JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Unsupervised Principal Component Analysis of NMR Metabolic Profiles for the Assessment of Substantial Equivalence of Transgenic Grapes (*Vitis vinifera*)

Gianfranco Picone,[†] Bruno Mezzetti,[§] Elena Babini,[†,#] Franco Capocasa,[§] Giuseppe Placucci,[†] and Francesco Capozzi*[,†,#]

[†]Department of Food Science, University of Bologna at Cesena, Piazza Goidanich 60, 47520 Cesena (FC), Italy

[§]Department of Environmental Science and Vegetal Productions, Marche Polytechnic University, Via Brecce Bianche, 60131 Ancona (AN), Italy

[#]Centre of Magnetic Resonance (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino (FI), Italy

ABSTRACT: Substantial equivalence is a key concept in the evaluation of unintended and potentially harmful metabolic impact consequent to a genetic modification of food. The application of unsupervised multivariate data analysis to the metabolic profiles is expected to improve the effectiveness of such evaluation. The present study uses NMR spectra of hydroalcoholic extracts, as holistic representations of the metabolic profiles of grapes, to evaluate the effect of the insertion of one or three copies of the *DefH9-iaaM* construct in plants of Silcora and Thompson Seedless cultivars. The comparison of the metabolic profiles of transgenic derivatives with respect to their corresponding natural lines pointed out that the overall metabolic changes occur in the same direction, independent of the host genotype, although the two cultivars are modified to different extents. A higher number of copies not only produces a larger effect but also modifies the whole pattern of perturbed metabolites.

KEYWORDS: transgenic grapes, ^{1}H NMR profile, PCA, metabonomics, metabolites, auxin

INTRODUCTION

Substantial equivalence between a transgenic food and its wild-type genotype, adopted by European Union and United States regulators in food safety assessment, is a concept aiming at demonstrating similarity and, therefore, safety, mainly through tests of physicochemical composition.[1]

However, much criticism was addressed to the concept of substantial equivalence between transgenic and natural foods: it emphasizes chemical composition at the expense of biological, toxicological, and immunological tests; it does not define the point at which a food is no longer substantially equivalent; and the concept actually impedes risk research.[2]

Presently, substantial equivalence is still considered a key step in the safety assessment process of transgenic foods, and it is not a safety assessment in itself.

When a transgenic food is compared with the conventional counterpart, "if the differences exceed natural variations, a nutritional and toxicological assessment is required for the transgenic food".[3]

Several strategies have been developed to identify differences in the composition of gentically modified (GM) food crops that may occur as a result of the genetic modification process.[4] These include profiling methods. Three main cell constituent groups are targeted by profiling technologies: RNA (microarray technologies), proteins (proteomics), and metabolites (metabonomics). Microarray technologies are feasible to monitor the expression of thousands of different genes simultaneously. For instance, expression microarray analyses were used to monitor the extension of unexpected transcriptome modifications obtained in rice by conventional plant breeding as compared with the ones obtained through genetic engineering.[5]

Proteomics is the large-scale study of a proteome, which is the entire complement of proteins, including the modifications made to a particular set of proteins, produced by an organism or system. Because proteins are direct products of gene transcription and translation, they are ideally suited for the detection of changes in the genome (e.g., insertional mutation), in gene regulation (pleiotropic effect), or in biochemical pathways (direct or pleiotropic effect) of a genetically modified plant.[6]

Metabonomics can be expressed as "the quantitative measurement of the dynamic multi-parametric metabolic response of living systems to patho-physiological stimuli or genetic variations and/or modifications".[7] However, a metabonomic study can provide significant information only if the metabolic change in the target group is significantly different from the biological variation of the control group.[8] In other words, metabonomics cannot provide any information when differences between groups are not meaningful, whereas it gives also a quantitative measure of the differences when detected.

Different applications of the metabonomic approach have been documented, especially in toxicity screening,[9] drug metabolism,[10] and functional genomics.[11] This kind of approach, nowadays, finds increasing development also in nutritional science (nutrigenomic approach),[12−18] food chemistry,[19−23] and plant genetic engineering technology.[24]

Received: May 24, 2011
Revised: August 1, 2011
Accepted: August 2, 2011
Published: August 02, 2011

Different analytical platforms have been employed to investigate metabolite profiles such as magnetic resonance (NMR) and mass spectrometry (MS). In particular, the NMR technique has the advantage of being less affected than MS by matrix effects and of providing at least one signal for each molecule, present in the mixture at a detectable concentration, and for this reason it is often indicated as the truly quantitative universal detector.[25−27]

The spectroscopic data are then explored by chemometric techniques, such as the unsupervised principal component analysis (PCA), to simplify and condense in a few parametric descriptors the global information given by the spectra that describe the whole chemical composition. Other supervised multivariate statistical analysis have been also developed (LDA, PLS-DA, ECVA, iECVA),[28−30] but they are mainly directed to highlight, within a complex mixture, those biomarkers indicating a clear difference among categories of samples. However, these supervised methodologies require a suitable number of samples to have a reliable predictive capability not affected by overfitting.[31,32]

Genes, increasing either auxin synthesis or sensitivity and altering auxin signal transduction, allow fruit set in the absence of pollination (parthenocarpy).[33−40] Thus, auxin might improve fruit set and consequently fruit number per plant. In perennial species, auxin also affects development of inflorescence. This is indicated by the increased number of inflorescences on raspberry and strawberry plants genetically engineered with an ovule-specific auxin-synthesizing gene, *DefH9-iaaM*.[38] This gene construct contains an ovule-specific regulatory region from *DefH9* isolated from *Antirrhinum majus* and the *iaaM* coding region from *Pseudomonas savastonoi*. The *iaaM* gene codes for a tryptophan-2-monoxygenase enzyme that converts tryptophan to indole-3-acetamide, which is then hydrolyzed to the auxin indoleacetic acid (IAA).[33]

DefH9-iaaM was introduced into the genome of two grape cultivars with different levels of fecundity mainly due to different average number of inflorescences per shoot.[41]

In Thompson Seedless cultivar the *DefH9-iaaM* gene doubles the average number of inflorescences per shoot, whereas shoot fruitfulness was unaffected in transgenic Silcora. Both transgenic cultivars had an increased number of berries per bunch, but more in Thompson Seedless (30%) in comparison with Silcora (15%). Berries of the transgenic cultivars had shown a substantial equivalent nutritional quality when analyzed by conventional techniques.[42]

With the aim to approach the food safety risk assessment of the new *DefH9-iaaM* table grape lines, the present paper proposes an analytical methodology, based on NMR spectroscopy, aimed at comparing and measuring the overall difference in the chemical composition of table grapes with respect to their transgenic derivative. Our approach statistically evaluates the metabolic profile of grapes, as determined on their hydroalcoholic extracts, and exploits the principles of metabonomics to analyze the changes induced by genetic modifications. Rather than focusing on a number of metabolites, the presented approach compares the compositional variability of grapes by representing it as patterns of spectral bins, each including compounds with similar chemical functionalities.

This paper represents the first application of metabonomics methods to the evaluation of the overall modifications occurring in the metabolic profiles of transgenic grapes.

MATERIALS AND METHODS

Plant Material. The *DefH9-iaaM* gene construct consists of the ovule-specific regulatory region from *DefH9* isolated by *A. majus* and the *iaaM* coding region from *P. savastonoi iaaM* coding for a tryptophan-2-monoxygenase enzyme that converts tryptophan to indole-3-acetamide. The latter is then hydrolyzed to IAA.[33] The *DefH9-iaaM* gene was introduced into the genome of two grape cultivars, that is, cv. Thompson Seedless, with low shoot fruitfulness, and cv. Silcora, with high shoot fruitfulness.[43] Transgenic *DefH9-iaaM* lines of both cultivars have been cultivated (1999−2006) under open-field conditions to compare their fecundity to that of control nontransgenic plants. The open-field experimental trial with transgenic and control clones was established at the Experimental Farm of the Marche Polytechnic University in March 2001, by following the EC (CE 2001/18) rules for transgenic plant field evaluation. All plants, both transgenic and wild type, were cultivated in the same vineyards according to identical agronomic practices. Field trial was set including four plots of four plants of each clone (control and transgenic lines).

Experimental Design for NMR Study. Forty-five NMR samples were prepared from the extracts of the two cultivars and their modified lines. At the 2007 harvest, berries of Thompson Seedless control line (T-WT) and one copy of the *DefH9-iaaM* gene line (T-GM1) were sampled. At the same time berries of Silcora from control line plants (S-WT) and from two *DefH9-iaaM* modified lines, one containing a single copy of the gene (S-GM1) and the other containing three copies of it (S-GM2), were also sampled.[42] At harvest, samples were freeze-dried at −80 °C and then shipped in dry ice to the laboratory for analyses.

Sample Preparation. For each genotype (control and transgenic lines of both cultivars) was prepared a bulk of 100 berries picked randomly from grape clusters of plants growing in different plots of the experimental vineyards. Both skin and pulp of each sample were homogenized under ice chilling using an Ultra Turrax T18 basic dispersing tool (IKA). Three aliquots of about 10.0 g of homogenate, poured in 50 mL Falcon tubes, were separately vortexed with 10.0 mL of a mixture of methanol and 50 mM acetate buffer (pH 5.0, 2:1) to perform the solvent extraction. The suspensions were centrifuged at 10600*g* for 20 min at 4 °C. The resulting hydroalcoholic solution was dispensed in different Eppendorf tubes, as 1 mL aliquots, and stored at −80 °C. Before the free induction decay (FID) acquisition, 10% (v/v) D_2O was added to each 1 mL extract and centrifuged at 20800*g* for 5 min at room temperature. A volume of 800 μL was transferred to a 5 mm NMR tube to acquire a single FID. Three extracts (E1−3) for each homogenate (H1−3) obtained from all genotypes (G1−5) were subjected to NMR analysis by alternating samples so that the five extracts E1:H1:G1−5 were first analyzed, followed by the series E1:H2:G1−5. When the last series, E3:H3:G1−5, was analyzed, in total 45 spectra were acquired. With this sampling scheme, the accuracy and precision of the instrumental analysis, as well as the storage effects, were assessed.

NMR Spectroscopy. The ^{1}H NMR spectra were recorded at 300 K on a Varian Mercury-plus spectrometer, operating at a ^{1}H frequency of 400 MHz; for each spectrum, 2048 scans were acquired, with data collected into 16K data points with a spectral width of 16 ppm, a pulse angle of 60°, a recycle delay of 1.0 s, and an acquisition time of 2.561 s. The water and methanol singlets were suppressed using the WET presaturation sequence, with irradiation at the water and methanol frequencies. Methanol satellites were suppressed by irradiation on the ^{13}C frequency of the solvent. The data were acquired under an automatic procedure, requiring about 2 h per sample. The FID were Fourier transformed, with Mestrelab software, by performing an exponential multiplication with a 1 Hz line broadening. The glucose's β-anomeric signal at 4.67 ppm was taken as chemical shift reference for all spectra.

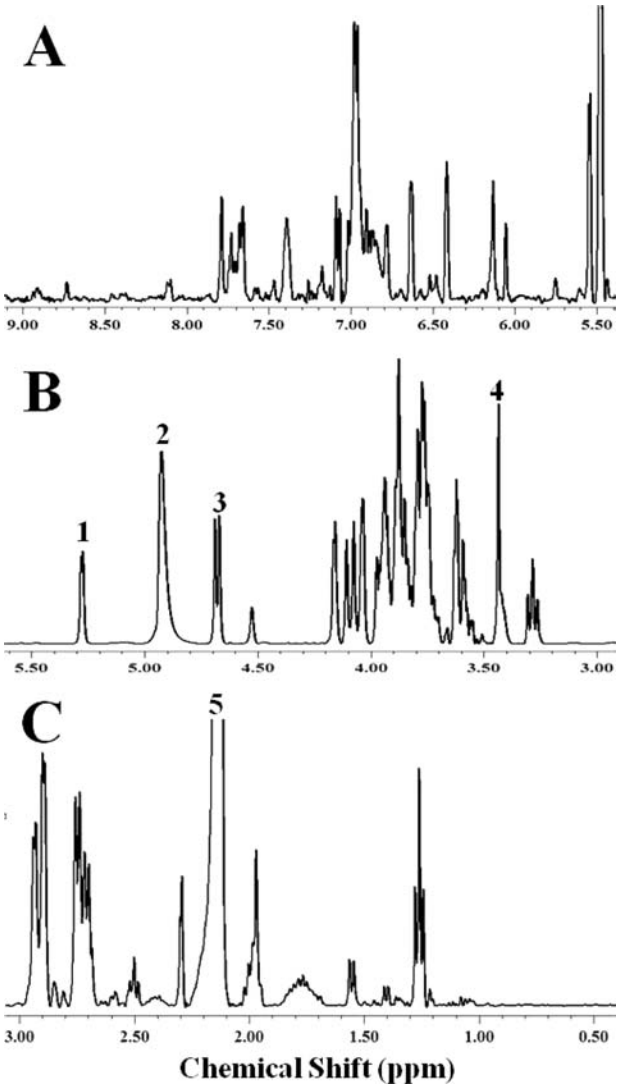

Figure 1. ^{1}H NMR spectrum of hydroalcoholic extract of grape (*Vitis vinifera*). The spectrum was subdivided into three spectral regions (A, downfield; B, midfield; and C, upfield region). Downfield and upfield regions were expanded on the vertical scale to appreciate the presence of small signals. Some signals, easily assigned, are labeled: 1, α-D-glucose; 2, residual water; 3, β-D-glucose; 4, residual methanol; 5, acetic acid belonging to the extraction buffer system.

Phase and multipoint manual baseline corrections were performed in duplicate for each FID to evaluate the errors due to the processing steps. Spectral data points were reduced from 16K (16384) to 8000 points, by deleting the edge parts of the spectra containing any signal above the noise and by cutting off the solvent and buffer signals (water, methanol, and acetic acid). The spectral information was further condensed by subdividing the spectra into 80 bins, each integrating 100 data points. The resulting binned spectra were saved as an ASCII file for the subsequent statistical analysis.

Prestatistical Processing of NMR Data. Strong resonances assigned to residual water (4.908 ppm), methanol (3.396 ppm), and acetic acid (2.107 ppm) are easily identified and excluded from the subsequent chemometric analysis. Although acetic acid is a grape metabolite, it has been excluded from the statistical analysis because it is component of the buffer system, necessary to keep the pH at the chosen value. Among several hundred signals belonging to grape's metabolites, the most intense ones arise from glucose in the midfield region (α-Glu-H1 at 5.252 ppm, β-Glu-H1 at 4.64 ppm, and H2−H6 in the range 3.5−4.20 ppm). Signals falling in the upfield region (<3.0 ppm) and in the downfield region (>5.49 ppm) were expanded on the vertical scale (×20 and ×400, respectively) to visually appreciate signals belonging to minority species (Figure 1).

Table 1. Summary of Student's *t* Test for PC1 and PC2 Applied to Silcora and Thompson Subsets[a]

PC	cultivar	$\|t\|$	df	*p* value
1	Silcora (WT vs GM1)	4.19	34	1.8×10^{-4}
	Silcora (WT vs GM2)	24.16	34	2.2×10^{-16}
2	Silcora (WT vs GM1)	10.59	34	2.6×10^{-12}
	Silcora (WT vs GM2)	3.35	34	1.9×10^{-3}
1	Thompson (WT vs GM)	4.6804	34	4.4×10^{-5}
2	Thompson (WT vs GM)	0.6228	34	5.4×10^{-1}

[a] $|t|$ is the absolute *t* value observed in the statistical analysis; df is degrees of freedom.

Prior to multivariate analysis, data underwent prestatistical improvement, such as normalization, aiming at minimizing unwanted sources of variation due to slightly different instrumental conditions and sample preparation artifacts. Moreover, the signals belonging to some titrable organic acids still show some variations of their chemical shift among different spectra, due to small differences (<0.05 unit) in the pH of the extracts being analyzed. To avoid such a detrimental effect, the spectral data were binned, thus converting each spectrum in a collection of 80 bins, each consisting of the integral area over 100 consecutive spectral data points.[44]

Multivariate Data Analysis. The multivariate analysis was carried out onto mean-centered and scaled binned spectra through PCA by using homemade algorithms written in the R 2.4.0 program language. PCA has been chosen as the "gold standard" for comparison, because it is an unsupervised method able to describe the total sample variance by projecting it in a condensed space.[45] Analysis of variance (ANOVA) was applied to the first 20 PCs, all together collecting 90% of the total variance, to find those PCs able to discriminate between GM and WT grapes, at a statistical *p* level of $<10^{-3}$.

Student's *t* test was carried out on PC1 and PC2 scores to test the statistical significance of the null hypothesis between WT and GM grapes (Table 1).

The line commands, in the R program environment, used for ANOVA and Student's *t* test are anova and *t*.test, respectively. Because the present study is aimed at evaluating the metabolic impact of the genetic modification and not the difference between the metabolic profiles of two cultivars, we first performed separately the PCA on two subsets of the NMR spectra acquired for the Silcora seedless cultivar (S-WT, S-GM1, and S-GM2) and Thompson Seedless cultivar (T-WT and T-GM1). A separate PCA for the two cultivars is required according to the principles of "metabonomics".[46,47] The differences in metabolic profile of two cultivars may be too pronounced to be summarized in a couple of parameters such as PC1 and PC2. Instead, it is correct to describe the differences between the experimental group and a control group by using PC scores.

RESULTS AND DISCUSSION

PCA of the Silcora Cultivar and Transgenic Line Subset. The application of the PCA on the Silcora subset results in the PC plot shown in Figure 2A, where the first principal component (PC1) describes 25% and the second one (PC2) 13% of the total subset variance. Although only 38% of the total variance is

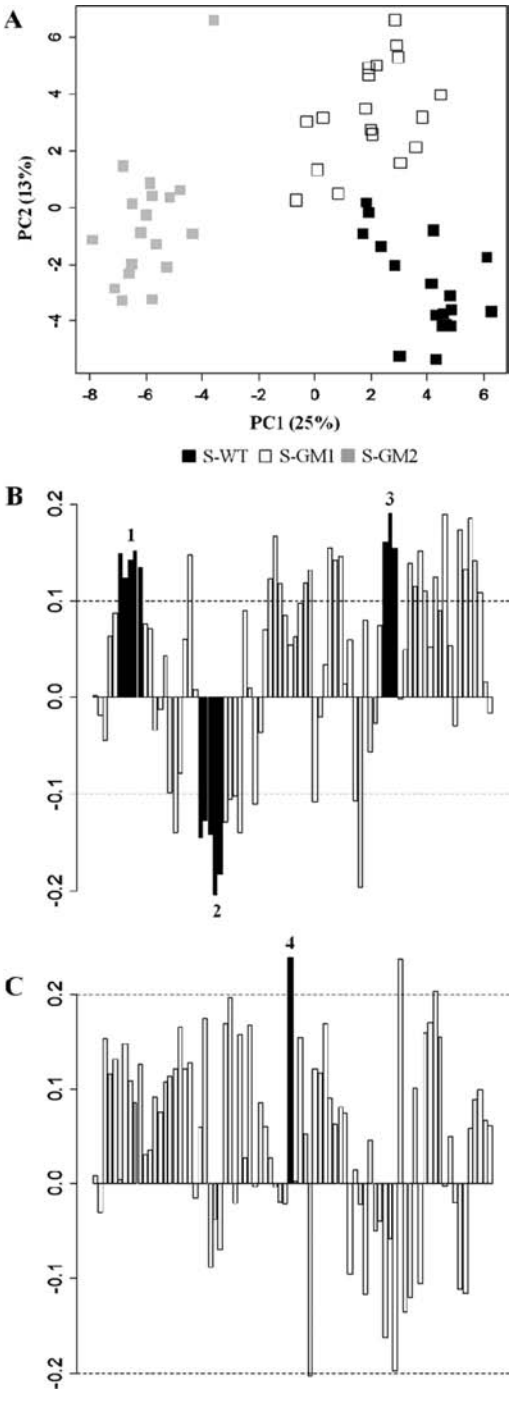

Figure 2. Silcora sample multivariate analysis. (A) Score plot obtained by application of PCA on mean-centered and scaled spectral bins recorded on berry extracts of the Silcora cultivar. The first two PCs explain 25% (PC1) and 13% (PC2), respectively, of the total variance. (B, C) Loading plots for spectral bins, along PC1 and PC2, respectively. The black bins labeled with numbers are described in the text as representative of meaningful bins responsible of separation along the respective PC dimension.

explained by using only the first two PCs, the application of the ANOVA test to the first 20 PCs (explaining 90% of the total variance) pointed out that only the first two PCs were markedly capable of discriminating between GM and WT, at the 10^{-16} p level. This means that all the other following PCs contain

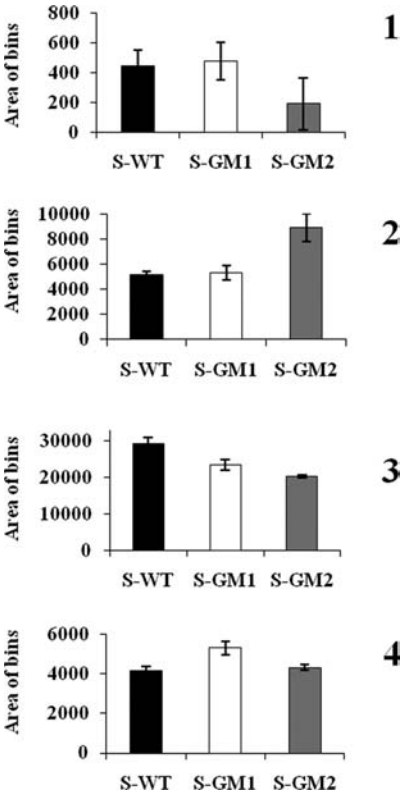

Figure 3. Bar plots representing the areas obtained by integrations of bin groups 1−4, here labeled according to loading plots shown in Figure 2B, C. Standard errors are also shown as black lines on the top of each bar.

information about the variance among samples, regardless of whether they are GM or WT.

As seen from a visual inspection of the PC plot, the variance among the GM and WT grapes is higher than the variance internal to each group. Moreover, the direction of separation between S-GM1 and S-WT is different (mainly along PC1) from the one along which S-GM2 is separated from S-WT (mainly PC2). This result is mainly interpreted by assuming that the spectral features responsible for the differentiation of the S-GM1 group from its nontransgenic clone are different from the ones differentiating S-GM2. Further details arise from the inspection of the PC loadings of the most meaningful components. In Figure 2, panels B and C, the loadings of PC1 and PC2, respectively, are shown, because they represent the only components exhibiting higher intergroup than intragroup variance. The loadings report the weight with which each bin influences the position of the sample within the PC plot.[48] For instance, increasing areas of the bins in the black group labeled 1 and 3, in panel B, push the sample score toward higher values of PC1. The opposite occurs with black group labeled 2, that is, lower values of PC1 corresponding to increasing areas. Similarly, the increasing area of the black bins labeled 4 in panel C increases the PC2 score. It is worth noting that the source of variation is not confined to a few bins and, thus, signals of a few molecules. Rather, the entire plant metabolite profile is subjected to changes. The absolute extent of such changes is, however, not directly interpretable from the analysis of PC loadings.

To understand the extent of changes in the amount of metabolites, the absolute areas of the black bin group labeled in Figure 2 (panels B and C) is reported in Figure 3. The area of

bin group 1 (including signals from aromatic substances[49] is clearly smaller in S-GM2 than in the other two genotypes of the same cultivar (S-WT and S-GM1). The same genotype, however, has a higher amount of other aromatic compounds, grouped in black bins labeled 2, compared with S-WT and S-GM1. This result justifies the decision, previously taken, to exclude the S-GM2 line from the complete field evaluation because of a nonhomogeneous phenotypic variability.[42]

A different trend is observed for black bins labeled 3, including signals from organic acids,[49] because their area decreases in the first variant (S-GM1) and even more in the second variant (S-GM2) with respect to that measured for S-WT. Data from S-GM1 confirm results from previous studies based on conventional chemical analysis.[42] The black bins labeled 4 in Figure 2C, including signals from the midfield spectra region, are involved in the separation along PC2. This bin, indeed, has a slightly higher area in the first variant S-GM1 than in the other genotypes and collects signals belonging to sugars. A further step in the chemometric analysis of the effect induced by genetic modification of grape berries is represented by the statistical description of the discrimination ability covered by each PC dimension.[50−52] The result of Student's *t* test, applied on the PC1 and PC2 scores of all variants, with respect to the WT genotype, is reported in Table 1. The *p* value represents the probability of its being correct to assume that the two compared populations are equivalent on the basis of their PC score.[53] For the Silcora cultivar, it emerges that the two genetically modified variants are both statistically different from the wild-type genotype, at least along one dimension of the PC space, precisely PC1 for S-GM2 and PC2 for S-GM1. Because the latter represents a large portion of the total variance attached to the metabolic profile described by the NMR spectrum, such a result provides evidence that a different copy number of the inserted gene induces different phenotypic changes related to the multiparametric metabolic response. These results confirm that the type and extent of the metabolic modifications occurring in transgenic plants depend on the number of copies of the inserted gene, so that it is of great importance to identify transgenic lines with the lowest number of inserted gene copies.[54,55]

For this reason, only S-GM1, between the two Silcora transgenic lines, was considered for more detailed field studies.[42]

PCA of the Thompson Cultivar Subset and Its Transgenic Line. The PCA approach was also applied to the Thompson cultivar subset, and the results can be summarized through the corresponding PC plot shown in Figure 4A.

The first consideration arises from the fact that only 27% of the total variance is described by the first two principal components, whereas up to 38% was obtained from the same number of PCs in the Silcora case. Such a result implies a lower descriptive power associated with the chosen PC plot which, however, can be raised by inspecting other dimensions of the PC space.[51]

The PC plot of Figure 4A allows us to assume that only PC1 has some tendency to discriminate between berries belonging to T-WT and transgenic genotypes, although this tendency is not as clear as in the other cultivar. Indeed, Student's *t* test gives results with higher *p* values (4.4×10^{-5} vs 2.2×10^{-16} or 2.6×10^{-12}) (Table 1).

Also, for the PC space calculated on the Thompson subset, the PC loadings give the weight of each bin in determining the extent of the separation along each PC dimension. In this way, black bins labeled 1 and 2 (both in the aromatic region) are responsible for the poor separation along PC1, so that a higher area is associated with the transgenic variant. This finding is also confirmed by integrating areas of such bins (Figure 5). The group of black bins labeled 3 exerts a slight decrease of its area, thus suggesting that the amount of organic acid is lower in the transgenic line than in its control line, which confirms the results based on chemical analyses performed by detecting the total

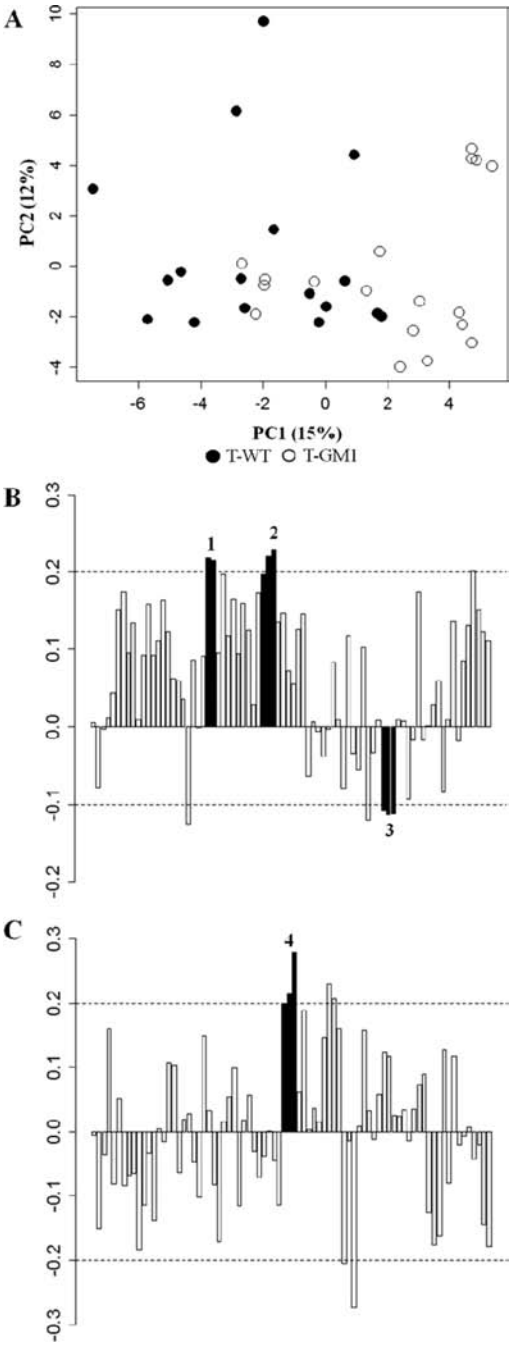

Figure 4. Thompson samples multivariate analysis. (A) Score plot obtained by application of PCA on mean-centered and scaled spectral bins recorded on berry extracts of Thompson cultivar. The first two PCs explain 15% (PC1) and 12% (PC2), respectively, of the total variance. (B, C) Loading plots for spectral bins, along PC1 and PC2, respectively. The black bins labeled with numbers are described in the text as representative of meaningful bins responsible of separation along the respective PC dimension.

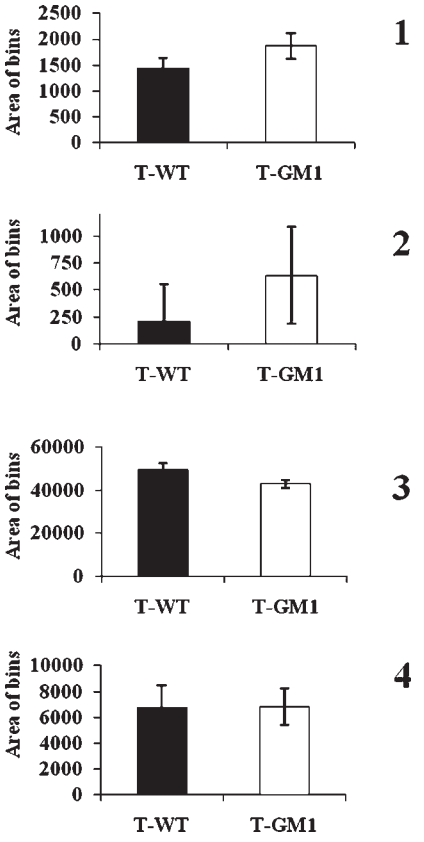

Figure 5. Bar plots representing the areas obtained by integrations of bin groups 1−4, here labeled according to loading plots shown in Figure 4B,C. Standard errors are also shown as black lines on the top of each bar.

titrable acidity.[42] On the contrary, although black bins labeled 4, containing sugar signals, have high weight on PC2, their cumulative area shows no meaningful difference when transgenic and control lines are compared. This is also expected on the basis of the fact that PC2 is, indeed, not able to discriminate between the two lines of the Thompson cultivar. Field evaluation of these two lines showed minor differences in fruit quality, whereas the effect of the gene was more evident in increasing plant productivity.[35,56] The high similarity of fruit metabolic pattern between control and transgenic lines of Thompson Seedless can be considered as a first indication of more substantial equivalence between the two lines, in comparison with the Silcora cultivar, which is quite important information for the evaluation protocols expected by the rules for the release of transgenic plants.

PCA Applied to Both Cultivars, Including Control and Transgenic Lines. Results from a global PCA including all samples belonging to both cultivars are shown in Figure 6. The purpose of this further analysis is not to explore the differences between the wild genotypes of both cultivars but, rather, to determine whether the genetic modification shifts the metabolite profile toward the same direction, independent of the cultivar.

Indeed, the application of PCA to separate cultivars emphasizes the metabolic impact of the genetic modification within each cultivar. However, there is no information about possible similarities between the metabolic profile changes occurring in Thompson Seedless with respect to the ones occurring in the Silcora cultivar. This can be ascribed to the fact that the direction of each PC in the variance space is just relative to the analyzed set

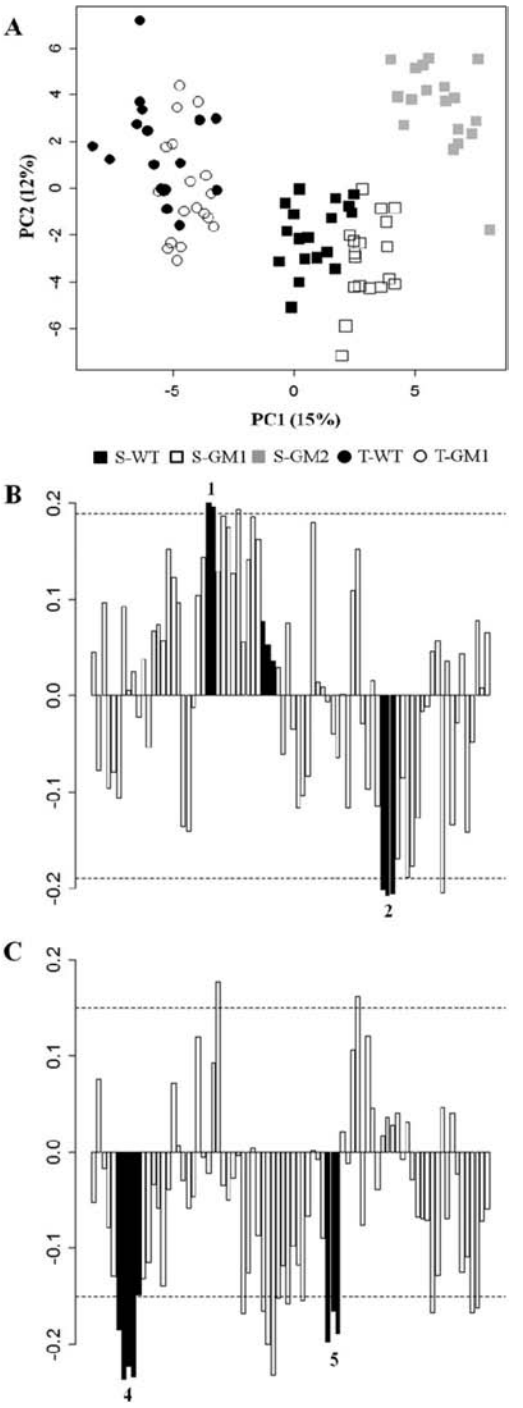

Figure 6. All grape sample multivariate analysis. (A) Score plot obtained by application of PCA on mean-centered and scaled spectral bins recorded on berry extracts of all cultivars and genotypes. The first two PCs explain 26% (PC1) and 12% (PC2), respectively, of the total variance. (B, C) Loading plots for spectral bins, along PC1 and PC2, respectively. The black bins labeled with numbers are described in the text as representative of meaningful bins responsible of separation along the respective PC dimension.

of spectra. Thus, the inclusion of other samples in the same set may affect the orientation of PCs in the new variance space.[57] When the highest variance in the comprehensive data set is between the two cultivars, PC1 must align along that variance

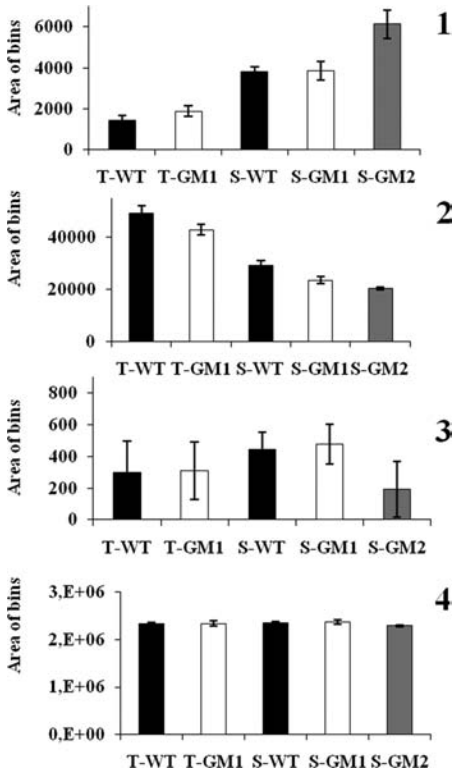

Figure 7. Bar plots representing the areas obtained by integrations of bin groups 1−4, here labeled according to loading plots shown in Figure 6B,C. Standard errors are also shown as black lines on the top of each bar.

direction, decreasing the effect and the importance of the genetic modification. For this reason, the extent and the direction, in terms of metabolite pattern, of the genetic modification should be evaluated both separately and comprehensively on all considered cultivars.

The PCA applied to all samples of both cultivars (Figure 6) points out that GM1 modification, that is, with one gene copy inserted, shifts the metabolic profile of the Silcora cultivar along the same direction of the PC score plot as for the Thompson cultivar, although to a lower extent.

Some further consideration can be extracted from the inspection of integral areas of selected bin groups (Figure 7). Aromatic compounds of bin group 1 are present at higher concentrations in the Silcora control line than in the Thompson control line. Moreover, all transgenic lines contain higher amounts of aromatic compounds belonging to group 1 than their corresponding WT line, except for GM of the Thompson variety, for which integrals show comparable amounts. It is intriguing that, in a previous study, the total amount of polyphenols was found to be 25% higher in S-GM1 than in its control line, whereas it is only 9% higher in the T-GM1 cultivar than in its corresponding control line.[42]

The opposite trend appears for the organic acids collected in the group of black bins labeled 2, because transgenic berries contain lower concentrations than their control counterpart. Moreover, also when only control lines of both cultivars are compared, organic acids included in the black bins labeled 2 show an opposite trend to that of aromatic compounds in those labeled 1. In fact, T-WT berries contain higher amounts of organic acids than S-WT. Again, the same trend was found in the chemical determination of total acids, although mostly due to tartaric acid variations rather than to changes in malic and citric acid concentrations.[42]

Interestingly, S-GM2 has decreased amounts of the aromatic compounds grouped in black bins labeled 3, whereas different aromatic compounds, grouped in bins 1, were subjected to an increase consequent to the same type of genetic modification.

According to the present results, it is possible to identify some NMR spectral regions that contain signals from metabolites that show statistically significant differences when the genetically modified lines and their respective control samples are compared. Such differences are mainly found in the aromatic region containing signals of tryptophan and indole derivatives, for which the metabolic pathway is affected by the genetic insertion. Moreover, meaningful variations are also observed in the organic acid regions, previously found to change their concentration. The critical role of auxin in plant growth and fruit development is a well-known issue,[58,59] and these results reveal the perspective offered by the NMR technology in identifying the entity of metabolic impact in specific tissues with modified auxin metabolism.

The chemometric analysis, which the metabonomics approach is based on, shows that the Silcora grape cultivar, more than the Thompson one, exerts statistically significant differences of their metabolic profile, as observed by NMR, when transgenic lines and control ones are compared.

According to the genetic modification, the results were supposed to be mainly related to metabolites involved in the ripening process. It is not the purpose of the present study to characterize individually the metabolites involved in the genetic modification. Rather, when substantial equivalence needs to be evaluated, the present holistic approach eliminates most of the constraints associated with the chemical analysis, for example, the quantification of only known components.

The combination of NMR spectroscopy together with unsupervised data analysis becomes an important tool able to identify differences in the metabolic profile also without looking for specific metabolites expected to change.[60] Such a nontargeted approach should be considered as a preliminary step in food benefit and risk assessment of a new transgenic plant or product, as previously considered in the identification step of unintended effects.[4]

Obviously, in the case of nonequivalence, the identification of all metabolites affected by the inserted genes must be performed, by applying both analytical, such as spectroscopic and chromatographic, and supervised statistical methodologies.

In the light of such considerations, the metabonomics approach applied to food science may result as a further tool in the hands of scientists that need more and more food descriptors to be taken under control during the risk assessment strategy. Moreover, the present approach may be used to define, together with other statistical and analytical tools, the point at which a transgenic plant/product is no longer substantially equivalent to the one that has not been modified, that is, when the metabolic profile of the modified line is statistically different from that of its control.

In fact, the two table grape cultivars showed different extents of metabolic modification, measured as distance in the PC score plot, when transformed with the same *DefH9-iaaM* gene, the level of variation observed between the transgenic lines and control of Silcora being much higher than that observed in Thompson Seedless control and transgenic lines. This confirms the importance of the genetic background in determining the

level of genetic and metabolic variations induced by the inserted gene.

Because the genetic modification has generated GM grapes that are statistically nonsubstantially equivalent to their corresponding wild genotype, the next study should characterize the chemical nature of metabolites undergoing major concentration changes, to rationalize and, possibly, to understand the main metabolic pathways being influenced by the genetic modifications.

It is worth remembering that physical−chemical analysis is not enough to ensure the safety of GM foods; animal and human feeding tests are also required by European and United States regulators. In this respect, unsupervised PCA performed on NMR spectra, or other comprehensive metabolic profiles of WT and GM derivatives, could assist in selecting samples to be included in such feeding tests, by picking up those showing the highest variance between groups, thus potentiating the effectiveness of such tests.

AUTHOR INFORMATION

Corresponding Author

*Phone: +39 0547 338 105. Fax: +39 0547 382 348. E-mail: francesco.capozzi@unibo.it.

Funding Sources

This work was supported by the Italian Ministry of University and Research (MIUR) through the PRIN 2007 Program. G.P. is supported by the Consorzio Interuniversitario Risonanze Magnetiche di Metalloproteine Paramagnetiche (CIRMMP) through his doctoral fellowship.

ABBREVIATIONS USED

NMR, nuclear magnetic resonance; FID, free induction decay; GM, genetically modified; S, Silcora cultivar; T, Thompson Seedless cultivar; WT, control line; GM1, transgenic line with one copy of *DefH9-iaaM* gene; GM2, transgenic line with three copies of *DefH9-iaaM* gene; IAA, indoleacetic acid; ANOVA, analysis of variance; PCA, principal component analysis.

REFERENCES

(1) Levidow, L.; Murphy, J.; Carr, S. Recasting "substantial equivalence": trans-Atlantic governance of GM food. *Sci. Technol. Hum. Val.* **2007**, *32*, 26–64.

(2) Millstone, E.; Brunner, E.; Mayer, S. Beyond "substantial equivalence". *Nature* **1999**, *401*, 525–526.

(3) Kuiper, H. A.; Kleter, G. A.; Noteborn, H. P. J. M.; Kok, E. J. Assessment of the food safety issues related to genetically modified foods. *Plant J.* **2001**, *27*, 503–528.

(4) Kuiper, H. A.; Kok, E. J.; Engel, K.-H. Exploitation of molecular profiling techniques for GM food safety assessment. *Curr. Opin. Biotechnol.* **2003**, *14*, 238–243.

(5) Batista, R.; Saibo, N.; Lourenço, T.; Oliveira, M. M. Microarray analyses reveal that plant mutagenesis may induce more transcriptomic changes than transgene insertion. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 3640–3645.

(6) Ruebelt, M. C.; Leimgruber, N. K.; Lipp, M.; Reynolds, T. L.; Nemeth, M. A.; Astwood, J. D.; Engel, K.-H.; Jany, K.-D. Application of two-dimensional gel electrophoresis to interrogate alterations in the proteome of genetically modified crops. 1. Assessing analytical validation. *J. Agric. Food Chem.* **2006**, *54*, 2154–2161.

(7) Nicholson, J. K.; Lindon, J. C.; Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* **1999**, *29*, 1181–1189.

(8) Abdel-Farid, I. B.; Kim, H. K.; Choi, Y. H.; Verpoorte, R. Metabolic characterization of *Brassica rapa* leaves by NMR spectroscopy. *J. Agric. Food Chem.* **2007**, *55*, 7936–7943.

(9) Zhang, X.; Wu, H.; Liao, P.; Li, X.; Ni, J.; Pei, F. NMR-based metabonomic study on the subacute toxicity of aristolochic acid in rats. *Food Chem. Toxicol.* **2006**, *44*, 1006–1014.

(10) Keun, H. C. Metabonomic modeling of drug toxicity. *Pharmacol. Ther.* **2006**, *109*, 92–106.

(11) Gavaghan, C. L.; Wilson, I. D.; Nicholson, J. K. Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. *FEBS Lett.* **2002**, *530*, 191–196.

(12) Solanky, K. S.; Bailey, N. J.; Beckwith-Hall, B. M.; Bingham, S.; Davis, A.; Holmes, E.; Nicholson, J. K.; Cassidy, A. Biofluid ^{1}H NMR-based metabonomic techniques in nutrition research − metabolic effects of dietary isoflavones in humans. *J. Nutr. Biochem.* **2005**, *16*, 236–244.

(13) Betoret, E.; Betoret, N.; Vidal, D.; Fito, P. Functional foods development: trends and technologies. *Trends Food Sci. Technol.* **2011** (doi:10.1016/j.tifs.2011.05.004).

(14) Bertram, H. C.; Knudsen, K. E. B.; Serena, A.; Malmendal, A.; Nielsen, N. C.; Fretté, X. C.; Andersen, H. J. NMR-based metabonomic studies reveal changes in the biochemical profile of plasma and urine from pigs fed high-fibre rye bread. *Br. J. Nutr.* **2006**, *95*, 955–962.

(15) Young, J.; Larsen, L.; Malmendal, A.; Nielsen, N.; Straadt, I.; Oksbjerg, N.; Bertram, H. Creatine-induced activation of antioxidative defence in myotube cultures revealed by explorative NMR-based metabonomics and proteomics. *J. Int. Soc. Sports Nutr.* **2010**, *7*, 9.

(16) Bertram, H. C.; Duus, J. Ø.; Petersen, B. O.; Hoppe, C.; Larnkjær, A.; Schack-Nielsen, L.; Mølgaard, C.; Michaelsen, K. F. Nuclear magnetic resonance-based metabonomics reveals strong sex effect on plasma metabolism in 17-year-old Scandinavians and correlation to retrospective infant plasma parameters. *Metabolism* **2009**, *58*, 1039–1045.

(17) Kemsley, E. K.; Le Gall, G. N. L.; Dainty, J. R.; Watson, A. D.; Harvey, L. J.; Tapp, H. S.; Colquhoun, I. J. Multivariate techniques and their application in nutrition: a metabolomics case study. *Br. J. Nutr.* **2007**, *98*, 1–14.

(18) Primrose, S.; Draper, J.; Elsom, R.; Kirkpatrick, V.; Mathers, J. C.; Seal, C.; Beckmann, M.; Haldar, S.; Beattie, J. H.; Lodge, J. K.; Jenab, M.; Keun, H.; Scalbert, A. Metabolomics and human nutrition. *Br. J. Nutr.* **2011**, *105*, 1277–1283.

(19) Castro, C.; Manetti, C. A multiway approach to analyze metabonomic data: a study of maize seeds development. *Anal. Biochem.* **2007**, *371*, 194–200.

(20) Bertram, H. C.; Weisbjerg, M. R.; Jensen, C. S.; Pedersen, M. G.; Didion, T.; Petersen, B. O.; Duus, J. Ø.; Larsen, M. K.; Nielsen, J. H. Seasonal changes in the metabolic fingerprint of 21 grass and legume cultivars studied by nuclear magnetic resonance-based metabolomics. *J. Agric. Food Chem.* **2010**, *58*, 4336–4341.

(21) de Vos, R. C. H.; Hall, R. D.; Moing, A. Metabolomics of a model fruit: tomato. *Annu. Plant Rev.* **2011**, *43*, 109−155.

(22) Bekiroğglu, S. An emerging powerful technique: NMR applications on quality assessments of fish and related products. In *Handbook of Seafood Quality, Safety and Health Applications*; Wiley-Blackwell: Oxford, U.K., 2011; pp 181−193.

(23) Rodrigues, J. A.; Barros, A. S.; Carvalho, B.; Brandão, T.; Gil, A. M. Probing beer aging chemistry by nuclear magnetic resonance and multivariate analysis. *Anal. Chim. Acta* **2011**, *702*, 178–187.

(24) Manetti, C.; Bianchetti, C.; Bizzarri, M.; Casciani, L.; Castro, C.; D'Ascenzo, G.; Delfini, M.; Di Cocco, M. E.; Laganà, A.; Miccheli, A.; Motto, M.; Conti, F. NMR-based metabonomic study of transgenic maize. *Phytochemistry* **2004**, *65*, 3187–3198.

(25) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **2005**, *56*, 255–265.

(26) Kyong Kim, H.; Choi, Y. H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protoc.* **2010**, *5*, 536–549.

(27) Colquhoun, I. J. Use of NMR for metabolic profiling in plant systems. *J. Pestic. Sci.* **2007**, *32*, 200–212.

(28) Savorani, F.; Picone, G.; Badiani, A.; Fagioli, P.; Capozzi, F.; Engelsen, S. B. Metabolic profiling and aquaculture differentiation of gilthead sea bream by ^{1}H NMR metabonomics. *Food Chem.* **2010**, *120*, 907–914.

(29) Nørgaard, L.; Bro, R.; Westad, F.; Engelsen, S. B. A modification of canonical variates analysis to handle highly collinear multivariate data. *J. Chemom.* **2006**, *20*, 425–435.

(30) Alonso-Salces, R. M.; Héberger, K.; Holland, M. V.; Moreno-Rojas, J. M.; Mariani, C.; Bellan, G.; Reniero, F.; Guillou, C. Multivariate analysis of NMR fingerprint of the unsaponifiable fraction of virgin olive oils for authentication purposes. *Food Chem.* **2010**, *118*, 956–965.

(31) Faber, N. M.; Rajkó, R. How to avoid over-fitting in multivariate calibration — the conventional validation approach and an alternative. *Anal. Chim. Acta* **2007**, *595*, 98–106.

(32) Llabjani, V.; Trevisan, J. L.; Jones, K. C.; Shore, R. F.; Martin, F. L. Binary mixture effects by PBDE congeners (47, 153, 183, or 209) and PCB congeners (126 or 153) in MCF-7 cells: biochemical alterations assessed by IR spectroscopy and multivariate analysis. *Environ. Sci. Technol.* **2010**, *44*, 3992–3998.

(33) Rotino, G. L.; Perri, E.; Zottini, M.; Sommer, H.; Spena, A. Genetic engineering of parthenocarpic plants. *Nat. Biotechnol.* **1997**, *15*, 1398–1401.

(34) Ficcadenti, N.; Sestili, S.; Pandolfini, T.; Cirillo, C.; Rotino, G. L.; Spena, A. Genetic engineering of parthenocarpic fruit development in tomato. *Mol. Breed.* **1999**, *5*, 463–470.

(35) Acciarri, N.; Restaino, F.; Vitelli, G.; Perrone, D.; Zottini, M.; Pandolfini, T.; Spena, A.; Rotino, G. Genetically modified parthenocarpic eggplants: improved fruit productivity under both greenhouse and open field cultivation. *BMC Biotechnol.* **2002**, *2*, 4.

(36) Pandolfini, T.; Rotino, G.; Camerini, S.; Defez, R.; Spena, A. Optimisation of transgene action at the post-transcriptional level: high quality parthenocarpic fruits in industrial tomatoes. *BMC Biotechnol.* **2002**, *2*, 1.

(37) Carmi, N.; Salts, Y.; Dedicova, B.; Shabtai, S.; Barg, R. Induction of parthenocarpy in tomato via specific expression of the rolB gene in the ovary. *Planta* **2003**, *217*, 726–735.

(38) Mezzetti, B.; Landi, L.; Pandolfini, T.; Spena, A. The defH9-iaaM auxin-synthesizing gene increases plant fecundity and fruit production in strawberry and raspberry. *BMC Biotechnol.* **2004**, *4*, 4.

(39) Wang, H.; Jones, B.; Li, Z.; Frasse, P.; Delalande, C.; Regad, F.; Chaabouni, S.; Latche, A.; Pech, J.-C.; Bouzayen, M. The tomato Aux/IAA transcription factor IAA9 is involved in fruit development and leaf morphogenesis. *Plant Cell* **2005**, *17*, 2676–2692.

(40) Yin, Z.; Malinowski, R.; Ziółkowska, A.; Sommer, H.; Plcader, W.; Malepszy, S. The DefH9-iaaM-containing construct efficiently induces parthenocarpy in cucumber. *Cell. Mol. Biol. Lett.* **2006**, *11*, 279–290.

(41) Mezzetti, B.; Pandolfini, T.; Navacchi, O.; Landi, L. Genetic transformation of *Vitis vinifera* via organogenesis. *BMC Biotechnol.* **2002**, *2*, 18.

(42) Costantini, E.; Landi, L.; Silvestroni, O.; Pandolfini, T.; Spena, A.; Mezzetti, B. Auxin synthesis-encoding transgene enhances grape fecundity. *Plant Physiol.* **2007**, *143*, 1689–1694.

(43) Mezzetti, B.; Landi, L.; Scortichini, L.; Rebori, A.; Spena, A.; Pandolfini, T. Genetic engineering of parthenocarpic fruit development in strawberry. *Acta Hortic.* **2002**, *567*, 101–104.

(44) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Anal. Chem.* **2006**, *78*, 2262–2267.

(45) Kozak, M.; Scaman, C. H. Unsupervised classification methods in food sciences: discussion and outlook. *J. Sci. Food Agric.* **2008**, *88*, 1115–1127.

(46) Cevallos-Cevallos, J. M.; Reyes-De-Corcuera, J. I.; Etxeberria, E.; Danyluk, M. D.; Rodrick, G. E. Metabolomic analysis in food science: a review. *Trends Food Sci. Technol.* **2009**, *20*, 557–566.

(47) Wishart, D. S. Metabolomics: applications to food science and nutrition research. *Trends Food Sci. Technol.* **2008**, *19*, 482–493.

(48) Jolliffe, I. Principal component analysis. In *Encyclopedia of Statistics in Behavioral Science*; Wiley: New York, 2005.

(49) Pereira, G. E.; Gaudillere, J.-P.; Van Leeuwen, C.; Hilbert, G.; Lavialle, O.; Maucourt, M.; Deborde, C.; Moing, A.; Rolin, D. ^{1}H NMR and chemometrics to characterize mature grape berries in four wine-growing areas in Bordeaux, France. *J. Agric. Food Chem.* **2005**, *53*, 6382–6389.

(50) Cañas, R. A.; Amiour, N.; Quilleré, I.; Hirel, B. An integrated statistical analysis of the genetic variability of nitrogen metabolism in the ear of three maize inbred lines (*Zea mays* L.). *J. Exp. Bot.* **2011**, *62*, 2309–2318.

(51) Sobolev, A.; Capitani, D.; Giannino, D.; Nicolodi, C.; Testone, G.; Santoro, F.; Frugis, G.; Iannelli, M.; Mattoo, A.; Brosio, E.; Gianferri, R.; D'Amico, I.; Mannina, L. NMR-Metabolic methodology in the study of GM foods. *Nutrients* **2010**, *2*, 1–15.

(52) Piccioni, F.; Capitani, D.; Zolla, L.; Mannina, L. NMR metabolic profiling of transgenic maize with the Cry1A(b) gene. *J. Agric. Food Chem.* **2009**, *57*, 6041–6049.

(53) Schervish, M. J. *P* values: what they are and what they are not. *Am. Stat.* **1996**, *50*, 203–206.

(54) van der Krol, A. R.; Mur, L. A.; Beld, M.; Mol, J. N. M.; Stuitje, A. R. Flavonoid Genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell Online* **1990**, *2*, 291–299.

(55) Gelvin, S. B. *Agrobacterium*-mediated plant transformation: the biology behind the "gene-jockeying" tool. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 16–37.

(56) Acciarri, N.; Restaino, F.; Vitelli, G.; Perrone, D.; Zottini, M.; Pandolfini, T.; Spena, A.; Rotino, G. L. Improved fruit productivity under both greenhouse and open field cultivation. *BMC Biotechnol.* **2002**, *2*, 4.

(57) François, O.; Currat, M.; Ray, N.; Han, E.; Excoffier, L.; Novembre, J. Principal component analysis under population genetic models of range expansion and admixture. *Mol. Biol. Evol.* **2010**, *27*, 1257–1268.

(58) Woodward, A. W.; Bartel, B. Auxin: regulation, action, and interaction. *Ann. Bot. (London)* **2005**, *95*, 707–735.

(59) Molesini, B.; Pandolfini, T.; Rotino, G. L.; Dani, V.; Spena, A. Aucsia gene silencing causes parthenocarpic fruit development in tomato. *Plant Physiol.* **2009**, *149*, 534–548.

(60) Picone, G.; Balling Engelsen, S.; Savorani, F.; Testi, S.; Badiani, A.; Capozzi, F. Metabolomics as a powerful tool for molecular quality assessment of the fish *Sparus aurata*. *Nutrients* **2011**, *3*, 212–227.