seed), a series of values relative to the different metabolite concentrations that could be obtained by different experimental techniques. Metabolites that can be evaluated by the present protocol are those that are soluble in the mixture of solvents used and whose concentrations are high enough to be revealed by NMR. PCA on these data sets allows metabolic information to be obtained that can be overlapped with that obtained on complete bucketed spectra, recognizing in this way the same latent variables (metabolic route) or giving further details about the metabolic network. This approach is called ASA.

Initially, to classify, PLS-DA was applied to the entire bucketed spectra of the hydro-alcoholic extracts of modified and unmodified seeds (ent-C+M); a PLS-DA model was obtained with one latent variable, explaining 44% of the model effect and 39% of dependent variable.

In Fig. 1, the PLS-DA score plot is represented. It is clearly evident that the first latent variable manages to successfully discriminate between the two kinds of maize seeds; in fact, all the modified samples have positive score values, while all the controls have negative score values. Furthermore, PCA was also used to group the data set ent-C+M in the expected classes; a six-component model was obtained, explaining 90% of the variance (for details see Table 2).

In Fig. 2, the score plot of PC (principal component) 1 versus PC2 is reported. A good classification was also obtained by PCA of data set ASA-C+M, which gave a three-component model, explaining 70% of the variance (for details, see Table 2).

In Fig. 3, the plot of the hydro-alcoholic extracts of maize seeds in the ASA space spanned by the score of the first and the second PCs is shown. The first PC manages to discriminate successfully between the two types

Table 1. *^1H chemical shifts of compounds from 1-H spectra of hydro-alcoholic extracts of maize seeds*

Signal	Multiplicity[a]	δ (ppm)
TSP	s	0.00
Ile (δCH_3)	t	0.93
Leu (δCH_3)	d	0.97
Leu ($\delta' CH_3$)	d	0.95
Val (CH_3)	d	0.99
Ile (γCH_3)	d	1.01
Val (CH_3')	d	1.04
Thr (γCH_3)	d	1.32
Ala (βCH_3)	d	1.48
Leu (βCH_2+γCH)	m	1.71
GABA (βCH_2)	q	1.89
Acetate (CH_3)	s	1.92
Pro (γCH_2)	m	2.01
N-acetyl CH_3	s	2.01
Pro (β'CH)	m	2.08
Glu (βCH_2)	m	2.09
Gln (βCH_2)	m	2.15
GABA (αCH_2)	t	2.31
Glu (γCH_2)	m	2.34
Pro (βCH)	m	2.35
Pyruvate (CH_2)	s	2.39
Succinate (α-βCH_2)	s	2.41
Gln (γCH_2)	m	2.46
Dimethylamine (CH_3)	s	2.73
Asn (βCH_2)	dd	2.87
Asn (βCH_2)	dd	2.95
GABA (γCH_2)	t	3.02
Choline [$N(CH_3)_3$]	s	3.21
β-Glucose (C2H)	t	3.25
α-Glucose (C4H)	dd	3.42
β-Glucose (C4H)	dd	3.4
Sucrose (G4H)	t	3.48
Sucrose (G2H)	dd	3.57
Sucrose (F1H)	s	3.68
Sucrose (G3H)	t	3.77
Sucrose (G5H+G6H)	c	3.82
Sucrose (F6H)	c	3.84
Sucrose (F5H)	m	3.90
Asn (αCH)	dd	4.00
Pro (αCH)	t	4.14
Sucrose (F3H)	d	4.22
Malate (αCH)	dd	4.29
Trigonelline (NCH_3)	s	4.5
β-Galactose (C1H)	d	4.60
β-Glucose (C1H)	d	4.65
Melibiose (Gal1H)	d	5.00
Melibiose (αG1H)	d	5.2
α-Glucose (C1H)	d	5.24
Sucrose (G1H)	d	5.42
Ferulic acid (HF)		6.81
Tyr (C3, 5H, ring)	d	6.9
His (C4H, ring)	s	7.09
Trp (C5H, ring)	t	7.15
Tyr (C2, H6, ring)	d	7.2
Trp (C6H, ring)	t	7.29
Phe	m	7.33
Phe	m	7.38
Phe	m	7.43
Trp (C7H, ring)	d	7.54
His (C2H, ring)	s	7.84
Trigonelline (HD)		8.08
Formate (CH)	s	8.46
Trigonelline (HB, HC)		8.84
Trigonelline (HA)		9.13

[a] s, singlet; d, doublet; dd, doublet of doublets; t, triplet; q, quartet; m, multiplet; c, complex.

of seeds; in fact, all the modified samples have positive score values, while all the controls have negative score values.

Bridge between classification and metabolic insight

In the next step, the most important metabolites for discrimination, i.e. variable importance for projection (VIP) values (Wold, 1994), were calculated from PLS-DA results to select among the 499 variables (buckets). This corresponds to the determination of the most suitable point from which to study the structure of an extremely big correlation matrix.

Figure 4 gives a representative number of buckets with their assignments and their VIP values. In this step, the buckets were assigned, and it was verified if all the metabolites corresponding to the signals selected on the entire bucketed spectra were present in the ASA data set. The correlation structure allows the right assignment of the different signals of the same metabolite to be checked. A test ANOVA was performed to identify the metabolites whose concentrations are significantly different between modified and control seeds (Table 3).

Metabolic insight

In a further analysis, correlations among the VIP metabolites were identified to highlight the metabolic pathway involved. In this analysis, data relative to modified and control samples are considered separately (data sets ent-C, ent-M, ASA-C, and ASA-M). The analysis of PCA loadings of entire spectra, focused on the metabolites selected on a VIP basis, highlighted important variations induced on the metabolic network, and was confirmed and integrated by the metabolic information obtained by PCA on assigned signals.

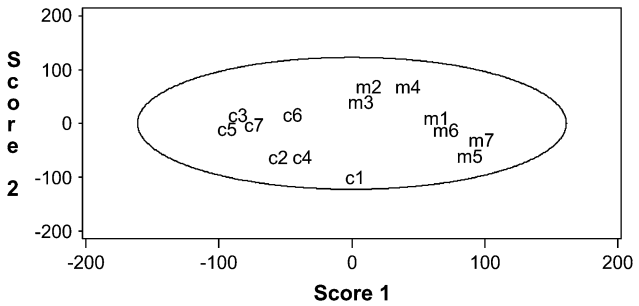

Fig. 1. Plot of the hydro-alcoholic extracts of maize seeds in the PLS-DA space spanned by the score of the first and the second latent variables. The ellipse represents the Hotelling T^2 with 95% confidence.

For the data set ent-C, a three-component model was obtained; this accounted for 85% of the variance, while for ent-M, a four-component model was obtained, explaining 90% of the variance (for details, see Table 2). A description of the results can start by comparing the two PCs at a higher correlation. The first PCs were excluded to avoid misinterpretation due to size effects. In particular, the comparison was performed calculating the correlation coefficient among the loadings obtained by PCA on the data sets ent-C, ent-M, and ent-C+M (Table 4). This choice allows the reference systems obtained by the PCs calculated on the two data sets to be compared (i.e. orientate one with respect to the other). This approach, one of the possible ways of treating the data, is a further way to observe multivariate analysis results easily and to identify correlations among metabolites. In fact, the correlation structure (the direction of PC axes) is determined by all the 499 variables, and the present aim is to understand where the VIP metabolites and their correlations are located in this reference system.

All the loading values of the VIP metabolites relative to the second PCs obtained on data sets ent-C and ent-M are reported in Fig. 5, even if only a few of them have a loading value >0.5 in absolute value. It is worth noting that Phe, Trp, Tyr, and His, in the modified samples, have a negative correlation with the other selected metabolites, while they always have a positive correlation in the control samples. Furthermore, in the control samples, Ile, Leu, and Val also have high loading values. In the modified samples, instead, these metabolites have a different position, and are identified on a different component (Fig. 6), in the reference system to which choline and Asn are the only other contributing metabolites. Therefore, metabolic

Table 2. *Variances explained (%) by each PCA and ASA model, respectively, and the model obtained on the entire bucketed spectra*

For both models the data were divided in three sets: the first contained all the samples (control+modified, C+M), the second contained only the control (C), and the third only the modified (M).

Principal component	Variance explained (%)					
	Entire bucketed spectra			Selected signals		
	ent-C+M	ent-C	ent-M	ASA-C+M	ASA-C	ASA-M
1	34.98	49.36	36.04	39.33	38.11	38.96
2	26.97	24.98	31.39	17.61	22.32	29.56
3	13.37	11.39	14.71	14.10	17.83	12.18
4	6.98	7.86	7.04	8.08	9.18	8.63
5	4.39	4.23	6.18	6.67	7.58	6.38
6	2.76	2.17	4.64	–	–	–

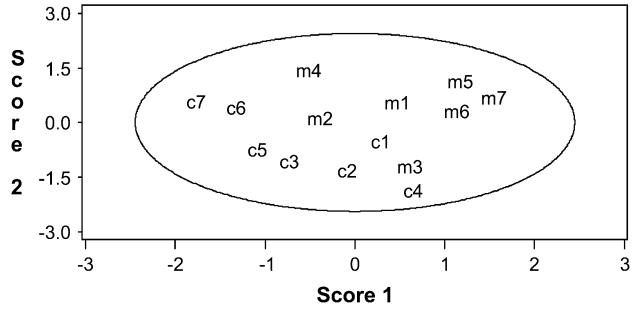

Fig. 2. Plot of the hydro-alcoholic extracts of maize seeds in the PCA space spanned by the score of the first and the second principal component. The ellipse represents the Hotelling T^2 with 95% confidence.

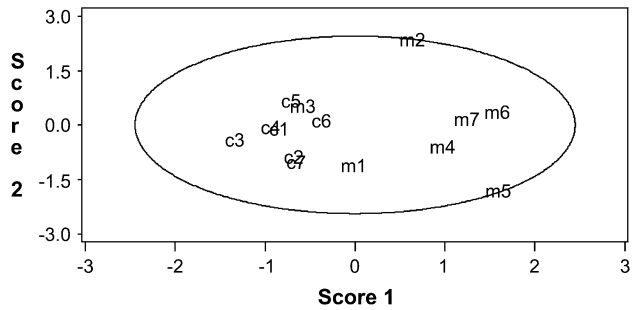

Fig. 3. Plot of the hydro-alcoholic extracts of maize seeds in the ASA space spanned by the score of the first and the second principal component. The ellipse represents the Hotelling T^2 with 95% confidence.

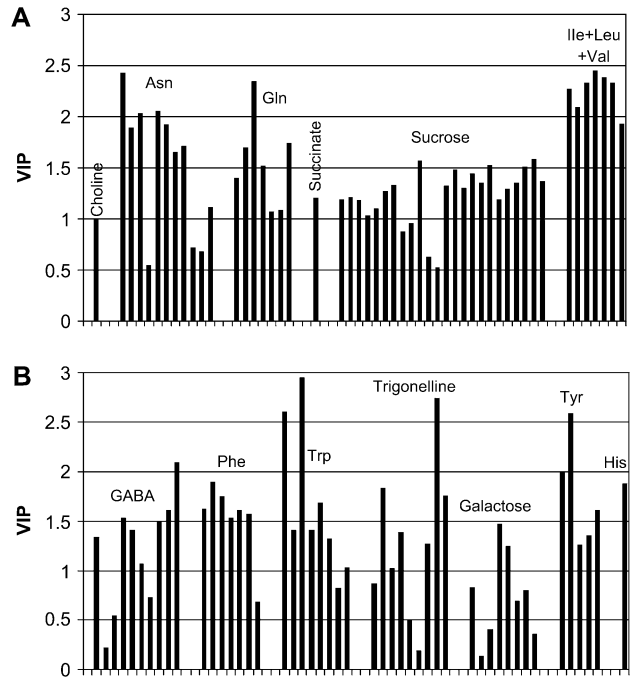

Fig. 4. Plot of the VIP (variable importance for projection) values for all the buckets of the most relevant metabolites.

Table 3. *ANOVA results for selected signals from control (c) and transgenic (t) maize seed spectra*

Metabolite (signal)	*F*-value[a]	Order
Acetate (βCH_3)	0.3	
Ala (βCH_3)	14.6**	t <c
α-Glucose (C1H)	12.6**	t >c
Asn (βCH_2)	18.6***	t <c
β-Glucose (C1H)	17.2**	t >c
Choline [$N(CH_3)_3$]	105.6***	t <c
Dimethylamine (CH_3)	4.0	
Ferulic acid (HF)	4.5	
Formate (CH)	0.2	
GABA (αCH_2)	28.8***	t >c
Gln (βCH_2)	18.5***	t >c
Glu (γCH_2)	1.5	
His (C2H, ring)	9.2**	t <c
Ile (γCH_3)	2.4	
Melibiose (Gal1H)	6.6*	t >c
Pyruvate (CH_2)	3.4	
Succinate (α-βCH_2)	44.5***	t >c
Sucrose (F1H)	7.1*	t >c
Thr (γCH_3)	0.3	
Trigonelline (HA)	0.3	
Tyr (C2, H6, ring)	1.9	
Val (CH_3')	0.5	

[a]*,**,***, Significant at the .05, 0.01, and 0.001 probability levels, respectively.

Table 4. *Correlation values for all the possible couples of principal components obtained on the complete bucketed spectra*

C 1						
−0.6610	C 2					
−0.0255	−0.0210	C 3				
0.0492	0.0438	−0.2368	M 1			
0.7653	−0.5826	0.3046	−0.2590	M 2		
−0.4968	0.5463	0.1799	−0.2444	−0.3147	M 3	
−0.1135	0.0294	−0.0647	−0.0358	−0.0461	−0.0435	M 4

differences were detected correlated to choline, Asn, Ile, Leu, Val, and GABA (γ-aminobutyric acid). This type of correlative analysis does not give straightforward elements to determine correlations among the non-VIP metabolites, correlations that PLS-DA show as important for discrimination.

In Fig. 6, the loading relative to the third component of data set M is given. It is worthwhile to note that the branched amino acids, choline and Asn, are positively correlated to this component. In Fig. 7, the loading relative to the third component of data set C is shown. It is interesting to note that choline, Asn, Gln, sucrose, and trigonelline are negatively correlated to this component. In Fig. 8, the loading relative to the fourth component of data set M is represented. Interestingly, Asn, Gln, and trigonelline are positively correlated to this component, while succinate is negatively correlated to it. Moreover, it is worth noting that the branched amino acids are important for

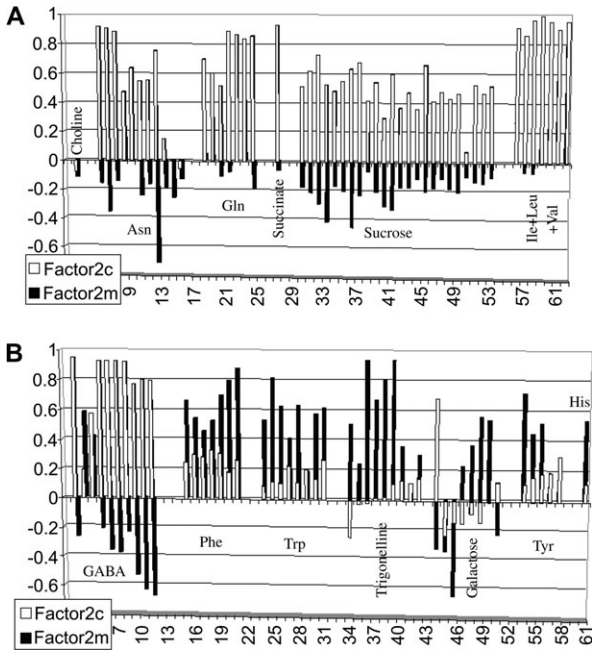

Fig. 5. Plot of loading values of the signals, corresponding to the most important metabolites, in the second principal components of PCA, constructed considering the data sets C (open columns) and M (filled columns). Phe, Trp, Tyr, and His, in the modified samples, have a negative correlation with the other selected metabolites, while they always have a positive correlation in the control samples. Furthermore, in the control samples, Ile, Leu, and Val also have high loading values.

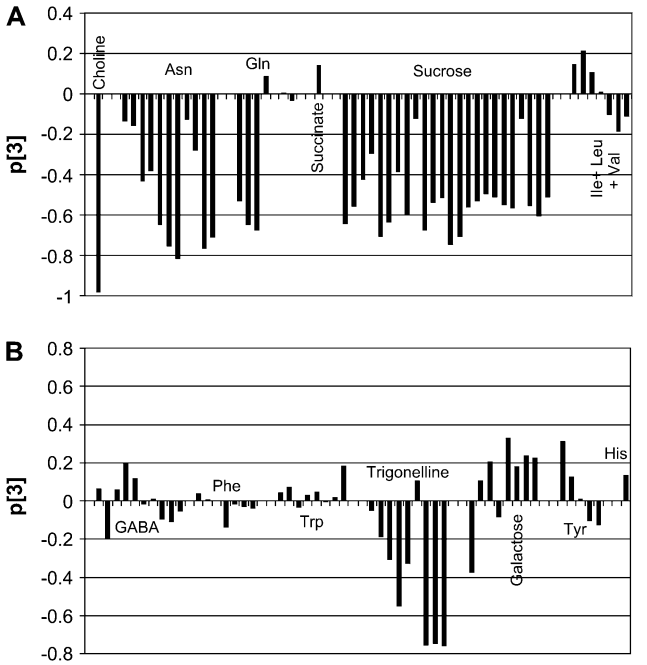

Fig. 7. Plot of loading values of the signals, corresponding to the most important metabolites, in the third principal component of PCA, constructed considering the C data sets. Choline, Asn, Gln, sucrose, and trigonelline are negatively correlated to this component.

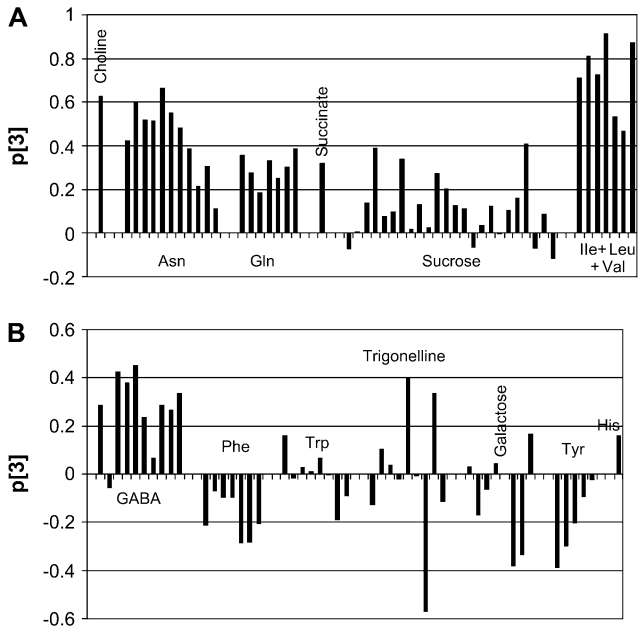

Fig. 6. Plot of loading values of the signals, corresponding to the most important metabolites, in the third principal component of PCA, constructed considering the M data sets. The branched amino acids, choline, and Asn are positively correlated to this component.

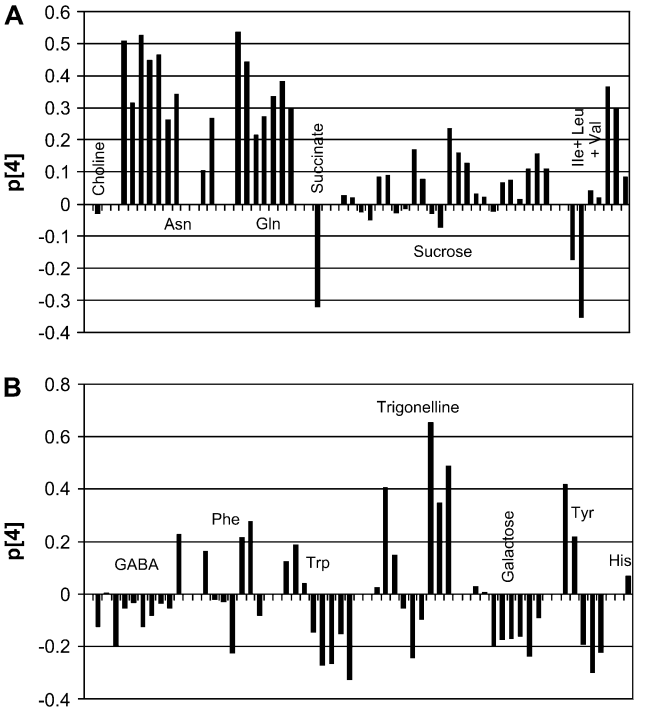

Fig. 8. Plot of loading values of the signals, corresponding to the most important metabolites, in the fourth principal component of PCA, constructed considering the C data sets. Asn, Gln, and trigonelline are positively correlated to this component, while succinate is negatively correlated to it.

the second component of data set C, and for the third component of M, whereas the other metabolites that are important for the second component of data set C are not significant for the third component of M. For this reason, it seems that, in data set M, branched amino acids are related to a metabolic variation, and they can constitute the hinge between two metabolic pathways.

For data set ASA-C, a four-component model is obtained, explaining 90% of the variance, while for data set ASA-M, a four-component model is also obtained, explaining 90% of the variance (for details, see Table 2).

In Table 5, the loading values obtained by performing PCA on data sets ASA-C and ASA-M are given. In the first PC of the control samples, signs are found of the correlation linking the metabolites involved in sugar production (in particular sucrose), the tricarboxylic acid (TCA) cycle, and primary nitrogen metabolism. Furthermore, in the first PC of the modified ones, many correlations are confirmed, even if variations concerning some metabolites, such as His, Thr, and Ile, are present. These metabolites do not correlate with the first PC, but with the second one.

The analysis of the first PC of the control samples becomes more informative in comparison with the third one obtained by the modified samples, which is characterized by the correlation with Tyr, Asn, Glu, and Ile. It is worth mentioning the correlation of the third PC of the control samples with trigonelline, choline, and Asn. This variation suggests important metabolic changes that can be represented in a graph, where the correlations are highlighted and they show the metabolic 'routes' (metabolic network).

Discussion

The present results highlight significant variations in the metabolism of maize seeds containing the MON810 event. In particular, looking at Fig. 4 which summarizes the most important metabolites for discrimination, it is interesting to note that Asn assumes great importance. Asn and Gln are the major forms of reduced nitrogen either transported in the developing seed or recovered from endosperm tissues in plants, and act as precursors for the amino acids found in storage proteins (Lea and Miflin, 1980; Sieciechowicz *et al*., 1988; Rabe, 1990). An alteration of its level in the seed may result from a perturbation of asparagine metabolism and/or amino acid metabolism. In PCA results (Fig. 5), in fact, signs can be found of the metabolic relationship between Asn and Gln due to the transformation:

$$\text{glutamine} + \text{aspartate} + \text{ATP} \rightarrow$$
$$\text{glutamate} + \text{asparagine} + \text{ADP} + \text{PPi}$$

Accordingly, the metabolic correlation among metabolites is direct; it is not always that the correlation among variables can be explained in such a direct way (product–parent compound). In fact, the metabolites involved in other pathways can guarantee the correlation. In any case, metabolic connections determine the correlation between the variables examined. For these reasons, the perturbation of the system can be detected by observing that the factor loading of some metabolites have significantly different values, i.e. they do not correlate with the same component in the control sample and in the modified one. This variation

Table 5. *Loading values obtained by ASA on the data set containing only the control seeds (C) and only the modified seeds (M)*

Metabolite (signal)	Factor 1 C	Factor 2 C	Factor 3 C	Factor 1 M	Factor 2 M	Factor 3 M
Trigonelline (HA)	−0.4059	0.4488	**0.6131**	−0.0211	**0.8968**	0.0767
Formate (CH)	**−0.7085**[a]	0.4748	0.1897	−0.3541	**0.8024**	−0.2259
His (C2H, ring)	**−0.6559**	0.4836	0.3086	**−0.8103**	0.5306	0.0386
Tyr (C2, H6, ring)	0.4759	**0.7838**	−0.3306	0.2961	**0.6252**	**−0.6327**
Ferulic acid (HF)	0.1605	**0.7360**	−0.4461	−0.1111	**0.6051**	−0.2999
α-Glucose (C1H)	−0.5789	**0.6880**	0.1446	**0.9176**	0.0540	−0.2836
Melibiose (Gal1H)	0.3554	0.0424	**0.8982**	**0.6834**	−0.0892	0.5294
β-Glucose (C1H)	−0.3596	**0.7582**	0.1662	**0.9249**	−0.0547	−0.3613
Sucrose (F1H)	**0.6244**	−0.5432	0.3352	**0.9039**	−0.0587	−0.1649
Choline [$N(CH_3)_3$]	0.3726	−0.2002	**0.7868**	0.5597	−0.5236	0.5009
Asn (αCH_2)	0.1820	0.1647	**0.6223**	0.5635	−0.1280	**0.6185**
Dimethylamine (CH_3)	0.4509	**0.7640**	−0.4236	0.3423	**0.6373**	0.1029
Succinate (α-βCH_2)	**0.8434**	0.3043	−0.3366	**0.8265**	0.1532	0.1476
Pyruvate (CH_2)	**0.6974**	0.3234	−0.4760	**0.7302**	**0.6603**	−0.1370
GABA (αCH_2)	**0.9484**	0.2916	−0.0234	**0.5870**	0.2165	0.0047
Gln (βCH_2)	**0.8255**	0.2685	0.3952	−0.2600	0.5003	**0.6106**
Glu (γCH_2)	**0.8322**	0.0561	0.3510	**0.8943**	0.3103	−0.2312
Acetate (βCH_3)	−0.4519	−0.5783	−0.2529	**−0.8683**	−0.2423	0.2236
Ala (βCH_3)	**0.7808**	0.2815	−0.0336	**0.9102**	−0.0929	0.2250
Thr (γCH_3)	**0.6039**	0.3586	0.1145	−0.3750	**0.7657**	−0.0805
Ile ($\gamma\gamma CH_3$)	**0.8633**	−0.2550	−0.2171	−0.0599	**0.7087**	**0.6068**
Val (CH_3')	**0.7626**	−0.4772	−0.1385	**0.5925**	0.5231	0.4362

[a] Values >0.6 are presented in bold. They indicate the metabolites important for the definition of the PC.

can be 'concerted', underlying the invariant characteristics of the relationship between metabolites that is resistant to the perturbation; groups of metabolites keep their correlation with the same component, while others do not keep it. Furthermore, the signature of the perturbation can be revealed by a change of the sign of factor loading; metabolites positively correlated in the control samples have loading with the opposite sign to the modified ones.

Analysis revealed that metabolites which exhibited a close correlation in a component of data set ASA-C, in data set ASA-M they are on a different PC. Migration from one PC to another involves many metabolites. Thus, for this reason it can be affirmed that correlation among these metabolites exists and is a system 'invariant'. These changes are the sign of the metabolic change due to the modification.

Analysing the first component, the invariant core represents the metabolic relationship between sugar production (in particular sucrose), the TCA cycle, and primary nitrogen metabolism. The fact that the correlations of these variables with the first PC are maintained in data sets C and M means that the metabolic network, based on correlations, represents a fundamental metabolic process for plant life, a process that remains unchanged in its relationships (invariance). Analysing data set M, it was noted that Gln is not correlated with the first PC, but it is correlated with the third, with Asn, Ile, and Tyr. Linking ASA results with the ones obtained by ANOVA, changes in the metabolic network of the modified seeds can be described.

In Figs 9–12, the metabolic pathways involved are depicted. In particular, in Figs 10–12, the invariant

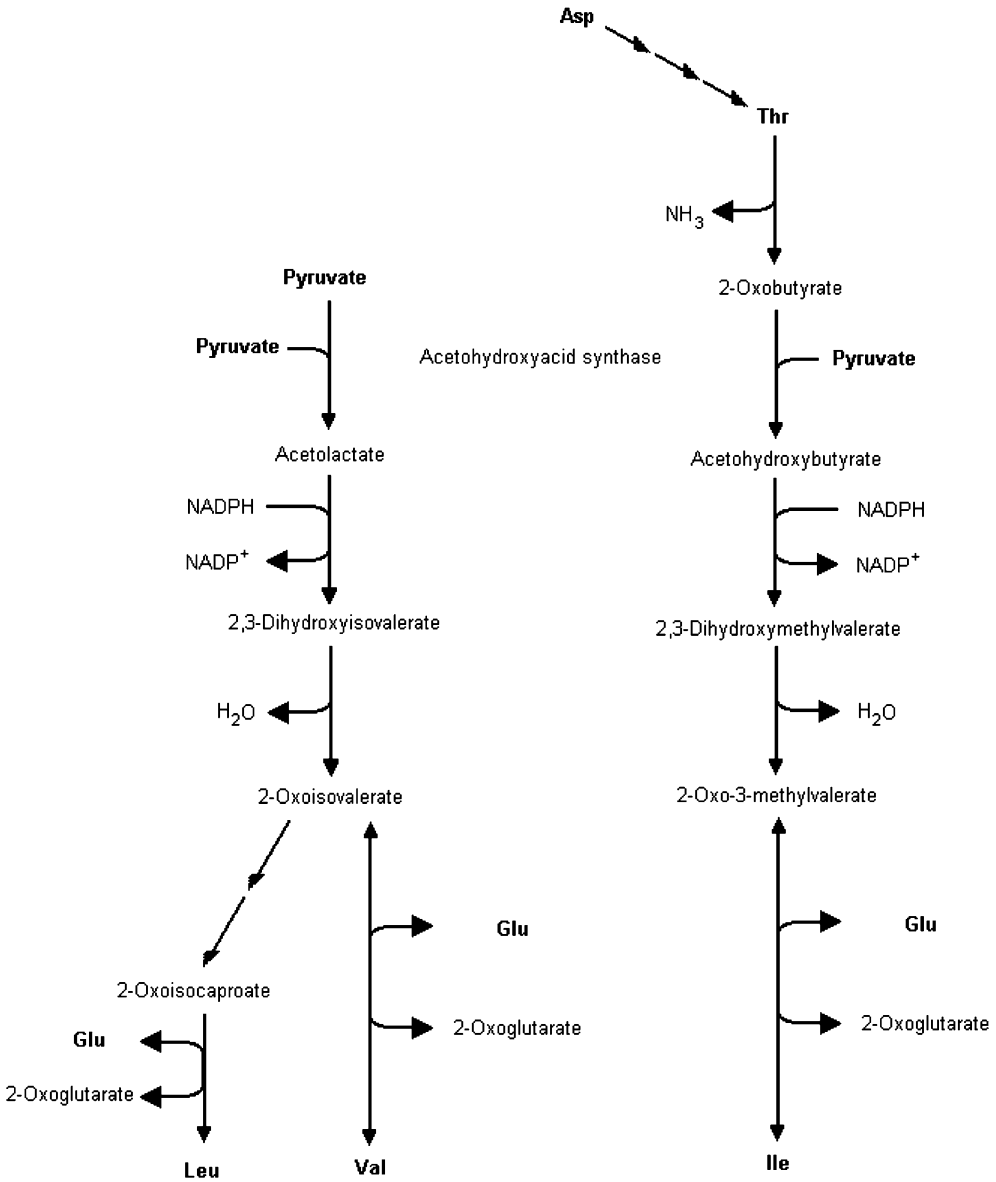

Fig. 9. A schematic representation of the synthesis of the branched chain amino acids.

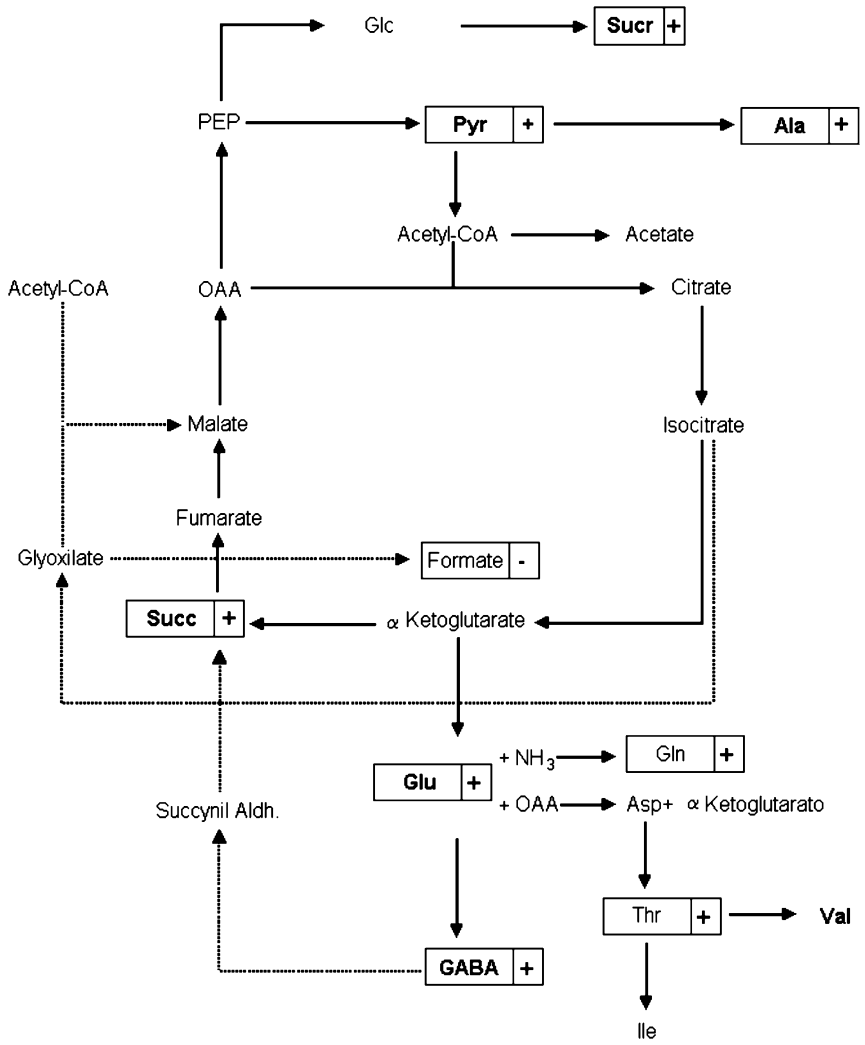

Fig. 10. A schematic representation of the pathway correlated to the first component, obtained by ASA on data set C. The invariant metabolites are indicated in bold and metabolites with high loading values (>0.6) have been boxed. Metabolites that are clearly involved in the metabolic pathway, but that cannot be revealed by NMR, are also indicated.

metabolites have been indicated in bold, and metabolites with high loading values (>0.6) have been boxed. It is worth noting that the correlation sign is always positive except for formate in the first PC of data set C and acetate in the first PC of data set M. The figures also contain indications of metabolites that are clearly involved in the metabolic pathway, but that cannot be revealed by NMR; this may be attributable to the sample preparation method for their low concentrations or to high overlapping of their signals.

Concerning the branched amino acids, a specific metabolic correlation on the third PC of data set M was detected; this result highlights a metabolic change due to the genetic modification. The pathways leading to the synthesis of these metabolites are considered to be biochemically parallel, being catalysed by enzymes with dual substrate specificities (Wallsgrove, 1990). The control of the pathway is achieved by feed-back inhibition of aceto-hydroxyacid synthase by leucine and valine (Fig. 9).

As shown in Figs 10 and 11, there is a metabolic modulation, with an increase in GABA, succinate, Glc, and Gln levels, and a decrease in Ala levels (as indicated by the arrows). GABA accumulation is in agreement with a reduction in GABA shunt, the conversion of Glu to succinate via the action of glutamate decarboxylase, GABA transaminase, and succinic semialdehyde dehydrogenase (Vandewalle and Olsson, 1983; Breitkreuz and Shelp, 1995). This shunt affords an alternative pathway for glutamate entry in the TCA cycle. It is worth noting that GABA also has a high VIP value, and succinate and Ala concentrations are significantly different (see ANOVA).

The metabolic pathway connecting Thr and Ile, indicated in Fig. 11, is the element that permits these analysis results to be overlapped with the one obtained by PCA on the complete bucketed spectra and shown in Fig. 9. Gln has no

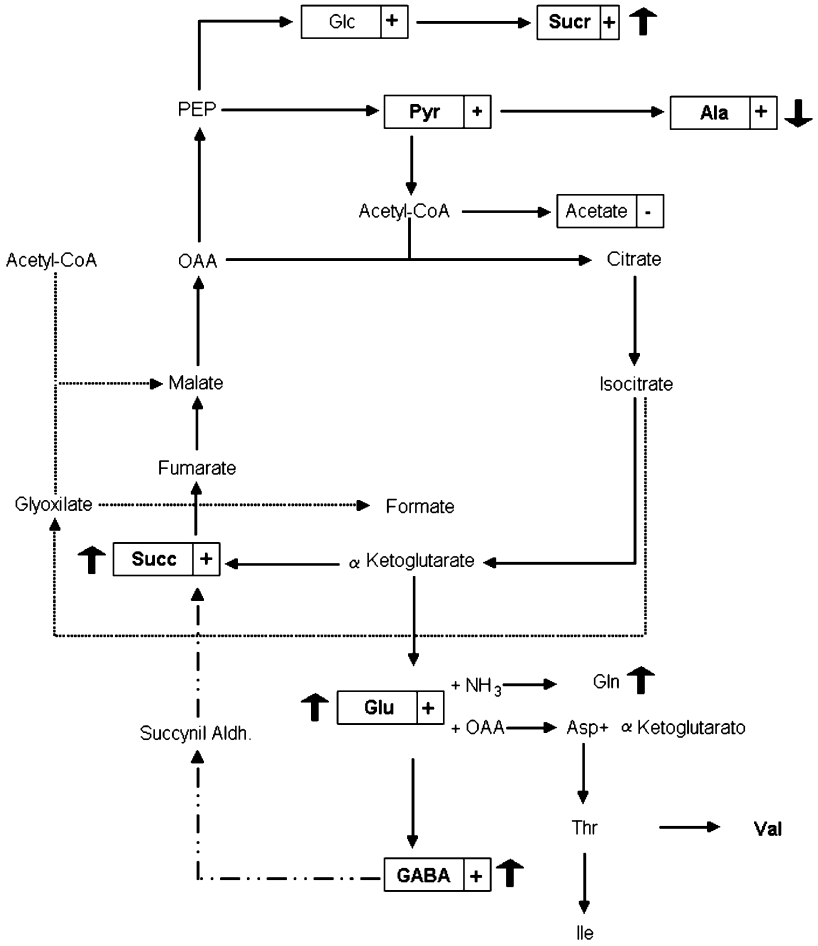

Fig. 11. A schematic representation of the pathway correlated to the first component, obtained by ASA on data set M. The arrows suggest an increase in GABA, succinate, Glc, and Gln levels and a decrease in Ala levels.

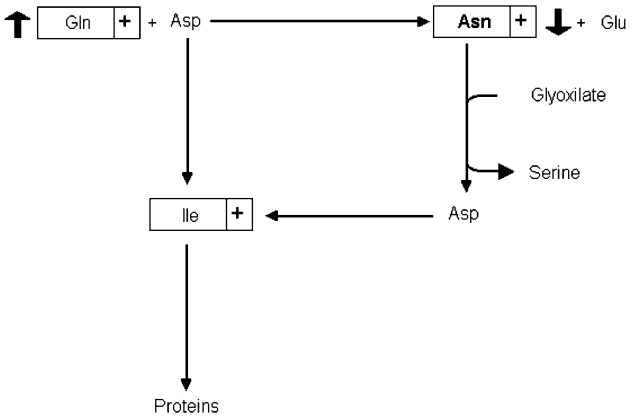

Fig. 12. A schematic representation of the pathway correlated to the third component, obtained by ASA on data set M.

high correlation (>0.6) with the first PC of data set M, while this correlation has been highlighted in the first PC of data set C. Therefore, the metabolic relationship is modified and Gln identifies a third PC with Asn and Ile. This component is correlated with the metabolic pathway shown in Fig. 12.

In this context, the GABA shunt has been recently associated in plants with various physiological responses, including carbon fluxes into the TCA cycle, nitrogen metabolism, osmoregulation, and signalling (Bouché and Fromm, 2004).

Metabolites of primary nitrogen metabolism are identified as the discriminant ones, in particular, isoleucine, leucine, and valine. Buckets relative to metabolites established to play important roles in osmotic adjustment and/or osmoprotection in a number of species, including higher plants (for a review of the roles of betaines and their sulphonio analogues as compatibles solutes and in cell volume regulation, see Yancey, 1994). Yancey (1994) indicates that these solutes are excluded from the hydration sphere and tend to stabilize the tertiary structure of proteins. In addition, they prevent or reverse the disruption of the tertiary structure caused by non-compatible (perturbing) solutes. It is worth noting that a high concentration of glycinebetaine is present in the seeds of the maize genotype La73. In this respect, Saneoka *et al.* (1995), in studying salt tolerance of glycinebetaine-deficient and -containing maize lines, have found that a single gene

conferring glycinebetaine accumulation (and/or a tightly linked locus) plays a key role in osmotic adjustment in maize leaves. Further research is needed to clarify if the alteration in these solutes in the seeds is also related to similar perturbation in green tissues.

Many recently deposited patents, concerning transgenic plants that exhibit improved agronomic and nutritional trials, are based on metabolic mechanisms that involve nitrogen assimilation, transport, and utilization, deliberately altering expression of key enzymes (as an example, see G Coruzzi and T Brears United States Patent 6864405, http://www.patentstorm.us/patents/6864405.html). In these cases, the response of the system is easy to verify searching for specific metabolite levels. In the present study, instead, it was possible to verify that the metabolomic approach, non-targeted, permitted unpredictable but reasonable changes to be highlighted; the genetic modification introduced to obtain the novel protein *Cry1Ab* also involved changes in nitrogen metabolism—the system reorganizes itself to produce a nitrogen-rich compound.

Conclusions

In conclusion, this work demonstrates that the metabonomic approach, constituted by PLS-DA and PCA steps, is a convenient way of immediately recognizing the discriminant pathways between genetically modified plants and their control. The results obtained can offer metabolite information which is very useful in the analysis of the new status of plants, looking not at mean values of the different components, but at the covariance matrix of the level of different metabolites—systemic information. Furthermore, this work raises novel points, both on methodology and on analysis of metabolic variations, as consequences of genetic modifications.

In the method used, a very important characteristic is the non-targeted capability that is very useful in the analysis of unpredictable metabolic variations and the identification of metabolites characteristic of the transformation. Moreover, the possibility of obtaining an improvement on the multivariate analysis by a different data normalization procedure (VAST) is highlighted. PCA on integrals of assigned signals of selected metabolites gives information complementary to the one obtained by PLS-DA and PCA on the complete bucketed spectra. Physiological information emerges after PCA and reveals that the genetic modification, in which the expression of Cry1Ab protein is introduced, induces metabolic variations involving osmolytes and branched amino acids.

Acknowledgement

The authors would like to thank Mr Danny Singh for his linguistic revision.

References

Aitkin R. 1974. *Mathematical structure in human affairs*. London: Heinemann.

Bailey NJC, Oven M, Holmes E, Nicholson JK, Zenk MH. 2003. Metabolomic analysis of the consequences of cadmium exposure in *Silene cucubalus* cell cultures via ^{1}H NMR spectroscopy and chemometrics. *Phytochemistry* **62,** 851–858.

Belton PS, Colquhoun IJ, Kemsley EK, Delgadillo I, Roma P, Dennis MJ, Sharman M, Holmes E, Nicholson JK, Spraul M. 1998. Application of chemometrics to the ^{1}H NMR spectra of apple juices: discrimination between apple varieties. *Food Chemistry* **61,** 207–213.

Betz FS, Hammond BG, Fuchs RL. 2000. Safety and advantages of *Bacillus thuringiensis*- protected plants to control insect pests. *Regulatory Toxicology* **32,** 156–173.

Bouché N, Fromm H. 2004. GABA in plants: just a metabolite? *Trends in Plant Science* **9,** 110–115.

Breitkreuz KE, Shelp BJ. 1995. Subcellular compartmentation of the 4-aminobutyrate shunt in protoplasts from developing soybean cotyledons. *Plant Physiology* **108,** 99–103.

Brindle JT, Antti H, Holmes E, *et al.* 2002. Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using ^{1}H-NMR-based metabonomics. *Nature Medicine* **8,** 1439–1444.

Camacho D, de la Fuente A, Mendes P. 2005. The origin of correlations in metabolomics data. *Metabolomics* **1,** 53–63.

Chilcutt CF, Tabashnik BE. 2004. Contamination of refuges by *Bacillus thuringiensis* toxin genes from transgenic maize. *Proceedings of the National Academy of Sciences, USA* **101,** 7526–7529.

Choi YH, Tapias EC, Kim HK, Lefeber AWM, Erkelens C, Verhoeven JTJ, Brzin J, Zel J, Verpoorte R. 2004. Metabolic discrimination of *Catharanthus roseus* leaves infected by phytoplasma using ^{1}H-NMR spectroscopy and multivariate data analysis. *Plant Physiology* **135,** 2398–2410.

Craddack JM. 1965. A meteorological application of Principal Component Analysis. *Statistician* **15,** 143–156.

Defernez M, Colquhoun IJ. 2003. Factors affecting the robustness of metabolite fingerprinting using ^{1}H NMR spectra. *Phytochemistry* **62,** 1009–1017.

de Maagd RA, Bosch D, Stiekema W. 1999. *Bacillus thuringiensis* toxin-mediated insect resistance in plants. *Trends in Plant Science* **1,** 9–13.

Fan TWM. 1996. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Progress in Nuclear Magnetic Resonance Spectroscopy* **28,** 161–219.

Fiehn O. 2002. Metabolomics: the link between genotypes and phenotypes. *Plant Molecular Biology* **48,** 155–171.

Gage FH, Dunnett SB, Bjorklund A. 1989. Age-related impairments in spatial memory are independent of those in sensory motor skills. *Neurobiology of Aging* **10,** 347–352.

Geiser M, Schweizer S, Grimm C. 1986. The hypervariable region in the genes coding for entomaphatogenic crystal proteins of *Bacillus thuringiensis*: nucleotide sequence of *hurdh1* gene of subsp. Kurstaki HD1. *Gene* **48,** 109–118.

Ghil M, Vautard R. 1991. Interdecadal oscillations and the warming trends in global temperature time series. *Nature* **350,** 324–327.

Giuliani A, Capuani G, Miccheli A, Aureli T, Ramacci MT, Conti F. 1991. Multivariate data analysis in biochemistry: a new integrative approach to metabolic control in brain aging. *Cellular and Molecular Biology* **37,** 631–638.

Giuliani A, Zbilut JP, Conti F, Manetti C, Miccheli A. 2004. Invariant features of metabolic networks: a data analysis application

on scaling properties of biochemical pathways. *Physica A* **337,** 157–170.

Haaland DM, Thomas EV. 1988. Partial least squares methods for spectral analysis. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Analytical Chemistry* **60,** 1193–1202.

James C. 2004. Global status of commercialized biotech/GM crops: 2004. *International Service for the Acquisition of Agri-biotech Applications* **32,** 1–11.

Keun HC, Ebbels TMD, Antti H, Bollard ME, Beckonert O, Holmes E, Lindon JC, Nicholson JK. 2003. Improved analysis of multivariate data by variable stability scaling: application to NMR-based metabolic profiling. *Analytica Chimica Acta* **490,** 265–276.

Knowles BH, Dow JAT. 1993. The crystal delta-endotoxins of *Bacillus thuringiensis*: model for their mechanism of action on the insect gut. *BioEssays* **15,** 469–476.

Krishnan P, Kruger NJ, Ratcliffe RG. 2005. Metabolite fingerprinting and profiling in plants using NMR. *Journal of Experimental Botany* **56,** 255–265.

Lea PJ, Miflin BJ. 1980. The transport and metabolism of asparagine and other nitrogen compounds within the plant. In: Miflin BJ, ed. *The biochemistry of plants*, Vol. 5. New York, NY: Academic Press, 569–607.

Lebart L, Morineau A, Warwick KM. 1984. *Multivariate descriptive statistical analysis*. New York, NY: Wiley.

Le Gall G, Colquhoun IJ, Davis AL, Collins GJ, Verhoeyen ME. 2003. Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H NMR spectroscopy as a tool to detect potential unintended effects following genetic modification. *Journal of Agricultural and Food Chemistry* **51,** 2447–2456.

Lindon JC, Holmes E, Nicholson JK. 2001. Pattern recognition methods and application in biomedical magnetic resonance. *Progress in Nuclear Magnetic Resonance Spectroscopy* **39,** 1–40.

Manetti C, Bianchetti C, Bizzarri M, *et al.* 2004. NMR-based metabonomic study of transgenic maize. *Phytochemistry* **65,** 3187–3198.

Miccheli A, Aureli T, Delfini M, Di Cocco ME, Viola P, Gobetto R, Conti F. 1988. Study on the influence of inactivation enzyme techniques and extraction procedures on cerebral phosphorylated metabolite levels by P-31 NMR spectroscopy. *Cellular and Molecular Biology* **34,** 591–603.

Preisendorfer RW. 1988. *Principal component analysis in meteorology and oceanography* (Vol. 17 in the series Development in Atmospheric Science). Amsterdam: Elsevier.

Rabe E. 1990. Stress physiology: the functional significance of the accumulation of nitrogen-containing compounds. *Journal of Horticultural Science* **65,** 231–243.

Rahman A. 1989. *One and two dimensional NMR spectroscopy*. Amsterdam: Elsevier.

Randall TD. 1995. An introduction to partial least squares regression. In: *Proceedings of the SAS User Group*. Cary, NC: SAS Institute Inc., 1250–1257.

Ricciolini R, Miccheli A, Di Cocco ME, Piccolella E, Marino A, Sammartino MP, Conti F. 1994. Dexamethasone-dependent modulation of human lymphoblastoid B cell line through sphingosine production. *Biochimica et Biophysica Acta* **1221,** 103–108.

Saneoka H, Nagasaka C, Hahn DT, Yang W, Premachandra GS, Joly RJ, Rhodes D. 1995. Salt tolerance of glycinebetaine-deficient and -containing maize lines. *Plant Physiology* **107,** 631–638.

Sieciechowicz KA, Joy KW, Ireland RJ. 1988. The metabolism of asparagine in plants. *Phytochemistry* **27,** 663–671.

Spearman CH. 1904. General intelligence objectively determined and measured. *American Journal of Psychology* **15,** 201–293.

Steuer R, Kurths J, Fiehn O, Weckwerth W. 2003. Observing and interpreting correlations in metabolomic networks. *Bioinformatics* **19,** 1019–1026.

Sumner LW, Mendes P, Dixon RA. 2003. Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry* **62,** 817–836.

Tabashnik BE, Carriere Y, Dennehy TJ, Morin S, Sisterson MS, Roush RT, Shelton AM, Zhao JZ. 2003. Insect resistance to transgenic Bt crops: lessons from the laboratory and field. *Journal of Economic Entomology* **96,** 1031–1038.

Vandewalle I, Olsson R. 1983. The gamma-aminobutyric acid shunt in germinating *Sinapis alba* seeds. *Plant Science Letters* **31,** 269–273.

Wallsgrove RM. 1990. Enzymes of primary metabolism. In Lea PJ, ed. *Methods in plant biochemistry*, Vol. 3. London: Academic Press, 325–333.

Wang X, Kruger U, Lennox B. 2003. Recursive partial least squares algorithms for monitoring complex industrial processes. *Control Engineering Practice* **11,** 613–632.

Ward JL, Harris C, Lewis J, Beale MH. 2003. Assessment of ^{1}H NMR spectroscopy and multivariate analysis as a technique for metabolite fingerprinting of *Arabidopsis thaliana*. *Phytochemistry* **62,** 949–957.

Weckwerth W. 2003. Metabolomics in systems biology. *Annual Review of Plant Biology* **54,** 669–689.

Wold H. 1966. Estimation of principal components and related models by iterative least squares. In: Krishnaiah PR, ed. *Multivariate analysis*. New York, NY: Academic Press, 391–420.

Wold S. 1994. PLS for multivariate linear modeling. QSAR: chemometric methods in molecular design. In: van de Waterbeemd H, ed. *Methods and principles in medicinal chemistry*. Weinheim: Verlag-Chemie, 195–218.

Wold S, Sjöström M, Eriksson L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems* **58,** 109–130.

Yancey PH. 1994. Compatible and counteracting solutes. In: Strange K, ed. *Cellular and molecular physiology of cell volume regulation*. Boca Raton, FL: CRC Press, 81–109.