JOURNAL OF
AGRICULTURAL AND
FOOD CHEMISTRY

Metabolite Profiling of Tomato (*Lycopersicon esculentum)* Using ^{1}H NMR Spectroscopy as a Tool To Detect Potential Unintended Effects Following a Genetic Modification

GWÉNAËLLE LE GALL,[†] IAN J. COLQUHOUN,[*,†] ADRIENNE L. DAVIS,[‡] GEOFF J. COLLINS,[‡] AND MARTINE E. VERHOEYEN[‡]

Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, United Kingdom, and Unilever Research, Colworth House, Sharnbrook, Bedfordshire MK44 1LQ, United Kingdom

The maize transcription factors *LC* and *C1* were simultaneously overexpressed in tomato with the aim of producing lines with increased amounts of flavonols. The metabolite composition of these genetically modified tomatoes has been compared with that of azygous (nonmodified) controls grown side-by-side under the same conditions. It has been possible to observe metabolic changes in both types at different stages of maturity. ^{1}H NMR spectra showed that the levels of glutamic acid, fructose, and some nucleosides and nucleotides gradually increase from the immature to the ripe stage, whereas some amino acids such as valine and γ-aminobutyric acid were present in higher amounts in unripe tomatoes. Apart from the significantly increased content of six main flavonoid glycosides (mainly kaempferol-3-*O*-rutinoside, with additional increases in kaempferol-3,7-di-*O*-glucoside (**1**), kaempferol-3-*O*-rutinoside-7-*O*-glucoside (**2**), kaempferol-3-*O*-glucoside, a dihydrokaempferol-*O*-hexoside (**3**), and naringenin-7-*O*-glucoside), the levels of at least 15 other metabolites were found to be different between the two types of red tomato. Among them were citric acid, sucrose, phenylalanine, and trigonelline. However, although statistically significant, these changes in mean values were relatively minor (less than 3-fold) and within the natural variation that would be observed in a field-grown crop. Nevertheless, this study clearly showed that NMR combined with chemometrics and univariate statistics can successfully trace even small differences in metabolite levels between plants and therefore represents a powerful tool to detect potential unintended effects in genetically modified crops.

KEYWORDS: **NMR; tomato; transgenic; GMO; metabolite profiling; metabolomics; ripening; maturity; *Lycopersicum esculentum***

INTRODUCTION

Genetically modified organisms (GMOs) have been developed increasingly during the past decade. Genetic engineering of plants holds promise in improving the quality of crops and enhancing the nutritional properties of the plants used for human and animal consumption. However, the introduction of the first generation of GM foods to the market during the mid-90s has given rise to public concern.

The safety testing of GMOs is a high priority for regulatory authorities and there is a need for techniques that are able to detect any unintended effects following a genetic modification. Conner and Jacobs (*1*) have outlined the mechanisms by which such unintended effects can occur in genetic engineering but pointed out that exactly the same mechanisms apply to traditional breeding procedures. It is unlikely, however, that there will be any relaxation in the requirements for scrutiny, especially for newer GMOs where genetic modification is used for engineering of metabolic pathways. Genomics, proteomics, and metabolomics are now making possible a range of nontargeted analyses at the gene, protein, and metabolite levels that may contribute to GMO risk assessment procedures (*2*) as well as to the characterization of new varieties developed by the traditional methods.

High-resolution ^{1}H NMR is a promising screening technique that could answer some of the concerns that GMOs are causing. The potential of NMR to quantify and identify a large number of compounds (technically any metabolite with a hydrogen atom, providing that the quantity of the compound is above the NMR detection limit) makes it a leading technique in the emerging area of metabolomic studies.

Tomato is a major food crop worldwide, and recently genetic modification has been used to up-regulate flavonoid biosynthesis (*3*, *4*) in order to enhance its antioxidant capacity. The transgenic tomatoes used in this study were generated by simultaneous

* Corresponding author. Tel.: +44 1603 255353. Fax: +44 1603 507723. E-mail: ian.colquhoun@bbsrc.ac.uk.
† Institute of Food Research.
‡ Unilever Research.

10.1021/jf0259967 CCC: $25.00 © 2003 American Chemical Society
Published on Web 04/01/2003

overexpression of two maize regulatory genes, *leaf color* (*LC*) and *colorless-1* (*C1*), which led to a significant increase of kaempferol glycosides in the flesh of the fruit (*3*). In addition, increases in naringenin- and dihydrokaempferol-glycosides were also observed (*5*). *LC* and *C1* are known to regulate biosynthesis of anthocyanins, a subclass of flavonoids. There are no reports of *LC* and *C1* regulating biosynthesis of compounds other than flavonoids and phenyl propanoids. Hence, the expectation was that modifications would be restricted to the flavonoid/phenylpropanoid biosynthesis pathway.

The characterization previously carried out concentrated on analyzing the target flavonoids (*3*, *5*). ^{1}H NMR, used for the work described here, offers the potential to analyze the content of sugars, amino and organic acids, or other compounds in the transgenic tomatoes in comparison with their controls using a combination of multivariate (PCA and PLS) and univariate (ANOVA) methods. The effect of the fruit maturity on metabolite composition is also reported.

MATERIALS AND METHODS

Materials. Unilever R&D Colworth (Sharnbrook, Bedford, UK) provided nontransgenic and transgenic tomatoes overexpressing the transcription factors *LC* and *C1*. The fruit used was from homozygous line 2059 (T5 generation) and had at least 10-fold increased levels of kaempferol-glycosides compared with fruit from the corresponding azygous control line. The plant transformation details have already been reported (*3*). Samples were obtained from five pairs of plants (transgenic and control) grown under identical conditions. Eight fruits were taken from each plant (six red plus one from each plant at both the green and turning stages). Eighty samples were prepared in total (*5*).

Methanol-d_4 and D_2O were purchased from Goss Scientific Instruments Ltd (Great Baddow, UK), and TSP (sodium 3-(trimethylsilyl)-propionate-d_4) was from Sigma-Aldrich (Gillingham, UK).

Standards of amino acids, organic and fatty acids, sugars, nucleosides and nucleotides, chlorogenic acid, and trigonelline were purchased from Sigma-Aldrich (Gillingham, UK). Flavonoid standards were purchased as in (*5*).

Methods. *Extraction*. Each sample (whole fruit) was freeze-dried and the product obtained was ground to a fine powder using a coffee grinder. Each sample was prepared by addition of 1.2 mL of 70% methanol-d_4/30% buffer (100 mM K_2HPO_4/KH_2PO_4, 1 mM EDTA (disodium ethylenediamine tetraacetic acid), 1 mM TSP) to 0.048 g of freeze-dried powder. The mixture was stirred at room temperature for 30 min and centrifuged at 10 000 rpm for 10 min (Jouan A14 centrifuge). Each NMR sample consisted of 750 μL of the supernatant, which was stored at −18 °C until required for analysis.

NMR Spectroscopy. ^{1}H NMR spectra were recorded at 27 °C on a 400-MHz JEOL GX spectrometer fitted with an autosampler. Methanol-d_4 was used as the internal lock. Each spectrum consisted of 304 scans of 8192 complex data points with a spectral width of 5000 Hz, an acquisition time of 1.64 s, and a recycle delay of 2 s per scan. The pulse angle was 50°. A presaturation sequence was used to suppress the residual water signal with low power selective irradiation at the water frequency during the recycle delay. Spectra were Fourier transformed with 1 Hz line broadening, phased, and baseline corrected using the JEOL (Delta) software. Spectra were converted to Felix 2000 software format and saved as ASCII files. Spectra were further transferred to a personal computer for data analysis.

Multivariate Analysis. *A. Principal Component Analysis (PCA)*. The application of the technique to NMR data, the use of PC scores in discriminant analysis, and the interpretation of PC loadings have already been described in previous work (*6*).

B. Partial Least Squares (PLS). PLS is used here as an alternative data compression technique to PCA (*7*).

Univariate Analysis. ANOVA is a statistical technique that permits testing of the hypothesis that two or more groups of samples are drawn from the same population (*8*). Results of multivariate analyses can be difficult to interpret in terms of specific compounds so we have applied ANOVA to selected NMR signals to determine whether there are significant differences between mean concentrations of individual compounds in the transgenic and control groups.

Software. PCA was carried out in Matlab, version 5.3.1.29215.a (The MathWorks Inc, Natick, Massachusetts). For each NMR spectrum in Felix ASCII format, 5580 points were extracted from the original 8192 points using a PASCAL program written in-house. Parts of the spectrum that do not contain any signals were excluded (the region between points 1301 and 6880 was kept for chemometric analysis).

F values and box plots were calculated using the Matlab macro "anova1" (Statistics toolbox).

RESULTS AND DISCUSSION

Signal Assignments. **Figure 1**A shows the ^{1}H NMR spectrum of a typical red tomato from the transgenic series. The spectrum was thoroughly analyzed using 2D experiments. The combined information gathered from COSY and HOHAHA spectra and the use of a library of ^{1}H spectra of reference compounds have allowed an almost complete assignment, as shown in **Figure 1B−D**. The overall view in **Figure 1A** also illustrates the relative vertical scales of the expansions shown in **Figure 1B−D**. **Table 1** summarizes the chemical shift information available for tomato from the 2D spectra and the reference standards. NMR spectra of reference compounds were run in the same solvent mixture as was used for the tomato extracts. In cases where further confirmation of the assignment was required, the tomato samples were spiked with appropriate standards to confirm that the chemical shifts were identical. The COSY spectra were used to select the "best" chemical shifts for quantification of individual compounds, since 2D spectra are able to reveal overlapping/interfering peaks. In such complex spectra, however, it is not always possible to find positions that are completely free of interference.

The flavonoid content of these tomatoes has previously been analyzed (*5*), and the main flavonoids expected in the ^{1}H NMR spectrum of the transgenic were kaempferol-3-*O*-rutinoside (the main contributing compound to the increase in flavonoid levels), naringenin-7-*O*-glucoside, kaempferol-3,7-di-*O*-glucoside (**1**), kaempferol-3-*O*-rutinoside-7-*O*-glucoside (**2**), a dihydrokaempferol-*O*-hexoside (**3**), and kaempferol-3-*O*-glucoside. We first identified the signals of these compounds that may be detected in the spectra of whole tomato extracts.

A standard of naringenin-7-*O*-glucoside in 70% methanol-d_4/30% D_2O indicated that the apparent singlet at 6.22 ppm in the red transgenic tomato ^{1}H NMR spectrum could be attributed to the H-6 and H-8 protons of the compound. **1** and **2** were previously isolated and fully characterized by NMR (*5*). Their chemical shifts (in methanol-d_4) in the low-field region were almost identical [6.50 (H-6), 6.78 (H-8), 6.89 (H-3′) and 8.09 (H-2′)]. In methanol-d_4, the chemical shifts of kaempferol-3-*O*-rutinoside were 6.40 (H-6), 6.62 (H-8), 7.02 (H-3′), and 8.08 (H-2′), while in 70% methanol-d_4/30% D_2O they were 6.25 (H-6), 6.45 (H-8), 6.95 (H-3′), and 8.04 (H-2′). The chemical shifts of kaempferol-3-*O*-glucoside in the same mixed solvent were identical to those of the rutinoside analogue, except for H-6 (6.28) and H-8 (6.49), which were slightly more deshielded. The COSY spectrum of the red tomato identified three pairs of coupled doublets (at δ 6.20−6.26, 6.29−6.50, and 6.55−6.81). It has been shown that the presence of a sugar linkage at the 7-position moves the kaempferol H-6 and H-8 chemical shifts downfield by approximately 0.2 ppm (*9*); hence, the 6.55−6.81 pair would correspond to **1** and **2** and the 6.29−6.50 pair to kaempferol-3-*O*-glucoside and kaempferol-3-*O*-rutinoside.

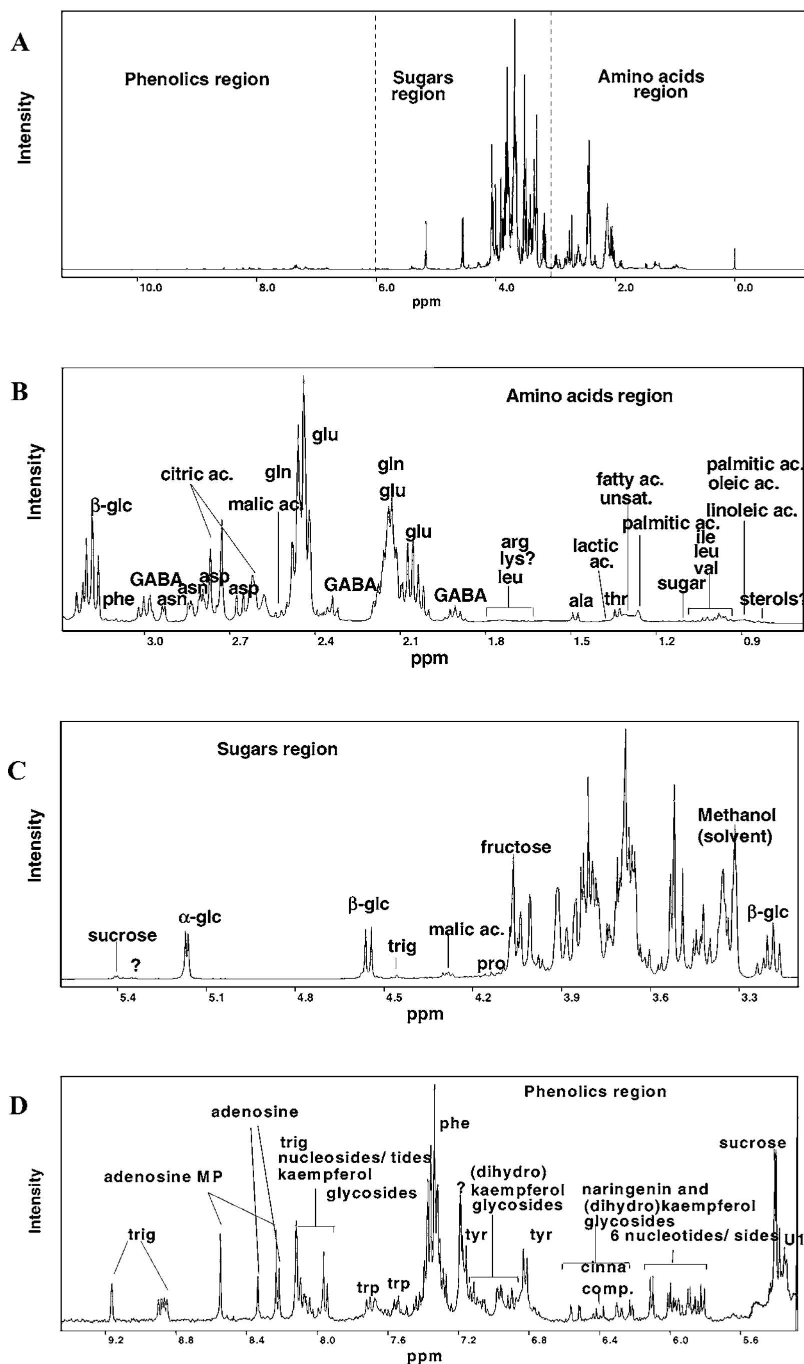

Figure 1. Details of ^{1}H NMR spectrum of a red modified tomato extract. Key: ac., acid; ile, isoleucine; leu, leucine; val, valine; unsat, unsaturated; ala, alanine; arg, arginine; lys, lysine; GABA, γ-aminobutyric acid; glu, glutamic acid; gln, glutamine; asp, aspartic acid; asn, asparagine; phe, phenylalanine; β-glc, β-glucose; pro, proline; cinna comp, cinnamic compounds; tyr, tyrosine; trp, tryptophane; trig, trigonelline, MP, monophosphate.

The H-6 and H-8 chemical shifts of the four kaempferol glycosides differ by approximately 0.2 ppm, but there is only a 0.06 ppm difference for the pair 6.20−6.26),indicating that the compound that gives rise to this pair is not a kaempferol-type molecule. The chemical shifts of protons H-6 and H-8 of dihydrokaempferol in methanol-d_4 are 5.93 and 5.88, respectively (*10*). With a 0.05 ppm difference between the two chemical shifts, there is a good indication that pair 6.20−6.26 may correspond to a dihydrokaempferol derivative. The chemical shifts of dihydrokaempferol-3-*O*-glucoside are very similar to those of the aglycon: 5.91 and 5.89, respectively (*10*), but those of the 7-*O*-glucoside, also in methanol-d_4, are more deshielded: 6.32 and 6.30, respectively (*11*). This is in agreement with the chemical shift trend mentioned above for H-6 and H-8 with a sugar linkage at position 7. The pair 6.20−6.26 is therefore most likely to correspond to **3**, the more abundant of the two dihydrokaempferol-hexosides identified by LC/MS (*5*), which appears to be a dihydrokaempferol-7-*O*-hexoside. The signals arising from its B ring protons are likely to be attributed to the pair 6.90−7.39 (**Table 1**), as this matches the signals previously found for dihydrokaempferol-7-*O*-glucoside (*11*).

Davies and Hobson (*12*) reported that between 13 and 38 μg g^{-1} (fresh weight) of chlorogenic acid, 97 μg g^{-1} of caffeic acid (as aglycon), and 16 μg g^{-1} of *p*-coumaric acid were found in tomatoes. However, Winter and Herrmann (*13*) and Fleuriet and Macheix (*14*) indicated that glycosides of caffeic and *p*-coumaric acids were at least as abundant as chlorogenic acid.

Table 1. ^{1}H Chemical Shifts of Compounds from 1-D and 2-D Spectra of a Red Modified Tomato Extract

compd[a]	chemical shifts (ppm)[b]								
sterols[c]	0.81	0.83							
palmitic acid only	0.86								
linolenic only	0.95								
all fatty acids	0.89	1.27	1.30	1.57	2.02	2.35	2.75	5.31	5.34
isoleucine	0.94	1.00	1.25	1.53	1.95	3.65			
leucine	0.95	0.98	1.75						
valine	0.99	1.04	2.27	3.50					
threonine	1.32	4.20							
lactic acid	1.35	4.20							
alanine	1.47	3.67							
arginine	1.71	1.91	3.23						
lysine	1.72	1.50	1.90	3.00	3.66				
γ-aminobutyric acid	1.89	2.32	3.00						
proline	2.05	2.45	4.10						
glutamine	2.12	2.43	3.67						
glutamic acid	2.03	2.12	2.41	3.66					
malic acid	2.50	2.75	4.27						
citric acid	2.59	2.74							
aspartic acid	2.62	2.80	3.81						
asparagine	2.78	2.94	3.90						
phenylalanine	3.05	3.30	3.87	7.32	7.38	7.42			
tyrosine	6.81	7.15							
tryptophan	7.09	7.11	7.53	7.70					
serine[c]	3.82	3.91							
rhamnose (glycoside)	1.08	3.41							
α-glucose	5.15								
β-glucose	4.53	3.17							
fructose	3.99	4.02	4.05						
U1	5.37								
sucrose	5.39								
uridine	4.17	4.22	5.80	5.88	7.94				
uridine-MP	4.29	5.95	5.89	8.03					
uridine-DP-glucose	3.44	4.29	5.60	5.93	5.97	7.97			
adenosine	4.73	5.99	8.21	8.32					
adenosine-MP	4.67	6.09	8.22	8.54					
guanosine	4.63	5.84	7.95						
cytidine	4.16	5.85	5.98	7.93					
cytidine MP	5.96	6.05	8.06						
cinnamic acid 1	6.38	7.47	caffeic acid?						
cinnamic acid 2	6.40	7.46	*p*-coumaric acid?						
chlorogenic acid	6.32	7.58							
naringenin glycoside	2.82	3.20	5.44	6.22					
dihydrokaempferol-7-*O*-hexoside[c]	6.20	6.26	6.90	7.39					
kaempferol-3-*O*-glucoside and rutinoside[c]	6.29	6.50	6.96	8.07					
compds **1** and **2**	6.55	6.81	6.97	8.08					
$H_{2',5',6'}$ system	7.06	7.20	7.51	quercetin glycoside (rutin?)					
trigonelline	4.46	8.08	8.85	8.88	9.17				

[a] Abbreviations: M; mono-; D, di-; P, phosphate; U, unknown. [b] Spectra referenced to methanol = 3.3 ppm. [c] Provisional assignment.

Of the three cinnamic compounds detected in the COSY spectrum (from the six doublets given by their olefinic protons, $J = 16$ Hz), one can be identified as chlorogenic acid and the other two as derivatives of caffeic and *p*-coumaric acids (**Table 1**).

A series of signals between 5.8 and 6.1 ppm consisted of two sets of doublets. The first set (all with $J \sim 5$ Hz) showed at least six COSY cross-peaks to signals in the range 4.1−4.75 ppm. These signals can be assigned to H-1 and H-2 of the ribose units of various nucleosides. The second set ($J \sim 8$ Hz) showed four cross-peaks to signals with shifts in the range 7.9−8.1 ppm. These signals arise from H-5 (5.9 ppm) and H-6 (8 ppm) in the base units of uridine and cytidine. Assignments of specific compounds were made by spiking the tomato extracts with reference standards (see **Table 1** for details). Several singlets in the downfield region could also be assigned to the base units of nucleosides or nucleotides with the help of spiking experiments. These were guanosine (7.95 ppm), adenosine (8.21 and 8.32 ppm), and AMP (8.22 and 8.54 ppm). A set of four signals at δ 9.17 (s), 8.88 (d), 8.85 (d), and 8.08 (dd) was shown to arise from a single compound by the COSY experiment and by correlated changes in the intensity of all the signals when comparing different samples. The pattern of these signals resembled that of niacin, a known constituent of tomato, but spiking experiments with nicotinic acid, nicotinamide, and various derivatives (NAD^+, $NADP^+$, nicotinate mononucleotide,

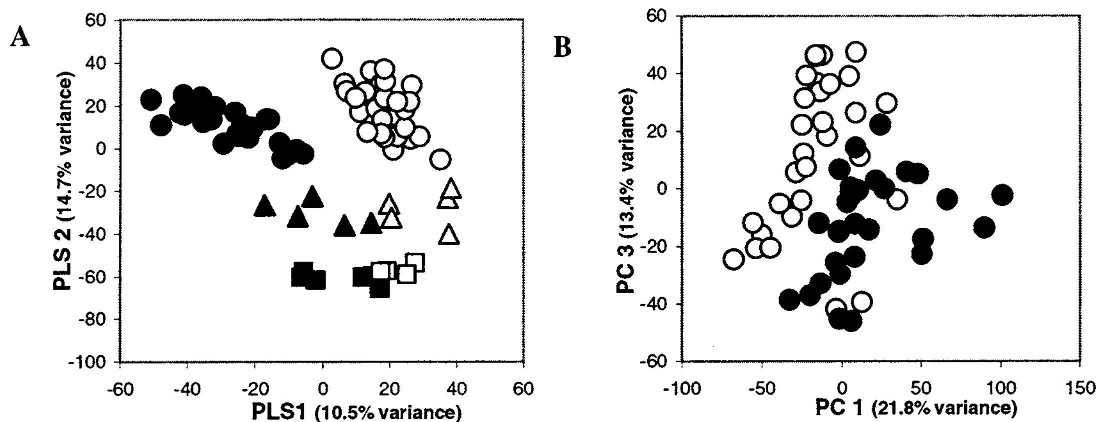

Figure 2. (A) First two PLS scores for 80 transgenic and control tomatoes at three stages of maturity and (B) PC1 and PC3 for 60 transgenic and control red tomatoes: ○, red transgenic; ●, red control; △, turning transgenic; ▲, turning control; □, green transgenic; ■, green control.

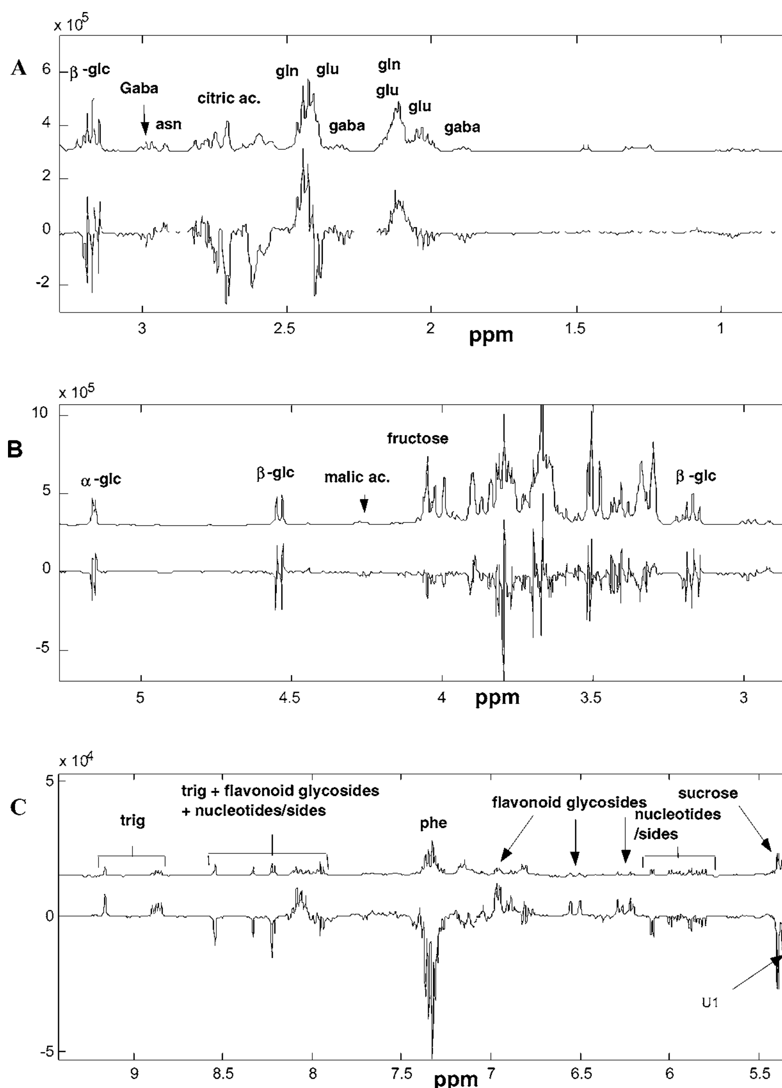

Figure 3. PLS loading 2 (lower line) and red transgenic mean spectrum (upper line), (A) high-, mid-, and (B) low-field regions. See Legend of **Figure 1** for key to assignments. Vertical scales refer to the loading trace.

etc.) showed that the observed chemical shifts did not coincide with any of the standards tested. A spiking experiment with trigonelline (*N*-methylnicotinic acid) then proved that it was the compound present in tomato, obviously at higher concentration than nicotinic acid or any of the related compounds. The four signals at 9.17, 8.88, 8.85, and 8.08 ppm were accompanied by a singlet at 4.46 ppm (*N*-methyl group).

Effects of Maturity on Metabolites. Both PCA and PLS compression techniques were applied to the NMR spectra of the 80 transgenic and control tomatoes covering the three stages of maturity—green, turning, and red. The PCA outputs are calculated without any class information input (PCA is an unsupervised technique, i.e., nonbiased). In contrast, the PLS-based data reduction technique does take into account the class

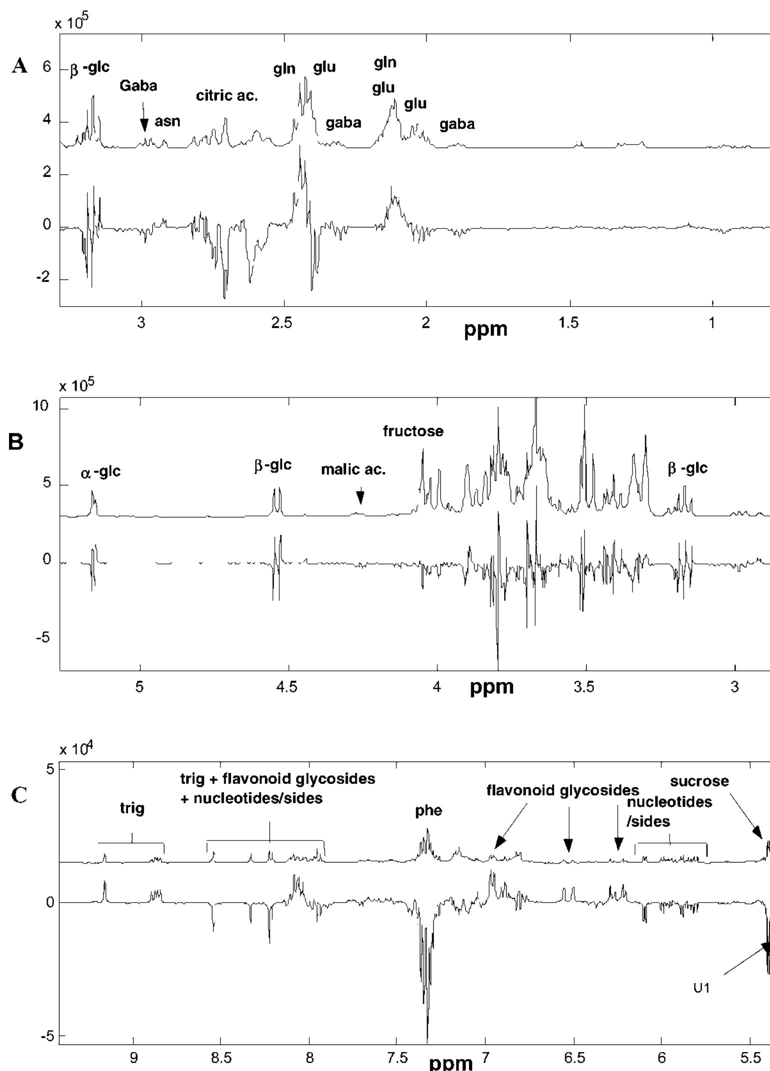

Figure 4. PLS loading 1 (lower line) and red transgenic mean spectrum (upper line), (A) high-, (B) mid-, and (C) low-field regions. See Legend of **Figure 1** for key to assignments. Vertical scales refer to the loading trace.

membership of the samples when carrying out the data compression. PLS is designed to maximize the differences between groups such that any differences are shown in the scores plots of the first factors, regardless of the amount of variance they express (*15*).

PLS clearly separated all six groups on the first two factors (**Figure 2**A), whereas a PCA analysis on the 60 red tomatoes displayed in **Figure 2B** showed a partial separation of transgenic and control groups on PC1 and PC3. In **Figure 2A**, most of the controls are located to the negative side of the PLS1 axis, while all the transgenic samples are located to the positive side. However, the difference between the two types of tomato sample is most clearly observed at the ripe stage. The division between transgenic and control groups at the turning and the green stages is not as clear along PLS1, and there are no systematic differences for the PCA scores (data not shown). This indicates that the levels of certain metabolites significantly diverge after the turning stage of maturity. This is in agreement with the conclusion reached from analysis of the flavonoid glycosides at the different ripening stages (*4*, *5*). The PLS2 axis relates mainly to changes of the metabolite content with ripening (the scores of the green tomatoes are located to the negative part of PLS2, while the scores become more positive through the turning and red stages). Since the scores are derived from the full NMR spectra, all of the compounds detected are potential contributors to the group separation that is observed (transgenic/control, ripe/nonripe). The scores plot on its own gives no indication of whether the transgenic/control separation results from "intended" or "unintended" effects or a combination of the two. A comparison of the appropriate PLS loading with the assigned NMR spectrum gives a first indication of which compounds contribute most to the separation in each case.

In the remainder of this section, the PLS2 loading will be used to indicate which metabolites change during ripening. Semiquantitative comparisons of amounts of individual compounds are made on the basis of the mean spectra of the different groups. Differences between transgenic and control tomatoes at the green and turning stages are also mentioned, where apparent. A statistical comparison is reserved for the red tomatoes in the following section, since many more red samples were available.

Loading 2 revealed that the green tomatoes contained on average more valine, isoleucine, leucine, γ-aminobutyric acid, malic acid, sucrose, phenylalanine, more of an unidentified compound (U1) at 5.37 ppm, and more chlorogenic acid (doublet at 6.32 ppm). The loading trace is negative at the position of those compounds' signals in **Figure 3A,B**. The levels of alanine, glutamic acid, fructose, the various nucleosides and nucleotides, flavonoid glycosides, and trigonelline appeared higher in the red tomatoes (positive loading trace at all these positions).

The mean spectra of green, turning, and red tomatoes (data not shown; transgenics and controls averaged separately giving six spectra in total) indicated that there was no detectable glutamic acid in the green tomatoes; its amount progressively increasing with the fruit ripening (no significant discrepancy was observed between control and transgenic). The high field region of the spectra showed that the levels of isoleucine at green and turning stages were comparable, dropping by 2-fold in the red tomatoes. Control levels were almost two times higher than in the transgenic at all ripening stages. This discrepancy was also observed for valine, which decreased in amount from green to turning to red by a factor of 2 at each step. The level of alanine decreased by a factor of 2 from green to turning but increased again in red tomatoes by 8-fold compared with turning. Control levels in green were again 2-fold higher than in the transgenic, but this discrepancy disappeared at the two subsequent stages. The levels of γ-aminobutyric and malic acids were 2 times higher in green than in the other tomatoes, with a slightly higher amount (less than 2-fold) in the control samples. The levels of asparagine and glutamine increased by a factor of 2 from green to turning but were unchanged at the red stage. Amounts of both compounds in transgenics were higher than in controls, but differences were less than 2-fold at all stages of maturity. Levels for the unknown compound U1 at 5.37 ppm (see **Figure 3B**) were highest at the turning stage and lower in red than in green tomatoes. Transgenics contained half as much of the compound compared with controls.

There was a constant decrease in the levels of sucrose (the amount halved at each stage) from green to red, and control levels were at least 2-fold higher than in the transgenic, the largest discrepancy being observed at the red stage. Glucose content remained constant from green to red stages for both types of tomato. The same was true for fructose, with the exception that in both types of green tomato the level was almost halved compared to the later stages of maturity.

The various nucleoside and nucleotide signals were hardly detected in the two types of green tomato, then their levels increased (2-fold between turning and red stages): control levels were slightly higher than those of the transgenic. Naringenin-7-*O*-glucoside and compounds **1** and **2** were not detected at the green stage but their amounts increased from green to turning to red (by 3-fold between the last two stages) for the transgenic fruits only. The same comment applies to kaempferol-3-*O*-glucoside and kaempferol-3-*O*-rutinoside and **3**, except that there was no change observed between turning and green stages.

A sharp decrease was observed for chlorogenic acid between green and turning stages (4-fold), the latter being barely detected in the ^{1}H NMR spectra of both types of red tomato. A slightly higher amount of chlorogenic acid was found in controls at the green stage, but the difference had disappeared at the turning stage. Davies and Hobson (*12*) reported that chlorogenic acid concentration falls during ripening while levels of caffeic and *p*-coumaric derivatives increase; therefore, this accords with the behavior of the doublet at 6.32 ppm, while the other two cinnamic compounds could correspond to caffeic and *p*-coumaric derivatives. The latter were hardly detected in green tomatoes. Their levels rose at the turning stage but dropped 2-fold when the fruits were ripe. A small but systematic discrepancy was observed between transgenics and controls with higher amounts in controls at both stages. The same level of phenylalanine was found in both types of green tomato, but this amount doubled at the turning stage in the control, remaining unchanged in the transgenic. The difference became even bigger in the red tomatoes, since the level in the transgenic dropped while the red control level remained constant (giving a final difference of 2−3-fold). Levels of trigonelline remained unchanged from green to red stages, but the transgenic tomatoes contained twice as much of the compound at all stages.

Table 2. ANOVA Results for Selected Signals from Red Transgenic and Control Tomato Spectra

	F value	
compd[a]	transgenic/control	order[b]
valine	16.1	c > t
rhamnose (glycoside)	324.9	t > c
threonine	0.3	
alanine	1.5	
arginine	14.8	c > t
γ-aminobutyric acid	9.9	c > t
glutamic acid	0.9	
glutamine	6.8	t > c
asparagine	7.2	t > c
citric acid	104.4	c > t
malic acid	cannot be aligned	
fructose furanose	1.7	
fructose pyranose	0.8	
α-glucose	3.3	
β-glucose	0.4	
trigonelline (4.46 ppm)	48.4	t > c
sucrose	56.2	c > t
U1 (singlet at 5.37 ppm)	21	c > t
uridine (5.88 ppm)	6.3	c > t
adenosine-MP (6.09 ppm)	6.1	c > t
cinnamic derivative 1	9.8	c > t
cinnamic derivative 2	1.2	
naringenin-7-*O*-glucoside	231.4	t > c
dihydrokaempferol-7-*O*-glucoside[c]	282.3	t > c
kaempferol-3-*O*-glucoside and rutinoside[c]	94.1	t > c
compounds **1** and **2**	170.9	t > c
phenylalanine	52.6	c > t
adenosine (8.32 ppm)	9.2	c > t
adenosine-MP (8.54 ppm)	29.4	c > t
trigonelline (9.17 ppm)	117.4	t > c

[a] Abbreviations: MP, monophosphate, U, unknown. [b] c, control; t, transgenic. [c] Provisional assignment.

Note that loading 2 was weighted at the levels of phenylalanine and trigonelline, because of the discrepancy between transgenic and control tomatoes at certain stages of maturity. Although, loading 2 relates mainly to the metabolite changes during ripening, it was actually the nature of the tomatoes and not their maturity that influenced the loading coefficient values in those two cases. This example shows the need for caution in an oversimplified interpretation of information from the loadings. Scrutinizing the mean spectra gives additional information and generally confirms the first indications given by the loading.

Overall green tomatoes contain less glutamic acid, less fructose, and in general less phenolic compounds than red ones. Some differences are observed between transgenic and control tomatoes at the preripe stages, but most of these do not exceed 2-fold, and some differences observed at the green stage later disappeared in the ripe fruits (e.g. alanine).

Effects of Genetic Manipulation on Metabolites of Red Tomatoes. As previously described, the scores of the red control tomatoes are located to the negative side of the PLS1 axis, while the transgenic ones are located to the positive side (**Figure 2A**). Loading 1 (**Figure 4**) showed that the controls on average contained more γ-aminobutyric, citric and malic acids, sucrose, phenylalanine, nucleosides and nucleotides, and more of the compound U1 at 5.37 ppm (negative loading trace) but less glutamine, asparagine, flavonoid glycosides, and trigonelline (positive loading trace) than the transgenics. Note that the

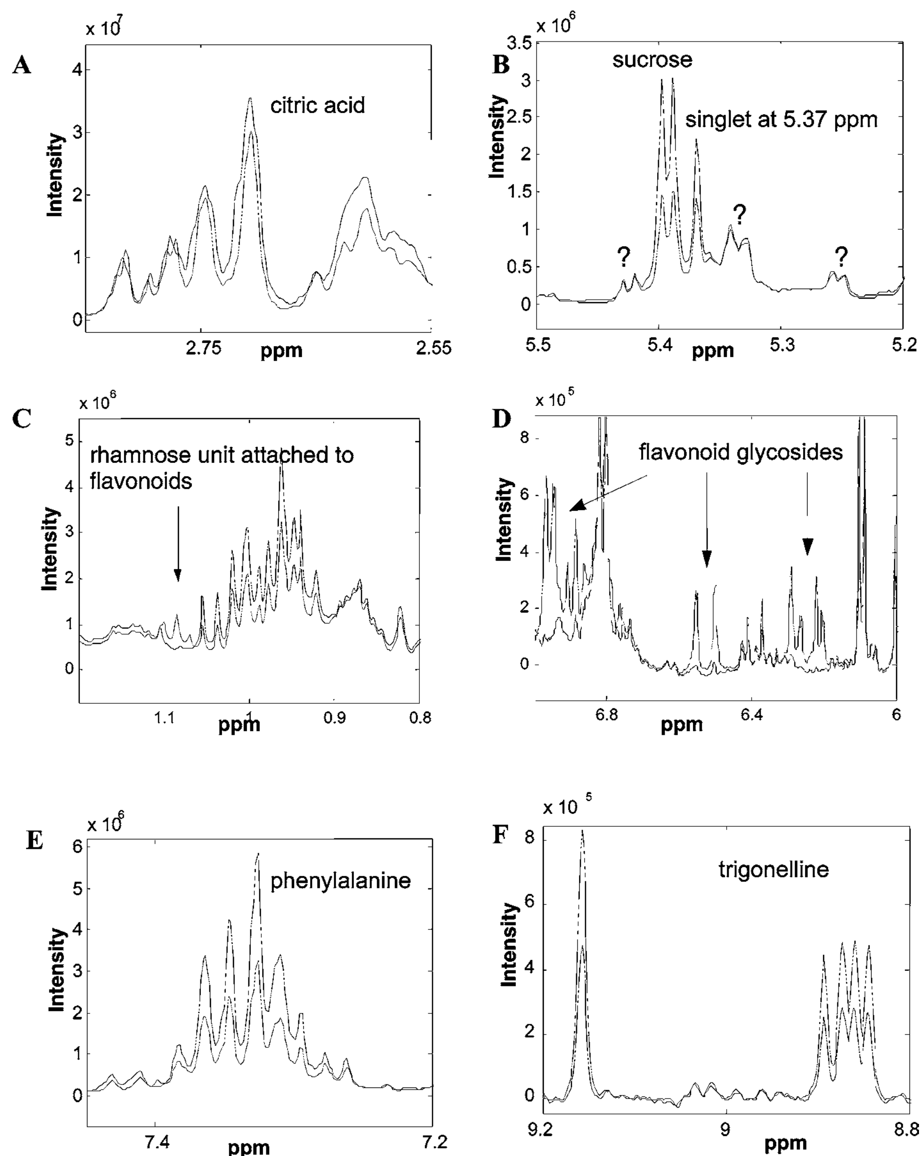

Figure 5. Details of the transgenic (dotted line) and control (solid line) tomato mean spectra: signals of a selection of metabolites.

loading trace was negative at the positions of glutamic acid and fructose in **Figure 4**, but as mentioned in the previous section, a look at the mean spectra showed that there was no real discrepancy between the control and transgenic levels; therefore, the loadings for these two compounds were not taken into account. Citric acid was not mentioned before because displacements of NMR signals between the spectra of tomatoes at the different stages of maturity were too great. [Citric acid chemical shifts are pH-sensitive and the buffer used could not be made strong enough to stabilize the pH completely across the range of samples studied.] However, considering just the red tomatoes, the signals were better aligned and loading 1 was clearly negative (**Figure 4A**). This showed that the average citric acid amount was higher in the controls.

To establish the significance of the differences observed between transgenics and controls, ANOVA was carried out on a selection of NMR metabolite signals. The calculation involves two groups each represented by a substantial number of samples (*3*). A baseline correction and an alignment program were applied to the selected peaks if necessary. Peaks were selected for ANOVA according to the level of difference previously detected from loadings and mean spectra. Additional peaks such as those of fructose or glucose were also included in the study in order to confirm their noninvolvement in the discrimination between red transgenic and control tomatoes. The degrees of freedom to consider for the calculation are 1 and 58. The critical F values for these degrees of freedom are $F_{0.05} = 4.0$, $F_{0.01} = 7.1$, and $F_{0.001} = 12.0$. It was decided not to consider as significant F values below 4 (**Table 2**).

As previously established, the levels of valine and γ-aminobutyric and citric acids, sucrose, nucleosides and nucleotides, phenylalanine, cinnamic derivative 1, and U1 were higher in control tomatoes (see F values for those compounds in **Table 2**). Similarly, ANOVA confirmed that transgenic tomatoes contain significantly more glutamine, asparagine, flavonoid glycosides, and trigonelline. Note that arginine (signal at 1.71 ppm) was not mentioned previously, as the loading did not point to any difference. However, as many compounds as possible were analyzed by ANOVA. It appeared that this result was significant, since comparison of the mean spectra (data not shown) depicted a small increase in amount of the compound in controls. It has not been possible to correctly align peaks of malic acid, but a slightly greater peak height was observed in the mean spectrum of the control tomatoes. As the levels of flavonoid glycosides were below detection level in the ^{1}H NMR of the control tomatoes, the calculated F values are somewhat

Table 3. List of Compounds with Approximate Changes in Mean Levels (red tomatoes, transgenic vs control)

compounds	degree of alteration	order[a]
valine, isoleucine, γ-amino butyric acid, arginine, malic acid, nucleotides/sides, cinnamic, and unknown U1 compounds	≤2-fold	c > t
glutamine, asparagine	≤2-fold	t > c
citric acid, phenylalanine, sucrose	2–3-fold	c > t
trigonelline	2–3-fold	t > c
kaempferol and naringenin glycosides	>10-fold	t > c

[a] c, control; t, transgenic.

arbitrary (the mean value of signals for those compounds in the control samples was actually the baseline). The *F* value for dihydrokaempferol-7-*O*-hexoside (**3**) is actually much higher than that of kaempferol-3-*O*-glucoside and kaempferol-3-*O*-rutinoside. This agrees with the fact that control tomatoes do contain some kaempferol-3-*O*-rutinoside but no dihydrokaempferol-7-*O*-hexoside (*5*). The rhamnose unit shown with a very high *F* value in **Table 2** belongs to the rutinoside units of **2** and kaempferol-3-*O*-rutinoside (the value was calculated from the rhamnose methyl group signal at 1.08 ppm). Threonine, alanine, glutamic acid, fructose, glucose, and cinnamic derivative 2 showed no significant difference between transgenics and controls.

In general, the compounds that showed differences of mean value not exceeding 2-fold have quite low *F* values (valine, arginine, γ-aminobutyric, glutamine, asparagine, the nucleosides and nucleotides, and U1). Details of the mean spectra of transgenic and control red tomatoes for compounds associated with larger *F* values (citric acid, sucrose, U1, phenylalanine, flavonoid glycosides, and trigonelline) are displayed in **Figure 5**. The difference in **Figure 5A** is somewhat smaller than the *F* value of citric acid suggests, but at least 2−3-fold differences are observed for the rest of the compounds selected (except for the targeted flavonoids, where larger differences were noted). These differences are statistically significant for the set of samples examined, but due to the limited dataset, they should be treated with caution. The two types of tomato plants were grown side-by-side in a glass house under hydroponic conditions with identical treatments (nutrients, light, etc.). Studies such as this one would benefit from the ability to assess the significance of differences observed within a wider context. It is well-known that metabolite contents can vary greatly according to parameters such as the soil nutrients, the climate, the season, etc. Noteborn et al. (*16*) found numerous significant differences ($P < 0.01$) between mean NMR amplitudes in a set of transgenic and control tomatoes grown under identical conditions, but the overwhelming majority of these cases showed less than 2-fold differences. Most were later found to be false positives when additional controls were analyzed.

NMR spectroscopy has been shown to provide, after an extensive assignment, a wealth of information about the main metabolites of the tomatoes studied. It has been possible to observe metabolic changes at different stages of maturity and, most interestingly, to detect both major and minor differences between the red transgenic and control tomatoes. This NMR analysis confirmed the main changes in metabolite levels already identified by HPLC, i.e., large increases (>10-fold) in naringenin- and kaempferol-glycosides. This is in line with current knowledge that maize *LC* and *C1* transcription factors specifically regulate flavonoid biosynthesis. Changes in other compounds were also identified. These were relatively minor (≤2−3-fold), although statistically significant. The main differences between transgenic and control red tomatoes are summarized in **Table 3**. This study shows that NMR combined with chemometrics and univariate statistics has a useful place in metabolomics research. Different profiling techniques are available (GC, HPLC, GC-MS, LC/NMR, and LC/MS) usually associated with greater sensitivity but needing more elaborate sample preparation and longer running times. ^{1}H NMR constitutes a consistent, quick, and informative screening technique.

ABBREVIATIONS

GC, gas chromatography; LC, liquid chromatography; NMR, nuclear magnetic resonance; MS, mass spectrometry; HPLC/DAD, high-performance liquid chromatography/diode array; *LC* and *C1*, *leaf color* and *colorless-1*; COSY, correlation spectroscopy; HOHAHA, homonuclear Hartmann-Hahn; PCA, principal component analysis; PLS, partial least squares; ANOVA, analysis of variance.

LITERATURE CITED

(1) Conner, A. J.; Jacobs, J. M. E. Genetic engineering of crops as potential source of genetic hazard in the human diet. *Mutation Res.* **1999**, *443*, 223−234
(2) 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.
(3) Bovy, A.; de Vos, R.; Kemper, M; Schijlen, E.; Almenar Pertejo, M.; Muir, S.; Collins, G.; Robinson, S.; Verhoeyen, M.; Hughes, S.; Santos-Buelga, C.; van Tunen, A. High-flavonol tomatoes resulting from heterologous expression of the maize transcription factor genes *LC* and *C1*. *Plant Cell* **2002**, *14*, 2509−2526.
(4) Muir, S. R.; Collins, G. J.; Robinson, S.; Hughes, S.; Bovy, A.; De Vos, C. H. R.; van Tunen, A. J.; Verhoeyen, M. E. Overexpression of petunia chalcone isomerase in tomato results in fruit containing increased levels of flavonols. *Nat. Biotechnol.* **2001**, *19*, 470−474.
(5) Le Gall, G.; DuPont, M. S.; Mellon, F. A.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E.; Colquhoun, I. J. Characterization and content of flavonoid glycosides in genetically modified tomato (*Lycopersicon esculentum*) fruits. *J. Agric. Food Chem.* **2003**, *51*, 2438−2446 (preceding paper in this issue).
(6) Le Gall, G.; Puaud, M.; Colquhoun, I. J. Discrimination between orange juice and pulp wash by H-1 nuclear magnetic resonance spectroscopy: Identification of marker compounds. *J. Agric. Food Chem.* **2001**, *49*, 580−588.
(7) Kemsley, E. K. Discriminant analysis of high-dimensional data: A comparison of principal components analysis and partial least squares data reduction methods. *Chemometrics Intell. Lab. Syst.* **1996**, *33*, 47−61.
(8) Wonnacott, T. H. W., Wonnacott, R. J. W. Introductory Statistics. In *Probability and Mathematical Statistics*, Wiley Series, Wiley: Chichester, UK, 1990.
(9) Budzianowski, J. Kaempferol glycosides from *Hosta-Ventricosa*. *Phytochemistry* **1990**, *29*, 3643−3647.
(10) Baderschneider, B.; Winterhalter, P. Isolation and characterization of novel benzoates, cinnamates, flavonoids, and lignans from Riesling wine and screening for antioxidant activity. *J. Agric. Food Chem.* **2001**, *49*, 2788−2798.
(11) Slimestad, R.; Andersen, O. M.; Francis, G. W. Ampelopsin 7-glucoside and other dihydroflavonol 7-glucosides from needles of *Picea-Abies*. *Phytochemistry* **1994**, *35*, 550−552.
(12) Davies, J. N.; Hobson, G. E. The constituents of tomato fruit−the influence of environment, nutrition, and genotype. *Crit. Rev. Food Sci. Nutr.* **1981**, *15*, 205−280.

(13) Winter, M.; Herrmann, K. Esters and glucosides of hydroxycinnamic acids in vegetables. *J. Agric. Food Chem.* **1986**, *34*, 616−620.

(14) Fleuriet, A.; Macheix, J. J. Derives glucoses des acides hydroxycinnamiques de la tomate: evolution au cours de la vie du fruit et formation in vitro. *Phytochemistry* **1980**, *19*, 1955−1958.

(15) Kemsley, E. K. *Discriminant analysis and class modelling of spectroscopic data*; Wiley: Chichester, UK, 1998.

(16) Noteborn, H. P. J. M.; Lommen, A.; Van Der Jagt, R. C.; Weseman, J. M. Chemical fingerprinting for the evaluation of unintended secondary metabolic changes in transgenic food crops. *J. Biotechnol.* **2000**, *77*, 103−114.

Received for review October 1, 2002. Revised manuscript received January 27, 2003. Accepted February 3, 2003. We thank the UK DEFRA Link program on Agro-Food Quality for funding (Project AFQ119).

JF0259967