JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

J. Agric. Food Chem. **2004,** *52,* 6075−6085 **6075**

NMR and HPLC-UV Profiling of Potatoes with Genetic Modifications to Metabolic Pathways

MARIANNE DEFERNEZ,[†] YVONNE M. GUNNING,[†] ADRIAN J. PARR,[†]
LOUISE V. T. SHEPHERD,[‡] HOWARD V. DAVIES,[‡] AND IAN J. COLQUHOUN*,[†]

Institute of Food Research, Norwich Research Park, Colney, Norwich NR4 7UA, United Kingdom, and Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, United Kingdom

Metabolite profiling has been carried out to assess the compositional changes occurring in potato tubers after genetic modifications have been made to different metabolic pathways. Most major features in the ^{1}H NMR and HPLC-UV profiles of tuber extracts have been assigned. About 40 GM lines and controls belonging to 4 groups of samples (derived from cv. Record or cv. Desirée and modified in primary carbon metabolism, starch synthesis, glycoprotein processing, or polyamine/ethylene metabolism) were analyzed. Differences were assessed at the level of whole profiles (by PCA) or individual compounds (by ANOVA). The most obvious differences seen in both NMR and HPLC-UV profiles were between the two varieties. There were also significant differences between two of the four Desirée GM lines with modified polyamine metabolism and their controls. Compounds notably affected were proline, trigonelline, and numerous phenolics. However, that modification gave rise to a very abnormal phenotype. Certain lines from the other groups had several compounds present in significantly higher or lower amounts compared to the control, but the differences in mean values amounted to no more than a 2−3-fold change: in the context of variability in the whole data set, such changes did not appear to be important.

KEYWORDS: ***Solanum tuberosum*** **L.; potato; metabolite profiling; metabolomics; NMR; HPLC; chemometrics; multivariate; transgenic; GMO**

INTRODUCTION

The safety assessment of genetically modified (GM) foods involves many aspects, but a compositional analysis of the food is always required. Any novel products intentionally introduced by the modification must be assessed individually for safety. The "substantial equivalence" approach has been adopted by regulatory bodies as the next step in the assessment of whole GM foods to ensure that they are as safe and nutritious as conventional counterparts (*1*, *2*). If the composition of the GM food, apart from those novel products, is not significantly different from that of comparable conventional foods, then the GM and conventional foods are considered to be substantially equivalent.

In substantial equivalence testing the GM food crop is grown side by side with the parental cultivar or other control in randomized plots, usually at more than one location. Then selected components are measured in GM and control plants and tested statistically for significant differences. Data from these field trials may also be compared with historical data for conventional varieties that are taken to indicate the safe range for each component. In practice, the analytes selected are the important nutrients plus specific antinutrients and toxicants known for that crop. The paper by Rogan et al. (*3*) provides an example of the substantial equivalence test applied to GM potatoes.

The selection of analytes represents an important decision. Efforts have been made to harmonize the procedure by compiling "consensus documents" for individual crops such as potato (*4*) that list the recommended analytes and the ranges of variation found in conventional varieties. However, there is a difficulty with predefined or targeted analyses that some unforeseen, unintended effects of the genetic modification may escape detection. Analysis by the emerging nontargeted molecular profiling ("-omics") techniques has been seen as one way of overcoming this difficulty (*5*). It should be emphasized that unintended effects also occur in conventional plant breeding and that they have traditionally been identified and eliminated by removing lines that show inferior characteristics (reduced yield, altered appearance, etc.) to established varieties. GM plants must meet the same criteria.

Almost all GM crops grown on a commercial scale at present are examples of "input trait" modifications. Novel gene products, for example, insecticidal proteins, have been introduced to enable changes in agricultural practices, generally with minimal interference to the plant's normal metabolism. The substantial

* Corresponding author (telephone +44 1603 255353; fax +44 1603 507723; e-mail ian.colquhoun@bbsrc.ac.uk).
[†] Institute of Food Research.
[‡] Scottish Crop Research Institute.

10.1021/jf049522e CCC: $27.50 © 2004 American Chemical Society
Published on Web 09/10/2004

Table 1. Description of Samples

variety	group	line description	line no.
Record	FK	1 control line, wild type	L1
		3 lines, 35S promoter, sense FK gene	L2–L4
		6 lines, 35S promoter, antisense FK gene	L5–L10
Desirée	W2GBSS	4 lines, GBSS promoter, sense W2 gene	L11–L14
		2 control lines, empty vector	L15, L16
	MAL1	7 lines, 35S promoter, antisense MAL1 gene	L17–L23
		1 line, 35S promoter, sense MAL1 gene	L24
		1 control line, empty vector	L25
	SAMDC	4 lines, 35S promoter, antisense SAMDC gene	L26, L27, L29, L30
	SAMTET	5 lines, 35S promoter, tet-repressor gene, antisense SAMDC gene	L32–L36
		1 control line, empty vector	L37
		1 control line, tissue culture	L38

equivalence approach is well suited for the safety assessment of such crops. There is, however, increasing development of crops, including potato, with "output trait" modifications, which have metabolite or biopolymer compositions altered in ways that cannot be readily achieved by conventional breeding. Output traits are manipulated by direct interventions in the plant's metabolism, for example, by introducing novel (foreign) enzymes or by changing the activity of endogenous enzymes.

Unintended effects can arise in crops with input or output trait modifications through disruption of host gene functions (the location of transgene insertion in the host genome is not controlled) or through somaclonal variation at the tissue culture stage of the transformation process. In addition, output trait modifications involve enzymes and substrates that are components of multibranched networks, so there may be widespread alterations in composition (both predictable and unpredictable) in response to an apparently simple intervention. Crops with altered output traits may contain secondary metabolites that would never be detected by substantial equivalence testing (*6*).

Our aim is to develop "nontargeted" molecular profiling methods that will be capable of detecting unintended effects in GMOs whether these involve the appearance of new compounds or changes in the amounts of existing ones. In this paper we report the application of NMR and HPLC metabolite profiling methods to four different groups of GM potato samples in which metabolic or developmental processes have been perturbed using sense or antisense transformation technologies. Choice of appropriate controls is also discussed. It is emphasized that the GM lines used are experimental ones that were selected to aid development of the profiling methodologies and are not intended for commercialization.

Several techniques have been developed for the nontargeted profiling of metabolites in plants. These include ^{1}H NMR (*7*, *8*) and HPLC-UV (*9*, *10*) as well as HPLC-MS (*11*), GC-MS (*12*, *13*), and direct injection FTMS (*14*). Present indications are that no single method will prove to be adequate if used in isolation. We used the first two of these methods to generate quantitative profiles, which proved to be largely complementary in terms of compounds detected. They were supplemented by HPLC-MS for compound identification. One of the important issues associated with metabolite profiling is the quantity of data that is generated. Our approach has been to use multivariate data analysis of whole NMR traces or extracted HPLC-UV peak intensities for an initial exploration of the data followed by univariate analyses of integrated NMR or HPLC peaks to confirm which compounds were mainly responsible for differences between GM and control samples.

MATERIALS AND METHODS

Plant Material. Transgenic lines selected for inclusion in the analyses have been developed at the Scottish Crop Research Institute (SCRI) over several years. GM material included has metabolic and/or developmental processes modified, in some cases very profoundly. By using such extremes we would expect unintended effects to emerge. All materials were planted as tubers, and the tubers were at least the second clonally propagated generation derived from the original transformation event. Where transgenic line numbers are provided, these indicate independent transgenic events with the constructs used. *Agrobacterium*-mediated transformation was used to generate the GM plants. **Table 1** summarizes the breadth of potato germplasm used in the study. Material includes two wild-type cultivars (cv. Desirée and cv. Record).

Record transgenics included three lines overexpressing fructokinase (FK) activity (sense) and six lines with FK activity down-regulated (antisense). The 35S CaMV promoter was used in all cases. Fructokinase phosphorylates fructose to fructose-6-phosphate, contributing to starch biosynthesis and glycolysis in tubers. Antisense lines show a 90% reduction in enzyme activity, and sense lines show an up to 3-fold increase in activity. ^{14}C labeling experiments indicate differential labeling of starch, albeit with modest changes in starch content. There is evidence of modified tuber numbers, but otherwise phenotypic changes are small (H. V. Davies et al., unpublished data).

For cv. Desirée, control lines included (a) wild-type tubers, (b) tubers generated from nontransgenic plants produced via tissue culture (which included a callus phase), and (c) transgenic tubers transformed with an "empty vector" construct—either vector pBIN19 (*15*) or its derivative pGPTV-Kan (*16*), containing the nptII gene but no target gene. It is well-known that tissue culture gives rise to somaclonal variation (*17*), a possible source of unintended effects independent of *Agrobacterium*-mediated gene transfer. Empty vector and tissue culture lines developed and grown alongside GM lines with targeted trait modifications represent important controls for vegetatively propagated plants when a comparison between matched homozygous and azygous plants is not possible.

Trangenic lines of cv. Desirée used were as follows:

(1) W2GBSS Series. These were transformed to express an *Aureobasidium* gene (designated W2) in sense orientation, driven by the granule-bound starch synthase (GBSS) promoter (*18*) and a plastid targeting sequence (*19*) in the binary vector pGPTV-Kan. The *W2* gene is derived from the filamentous fungus *Aureobasidium pullulans* and is believed to encode a glucan branching enzyme. Although the tubers show a waxy phenotype (high amylopectin content), this is not due to the *W2* gene, which is not expressed (M. A. Taylor et al., unpublished data). As the GBSS gene and protein are barely detectable in the W2GBSS tubers, it is considered that cosuppression of the endogenous *GBSS* gene is responsible.

(2) MAL1 Series. These contained the potato *MAL1* gene introduced in sense or antisense orientation under control of the 35S CaMV promoter in the binary vector pBIN19. Antisense lines show extremely stunted growth in the field, but less so when grown under containment.

The antisense lines have reduced activity of a glycoprotein-processing type II enzyme in tubers and show distinctive changes in leaf morphology caused by changes in cell wall structure (*20*). Sense lines have no obvious phenotype.

(3) SAMDC and SAMTET Series. These contained the potato *S*-adenosylmethionine decarboxylase (SAMDC) gene in antisense orientation under control of either the 35S CaMV promoter or the tetracycline-inducible promoter (SAMTET lines) (*21*), both in the binary vector pBIN19. The 35S CaMV lines show a stunted phenotype with reduced tuber numbers and dry matter content due to modified ethylene/polyamine metabolism following down-regulation of SAMDC activity. SAMTET lines were not treated with tetracycline and can therefore be considered as a "control".

General Sample Description and Preparation. Plants from two varieties of potatoes were used in this study: *Solanum tuberosum* L. cv. Record (10 lines) and cv. Desirée (26 lines). Four plants from each potato line were grown in compost in 30 cm^2 pots maintained under polytunnels at SCRI in 1999. Plants were grown through natural senescence, and tubers from each plant were harvested separately and maintained at 10 °C in the dark for 2 weeks prior to subsampling. For each plant average-sized tubers [usually between 80 and 100 g of fresh weight (FW) per tuber, depending on the line and construct] were cut in half longitudinally and each half cut longitudinally again. Sections were aligned and cut transversely to provide eight tuber sections overall. "Opposite eighths" were removed for freeze-drying. This sampling process minimizes variation due to the metabolite gradients that exist in tubers. Sufficient opposite eighths were combined to provide ~100 g FW of tuber. After cutting and selection, tuber material was immediately frozen in liquid nitrogen and maintained at −20 °C prior to freeze-drying. Freeze-dried material was powdered using a Retsch mill with a 1 mm sieve and stored in sealed containers at −20 °C in the dark until analyzed. This protocol provided one randomized replicate set of tuber samples for each plant.

Sample Preparation for NMR Spectroscopy. Each extract was prepared by the addition of 1 mL of 70% methanol-d_4/30% buffer (100 mM K_2HPO_4/KH_2PO_4, 1 mM TSP in D_2O) to 0.04 g of freeze-dried potato powder. The mixture was stirred at room temperature for 30 min and centrifuged at 10000 rpm for 10 min (Jouan A14 centrifuge). Each NMR sample consisted of 700 μL of the supernatant, which was stored in the NMR tube at −20 °C until required for analysis.

NMR Spectroscopy. ^{1}H NMR spectra were recorded at 27 °C on a 400 MHz JEOL GX spectrometer. Tuning of the spectrometer and manual shimming were carried out on the first sample before the start of the autosampler run for each batch of samples (~20 at a time). Methanol-d_4 was used as the internal lock. Each spectrum consisted of 300 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°. The receiver gain was set at the same value for all samples within the series. A presaturation sequence was used to suppress the residual water signal with low power selective irradiation at the water frequency during the recycle delay. Spectra were Fourier transformed with 1 Hz line broadening, phased, and baseline corrected using the 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.

Sample Preparation for HPLC. Powdered freeze-dried samples (0.12 g) were extracted with 0.5 mL of methanol containing 2 mM trifluoroacetic acid (TFA) at −20 °C for 15 min; a 0.5 mL aliquot of water precooled to 4 °C was then added, and the samples were left at −20 °C for a further 15 min with intermittent shaking. Extracts were then centrifuged (two times) at 13000 rpm in an MSE Micro Centaur benchtop centrifuge for 1 min, and the supernatants were analyzed directly by HPLC.

HPLC Analysis. HPLC was carried out on a Spectra-Physics SP8800/AS100 system complete with a SpectraFocus scanning UV detector (Spectra-Physics Analytical, Fremont, CA) set to 200−365 nm. Samples (50 μL) were injected using the filled loop mode onto a Columbus (Phenomonex Ltd., Macclesfield, U.K.) 5 μm C_{18} reverse phase column (250 mm × 4.6 mm) freshly equilibrated with 1 mM aqueous TFA and eluted with a gradient of increasing acetonitrile. The solvent profile was as follows (constant flow rate of 1 mL/min): 0 min, 100% 1 mM TFA, 0% acetonitrile; 40 min, 68% 1 mM TFA, 32% acetonitrile; 45 min, 50% 1 mM TFA, 50% acetonitrile; 50 min, 100% 1 mM TFA, 0% acetonitrile.

Quantitation of metabolites was based on peak area at 220 nm. Integration was carried out using PC1000 software from Thermo Separation Products (Riviera Beach, FL). This will directly export tables of results for data manipulation, although individual analyses were first scrutinized visually to ensure optimal peak detection.

HPLC-MS Analysis. HPLC-MS was carried out on a Micromass Quattro II triple-quadrupole mass spectrometer (Micromass, Manchester, U.K.) coupled to a Jasco PU-1585 triple-pump HPLC equipped with an AS-1559 cooled autoinjector, using HPLC conditions identical to those for HPLC-UV. The column eluent was passed through an ASI 600 fixed ratio splitter valve (Presearch, Hitchin, U.K.) and the majority sent to waste via a UV detector with the remaining 200 μL/min entering the mass spectrometer. Spectra were obtained in positive ion electrospray mode using a Micromass Z-spray ion source. The electrospray probe was operated at 3.5 kV and a cone voltage of 28 V. The source and desolvation temperatures were 140 and 350 °C, respectively. The nitrogen nebulizing and drying gas flow rates were optimized at 15 and 500 L/h, respectively. Spectra were recorded (in centroid mode) between m/z 50 and 1500 with a scan duration of 2 s/scan and an interscan time of 0.1 s. MS1 was set to unit mass resolution or better (LM and HM resolution parameters both set to 15.0).

Data Analysis Approaches. All profiling methods produce large quantities of data. Two broad approaches can be adopted to deal with this: either each compound is examined separately using univariate analysis, or the data for all compounds and all samples (or subsets of the two) can be analyzed simultaneously using multivariate analyses. We use both statistical approaches. The former broadly involves the study of the distribution of the compounds in the samples, for example, with histograms, box-and-whiskers plots, or comparisons between the distributions in various subgroups of samples. The main advantage of this approach is that it is conceptually easy, but potential limitations are that no account is taken of interactions between components and it can only be applied to signals that can be integrated. On the contrary, in multivariate analysis all compounds are taken into account at once, giving an overall view of the data. This can be a great advantage when very many compounds are being analyzed, even if the interpretation is sometimes less straightforward.

Principal Component Analysis (PCA). PCA (*22*) is one of the most useful techniques for analyses when the aim is to look for compositional similarities and explore the overall natural variability in a population of samples. The principle is to characterize each sample, not by directly analyzing every item of the original data but by transforming the data to a much smaller set of variables or PC scores. These new variables are combinations of the initial measurements but highlight the variance within the dataset and remove redundancies. Successive PCs account for decreasing amounts of variance, and most of the information is contained in the first few PCs. The explanation of what each PC represents in relation to the original measurements lies in the loadings, a set of weights given to each of the original measurements. There can be nontrivial issues related to the data preparation prior to using such techniques (*23*), and for the NMR spectra we have used whole (partly aligned) traces, whereas for the HPLC data we have used integrated values.

Data Analysis of NMR Spectra. The analyses were carried out using Matlab 6.1 (The Mathworks, Inc.) running on a desktop computer. For the multivariate analyses (i.e., PCA) regions that exhibited slight peak shifts (2.40−2.69, 2.72−2.80, 2.86−2.95, and 3.82−3.90 ppm) were aligned by a procedure based on that described by Vogels et al. (*24*, *25*). PCA was applied to the whole spectrum, except the baseline in the high- and low-field regions and the methanol signal (i.e., using regions 0.72−3.25 and 3.35−9.58 ppm, representing ~5740 datapoints). The average spectra of each triplicate measurement were used in the analyses, and the data were mean-centered prior to PCA. For the univariate analyses the peak height or area was calculated, box-and-whisker plots were drawn, and analysis of variance (ANOVA) was carried out.

Data Analysis of HPLC Chromatograms. The integrated peak areas of 33 compounds were collated in an Excel speadsheet, which

Table 2. ^{1}H NMR Chemical Shifts of Potato Extracts

compound	chemical shifts (ppm)								
fatty acid chain	0.86	1.25	1.57	2.03	2.15	2.33	2.75	5.34	
Ile	0.93	1.00	1.25	1.53	1.95	3.55			
Leu	0.96	1.63	1.74						
Val	0.99	1.04	2.27	3.48					
ethanol	1.17								
Thr	1.32	3.44	4.18						
Ala	1.47	3.66							
Lys	1.50	1.71	1.88	2.98	3.64				
Arg	1.71	1.88	3.22	3.65					
γ-aminobutyric acid	1.88	2.29	2.98						
U1	1.85	1.96	3.94						
quinic acid	1.88	1.98	4.06						
Pro	1.98	2.08	2.32	3.39	4.03				
Glu	2.02	2.14	2.39	3.63					
Gln	2.11	2.44	3.65						
pyroglutamic acid	2.06	2.34	2.43	4.09					
Met	2.06	2.18	2.63	3.74					
malic acid	2.37	2.68	4.26						
citric acid	2.53	2.71							
Asp	2.57	2.80	3.78						
Asn	2.75	2.94	3.88						
Tyr	2.97	3.20	3.79	6.81	7.15				
Phe	3.04	3.30	3.85	7.29	7.38				
choline	3.20								
Trp	3.21	3.50	3.93	7.09	7.16	7.25	7.42	7.70	
ascorbic acid	4.39								
trigonelline	4.44	8.08	8.84	8.86	9.15				
α-Glc	5.14	3.42							
β-Glc	4.53	3.16							
Suc	5.39	3.47	3.39	3.72	3.81	4.01	4.13	3.79	
U2	5.24	4.46	4.22	3.97					
U3	5.83	6.83							
UDP-Glc?	5.93	7.99							
caffeic acid	6.27	7.27	6.80	6.92	7.04				
chlorogenic acid	6.32	7.59	6.83	7.00	7.10	5.32	3.74	4.17	2.0
U4	6.40	7.65							
fumaric acid	6.54								
His	7.02	7.71							
U5	8.19	9.03	9.10	9.36					
formic acid?	8.48								

was then exported as an ASCII file to Matlab for multivariate (PCA) and univariate (box-and-whisker plots and ANOVA) analyses. In the univariate analysis missing data were simply discarded. In PCA, missing values were estimated: (i) for compounds measurable only in one variety, they were replaced by a value set to 20% of the minimum value observed for the other variety (this was done only in the PCA involving both varieties; in PCAs of individual varieties the compounds were discarded); (ii) otherwise, missing data were replaced by the average value for the compound of interest, calculated for samples of the same line when available or of the same construct if not. The variables were mean-centered and variance-scaled prior to PCA.

RESULTS

NMR Assignments. The use of NMR for metabolite profiling of potato extracts has not been reported before, although ^{1}H NMR has been used (*26*) to determine the degree of chain branching in starches isolated from potato (including GM varieties). Potato samples were prepared for ^{1}H NMR as described previously for tomato (*7*), and the NMR signals were assigned in the same way using a combination of 2D NMR experiments, comparison with spectra of reference standards, and spiking of extracts with test compounds. Chemical shifts of compounds that were identified are listed in **Table 2** together with a few unknowns that appeared consistently in the spectra. Many of the compounds identified are the same as those identified in tomato, but the overall appearance of the spectrum is quite different because of the different balance of components. In the high-field region (0.5−3.2 ppm) the spectra of ripe tomatoes are dominated by signals of Gln and Glu; in potato the signals of Val, γ-aminobutyric acid (Gaba), Glu, citric acid, and (especially) Asn are most prominent, together with a singlet at 3.20 ppm assigned to choline. In the mid-field region (3.2−5.5 ppm) the anomeric signals of the sugars show that the ratio of Glc to Suc is much reduced in potato compared with tomato (although the Glc level is quite variable). Two minor but clearly resolved signals were assigned to trigonelline (*N*-methyl group singlet at 4.44 ppm) and ascorbate (doublet at 4.39 ppm, J = 2.6 Hz). In the low-field region (5.5−10 ppm) the major signals are from Tyr and Phe, then at a lower level His and Trp, chlorogenic and caffeic acids, and in some samples a related unknown (U4, doublets at 6.40 and 7.65 ppm, J = 16 Hz indicating the presence of a trans double bond, possibly an isomer of chlorogenic acid). At a much lower level and detectable in only some samples were the aromatic signals of trigonelline and an unknown (U5) with a pattern of signals similar to that of trigonelline (8−9.4 ppm). Signals of nucleosides and nucleotides were much weaker than in tomato and hardly detectable in most samples.

Multivariate Analysis of NMR Data. The NMR spectra of each potato sample were recorded in triplicate. PCA was then used to explore the relative variability of different factors such as variety, line, type of genetic modification, and plant. NMR replicates were found to be highly reproducible, as the scores of replicate measurements were virtually superimposed. Hence, the analyses were repeated using average spectra of the NMR repeats in order to get a better signal-to-noise ratio and clearer graphs. A PCA encompassing both varieties and separate PCAs for each were carried out.

The PCA involving both varieties clearly highlights the differentiation between Record and Desirée potatoes, because the scores form two clusters separated on PC1 (**Figure 1a**). The loadings of the first PC were negative across the whole spectrum, and the PC1 scores were highly correlated with the overall NMR intensity (data not shown). Whereas both solvent (methanol) and reference (TSP) intensities were very reproducible between spectra, the total intensity over the rest of the spectra varied from sample to sample. However, it was consistent for independent extracts and NMR runs of a given sample, indicating that the NMR signal, and hence the concentration of compounds in the extract, was sample-dependent and inversely related to characteristics such as the percentage of dry nonextractable matter (mainly starch) in the samples. Total intensity for Desirée samples was generally higher than for Record.

Apart from this immediate overall concentration difference, there was also separation due to other factors between Record and Desirée potatoes in further PCs (see PC3/ PC4 scores in **Figure 1b**). A PCA carried out after area-normalizing the spectra (to compensate for the overall concentration difference) also showed a separation between Desirée and Record samples, this time on PC1/2 (data not shown). The difference between average spectra for Record and Desirée potatoes showed that many metabolites (e.g., Tyr, Glc, Fru, and Val) were present at higher concentration in the Desirée samples, but for some of the major metabolites (Asn, Gln, and citric acid) the difference was not great. Very few metabolites (trigonelline and choline) had a higher average concentration in Record than in Desirée.

Further investigation was carried out with the examination of the two varieties separately. Again, when using raw spectra, the first PC scores were strongly correlated to the overall NMR intensity and ultimately to the extract's concentration, again most likely due to differences in nonextractable matter. Area nor-

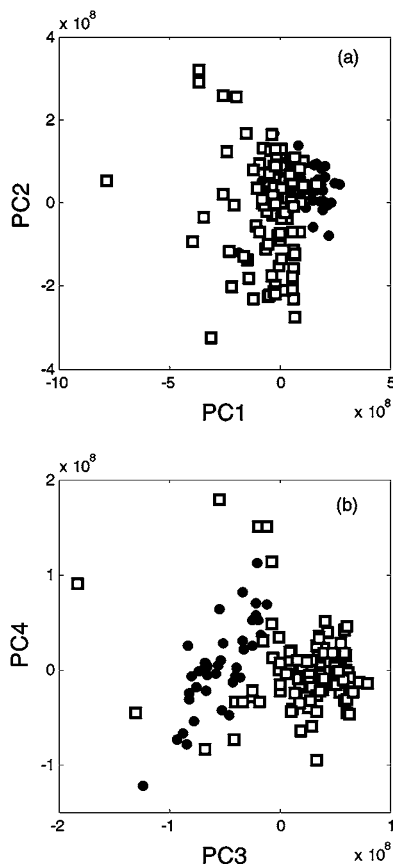

Figure 1. PCA of NMR spectra of Record (●) and Desirée (□) potatoes; scores on (**a**) PC1/PC2 and (**b**) PC3/PC4.

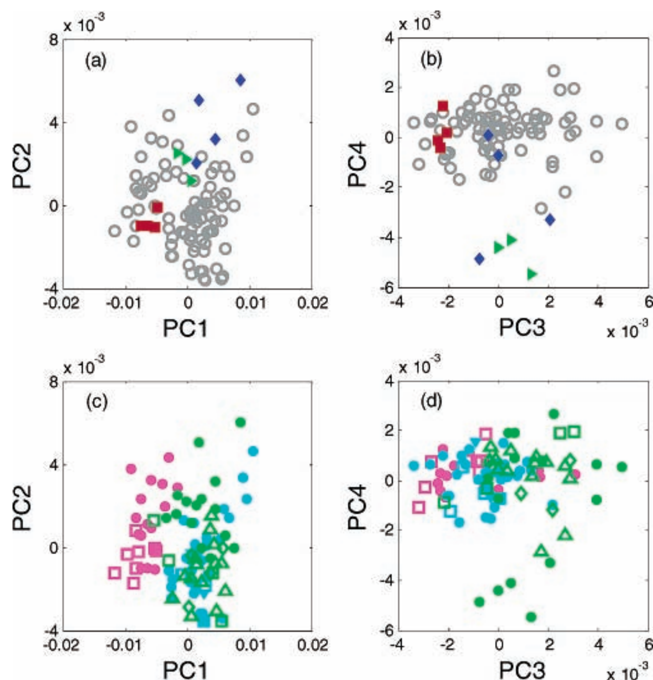

Figure 2. PCA of NMR spectra (area normalized) of Desirée potatoes; scores on PC1/PC2 and PC3/PC4 highlighting (**a**, **b**) lines L14 (solid red square), L29 (solid green triangles), and L30 (solid blue diamonds) and (**c**, **d**) the different constructs (W2GBSS-s, solid magenta circles; W2GBSS-ctrl ev, open magenta squares; MAL-a, solid blue circles; MAL-s, solid blue triangles; MAL-ctrl ev, open blue squares; SAMDC-a, solid green circles; SAMDC-ctrl tet rep, open green triangles; SAMDC-ctrl ev, open green diamonds; SAMDC-ctrl tc, open green squares).

malization could be used to lessen this effect. The PC score plots showed that the variation between replicate plants could be very different from one line to another: whereas the replicate plants were tightly clustered in some cases, the spread between replicate plants could also be as large as between lines in other cases (**Figure 2a**,**b**). Despite this, clustering of samples from the same modification was observed (**Figure 2c**,**d**); for instance, for the Desirée potatoes the W2GBSS samples, both modified and control, had lower PC1 scores than the MAL1 and SAMDC groups. The loading corresponding to PC1 suggested that the W2GBSS group contained higher levels of Glc and malic acid than the other groups but lower levels of other metabolites including Phe, Tyr, Asn, Gln, and Gaba.

A more detailed examination of the PC scores plots showed some clustering of the independent replicates from individual lines. One FK antisense line (L5) was separated from the Record WT, all of the FK sense lines, and the remaining FK antisense lines (PCA of Record samples only, data not shown). In the Desirée-only PCA the W2GBSS lines L11 and L13 were separated from L12, L14, and the two EV controls (L15 and L16) by their higher scores on the PC2 axis (**Figure 2c**). MAL1 L22 was separated from the other MAL1 antisense lines, sense line (L24), and the EV control (L25) also on PC2 (**Figure 2c**). SAMDC L29 and L30 were separated on PC4 and PC2, respectively, from a cluster containing SAMDC L26 and L27 plus the EV (L37), TC (L38), and SAMTET controls (see **Figure 2a**,**b** for identification of L29 and L30 replicates). Apart from the one MAL1 and two SAMDC lines indicated above, all of the other MAL1, SAMDC, and SAMTET lines and controls formed one cluster occupying roughly the bottom right quadrant in **Figure 2c**. The large number of groups involved together with the relatively small intergroup separations suggested that attempts to interpret the loadings for PC2 and the higher axes could be misleading. Instead, a univariate analysis of the NMR data was carried out for those compounds with resolved peaks that could be quantified with confidence. Almost all of the compounds that were thought to be of significance from an examination of the loadings were included.

Univariate Analysis of NMR Data. The NMR intensities of signals from 21 compounds were calculated for the four replicate plants of each line. For this, either the peak height or a sum of heights was used depending on how well separated the peaks were.

ANOVA was used to compare each line with its associated control, that is, with the WT (Record) or the EV line(s) that had undergone the same transformation and planting history as the GM line under examination (Desirée). It was observed that a limited number of samples showed statistically significant differences compared to their controls for at least some of the compounds (**Table 3**). For instance, within the Record samples, L5 has a number of compounds present in significantly higher concentrations than its control. Within the Desirée samples, ANOVA showed that L29 and L30 exhibited numerous significant differences from the EV control L37.

Box plots displaying one compound at a time for all lines simultaneously placed these differences in a wider context. They showed that not only were L29 and L30 different from their associated controls, but they were also very different from all other Desirée and Record lines; this is most strongly marked for the Pro signal, but also for other signals such as trigonelline as shown in **Figure 3**. It is interesting to note that these two lines had also been found to have PC scores different from those of all other samples, including other lines with the same construct, in the PCA of the whole NMR traces (**Figure 2**).

Table 3. Summary of ANOVA Results on Selected Compounds Based on NMR Peak or Multiplet Intensities[a]

line	compound																					
	Cga	Phe	Tyr	αGlc	βGlc	Trg	Asc	Mal	Suc	Asn	Cho	Cit	Gln	Glu	Pro	Gaba	Ala	Thr	Fa	Eth	Val	Ile
chemical shift for intensity evaluation, ppm	7.58	7.32	6.80	5.15	4.53	4.44	4.39	4.25	4.13	3.86	3.20	2.53	2.44	2.39	1.98	1.88	1.47	1.32	1.25	1.17	1.05	0.94
Record	FK Sense: Control Line 1 (WT)																					
2	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	+	•	•	•
3	•	•	•	•	•	•	•	•	•	•	•	•	•	+	•	•	•	•	•	•	•	•
4	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	FK Antisense: Control Line 1 (WT)																					
5	•	+	+	+	+	+	−	+	•	+	+	•	+	•	•	•	+	+	+	•	+	+
6	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	−	•	•	•
7	•	•	•	•	+	•	•	•	•	•	•	•	•	•	•	•	•	•	−	•	•	•
8	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	+	+	−	•	+	•
9	•	•	•	•	•	+	•	•	•	•	•	•	•	•	•	+	+	•	−	•	•	•
10	•	•	•	•	•	•	−	•	•	•	•	•	•	•	•	•	•	•	−	•	•	•
Desirée	W2GBSS: Control Mean of Lines 15 + 16 (EV)																					
11	-	•	•	•	•	+	+	+	•	+	+	•	+	+	•	+	+	+	•	•	•	•
12	•	•	•	•	•		+	•	•	•	•	−	+	+	+	•	+	+	•	•	•	•
13	+	•	+	•	•	+	+	+	+	+	+	−	+	+	+	+	•	•	+	•	•	•
14	-	•	•	−	−	•	•	•	•	•	•	•	•	•	•	•	•	•	−	•	•	•
	MAL1 Antisense: Control Line 25 (EV)																					
17	•	•	•	+	+	•	•	•	•	•	•	−	•	•	•	•	+	•	•	−	•	•
18	•	•	•	•	•	•	•	•	•	+	•	−	+	•	•	•	+	•	•	−	+	+
19	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	+	•	+	−	•	•
20	•	•	•	•	•	•	•	−	•	•	•	−	•	−	•	•	+	•	•	−	•	•
21	•	•	•	•	•	•	+	•	•	+	•	•	+	•	•	•	+	•	•	•	+	+
22	-	•	•	−	−	•	−	−	•	+	+	•	+	•	•	•	•	•	•	•	+	•
23	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	+	•	•	•	•	+
	MAL1 Sense: Control Line 25 (EV)																					
24	•	+	•	•	•	•	+	•	•	•	•	•	•	•	•	+	+	•	•	•	•	•
	SAMDC Antisense: Control Line 37 (EV)																					
26	•	•	•	•	•	+	•	•	•	+	•	•	•	•	•	•	•	+	•	−	•	•
27	+	•	+	•	•	+	•		•	•	•	•	•	•	•	•	•	•	•	−	•	•
29	+	•	•	•	•	+	•	+	+	+	+	•	•	•	+	+	•	+	+	−	+	+
30	+	•	−	•	•	+	−	•	•	+	+	•	+	+	+	•	•	+	+	−	•	+
	SAMTET (Tet-repressor + SAMDC Antisense): Control Line 37 (EV)																					
32	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
33	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	−	•	•
34	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
35	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	−	•	•	•	•	•	•
36	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•
	DES TC: Control Line 37 (EV)																					
38	•	•	•	•	•	•	•	•	•	•	•	•	•	•	•	−	•	•	•	•	•	•

[a] Differences from the appropriate control at the 5% significance level are indicated as + (increase), – (decrease), and • (no significant difference) with respect to the control. Cga, chlorogenic acid; Trg, trigonelline; Cho, choline; Fa, fatty acid.

They are two of a group of four antisense SAMDC lines and are clearly outliers. In contrast, the two other lines (L26 and L27) with the same construct showed fewer differences, and the SAMTET (antisense gene present but not expressed) and TC lines did not show a metabolite profile significantly different from that of the control. Box plots provide additional useful information. For example, **Table 3** indicates that all four SAMDC lines had levels of trigonelline that were significantly different from that of the EV control, L37. However, the box plot (**Figure 3b**) shows that the level of trigonelline was not unusually high in L26 and L27 when compared with all of the other samples (especially Record), whereas L29 and L30 had the highest levels recorded. Other lines that were shown by PCA to be slightly different from their controls and the other members of their group were L11, L13, and L22: all of these lines showed somewhat higher levels of choline, Asn, and Gln than comparable samples in their groups (**Table 3**).

In addition to the differences associated with these specific lines, the box plots also revealed differences that affected whole groups of samples comprising both controls and modified. For instance, Tyr was present in larger quantities (2−3 times) in Desirée than in Record, and the W2GBSS samples, both modified and control, contained higher amounts of Glc and malic acid and lower amounts of Gaba than other Desirée samples (confirming observations based on the PC loadings).

HPLC Peak Identification. Reverse phase HPLC with diode array detection resulted in chromatograms of moderate complexity, with some 40 peaks being readily observed at standard sensitivity (**Figure 4**). Closer examination revealed the presence of considerably more minor peaks, although not all peaks were fully resolved. Peaks were tentatively identified on the basis of the UV spectra, by reference to phytochemical literature on potatoes, and by comparison of retention times with those of commercially available standards. These identifications were then confirmed by mass spectrometry (either HPLC-MS or HPLC-MS/MS). The aromatic amino acids Tyr, Phe, and Trp were identified as significant components in the UV chromatogram, as were phenolics of various classes including hydroxycinnamate esters, hydroxycinnamate amides, and flavonoids (flavonols and anthocyanins). Nucleosides were observed at rather low levels, although they can be much more obvious in some solanaceous foods (e.g., tomato). A number of the more

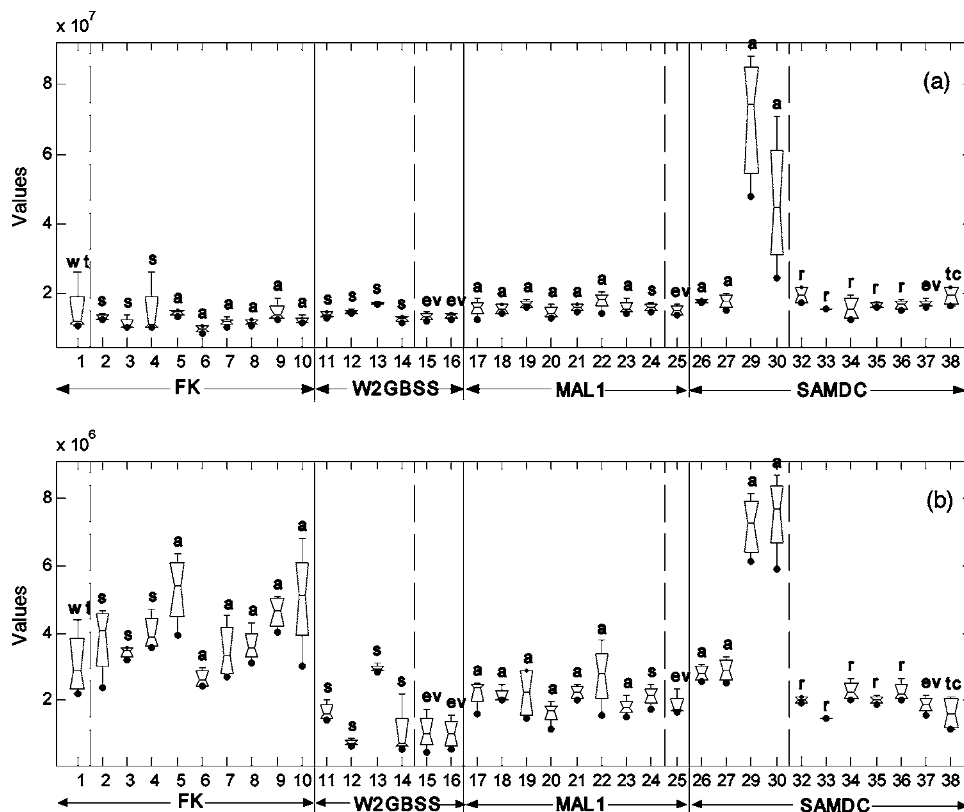

Figure 3. Box plot of the NMR signal intensities of Record and Desirée potatoes for (**a**) Pro and (**b**) trigonelline. The numbers on the *x*-axis are the line numbers.

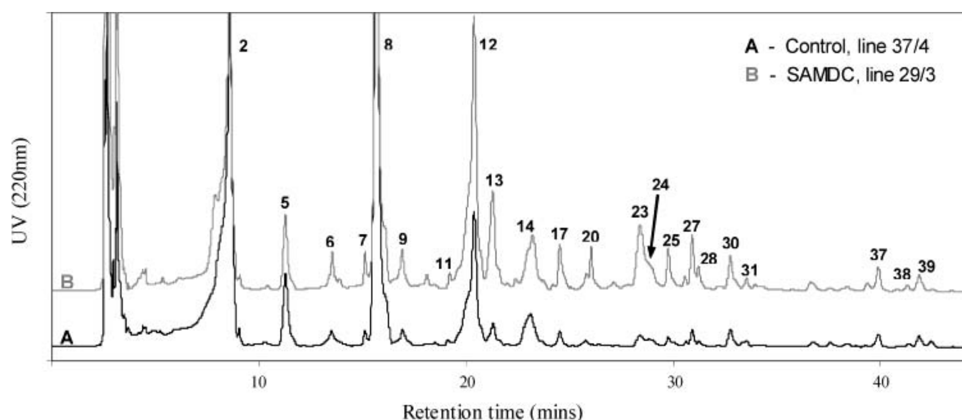

Figure 4. Comparison of HPLC traces (UV_{220nm}) from SAMDC antisense line L29 and control line L37. Numbering of compounds corresponds to codes in **Table 4**.

hydrophobic peaks showed high molecular weights (MW > 4000) and produced multiply charged ions during HPLC-MS; these may indicate polyphenolics or, perhaps more likely, peptides. Glycoalkaloids were not clearly visualized in the UV chromatograms, due to low UV absorbances and the presence of other materials in the same general area of the chromatogram. They were, however, very readily detected during HPLC-MS separations. The compounds' identification is summarized in **Table 4**, together with codes used in the analysis below.

Multivariate Analysis of HPLC Data. The integrated peak area of 33 compounds, 24 of which were at least in part identified, was collated for all of the samples, although some of the data were missing. In particular, four of the compounds, which were clearly present in the Record potatoes, had intensities that were too low for reliable quantitation (**100**−**103**) in Desirée. Conversely, seven compounds (**17**, **23**−**26**, **28**, and **31**) were clearly present in Desirée, but it was not possible to obtain any meaningful values for the Record potatoes (among others these included the anthocyanins—Record being both a pale-fleshed and pale-skinned variety). In addition to this, for the Desirée set 49 of the total of 2378 values (29 compounds × 82 samples) were missing due to the peaks not being well enough resolved for quantitation. For the Record set there were 81 such values of 910 (26 compounds × 35 samples). This value was higher than for Desirée because of the very minor nature of some of the peaks involved.

The data were analyzed by PCA to explore the variability and assess the variation within the set of samples as a whole. Both varieties were first analyzed together, and then separate analyses were also carried out. To lessen the influence of major peaks over that of more minor peaks, the variables were scaled by mean-centering and variance-scaling prior to PCA. There is no easy way to deal with missing values in PCA, as the whole matrix of data is subjected to the analysis. Hence, missing data

Table 4. HPLC Peak Identification (Peaks Are Listed in Order of Retention Time)

code	identity	ID method[a]	lit. ref
2	Tyr	A, C	*27*
4	[unknown]		
100	[unknown]		
5	Phe	A, C	*27*
6	caffeoylputrescine	A, C	*28*
7	bis(dihydrocaffeoyl)spermine	A	*b*
8	Trp	A, C	*27*
9	feruloylputrescine	C	*28*
10	3-*O*-caffeoyl quinate	C	*29*
11	bis(dihydrocaffeoyl)spermidine	A	*b*
12	chlorogenic acid	A, C	*29*
13	4-*O*-caffeoyl quinate	C	*29*
14	Caffeic acid	A, C	*29*
15	quercetin-GlcGlcRha	B	
16	[unknown]		
17	tris(dihydrocaffeoyl)spermine	A	*b*
19	ferulate derivative	B	
20	kaempferol-GlcGlcRha	B	
23	pelargonidin 3-*O*-(coumaroyl-rutinose)-5-*O*-glucoside	C	*30*
24	mixed anthocyanins, MW = 917	B	*30*, *31*
101	[unknown]		
25	anthocyanin-like, MW = 887	B	
102	[unknown]		
26	hydroxycinnamate	B	
27	ferulic acid amide, MW = 311?	B	
28	kaempferol-GlcRha	B	*30*
30	tris(dihydrocaffeoyl)spermidine	A	*b*
31	[unknown]		
103	[unknown]		
37	MW ~ 4235, polypeptide??		
38	MW ~ 4308, polypeptide??		
39	[unknown]		
40	[unknown]		

[a] Identification method: A, chromatographic and spectroscopic (UV, MS) properties as for pure standard; B, spectroscopic (UV, MS) properties indicative of this general class; C, as B, but more detailed assignment based on the known phytochemistry of potato (see refs quoted). [b] Parr, A. J. (unpublished data).

were replaced by the average value for the compound concerned, calculated for samples of the same line when available or of the same construct if not. Compounds that were not measurable for any of the Record potatoes were not included in the separate PCA analysis of this variety and were replaced by a small value in the PCA of both varieties together (value set to 20% of the minimum value observed for the Desirée potatoes). A similar treatment was applied for compounds not measurable for any of the Desirée potatoes.

The inclusion of both Desirée and Record potato samples in a PCA unequivocally discriminated the two varieties, which was in great part due to the compounds accurately quantitated in only one of the species (**Figure 5**). L29 and L30 were also clearly distinct from all other lines.

The first four PCA scores obtained by analyzing the Desirée samples only are shown in **Figure 6a**,**b**. The analysis clearly identifies L29 and L30 as different, as their scores on PC1 (39.9% variance) are much lower than those of all other samples. The W2GBSS group is also clearly separated from the other samples on PC2 (13.2% variance). In addition, SAMDC L26 and L27 are somewhat separated on PC3 and PC4 (8.9 and 5.3% variance, respectively), although this is not as clearly marked as for the above (samples in lower right quadrant of **Figure 6b**). Examination of the loadings suggests that the difference between lines L29 and L30 and other lines involves a number of compounds. As L29 and L30 have negative scores on the first PC, they contain higher amounts of the compounds with high, negative PC1 loadings. This includes various flavonoids (**20** and **28**), anthocyanins (**23** and **24**) and anthocyanin-like compounds (**25**), chlorogenic acid (**12**), 4-*O*-caffeoyl quinate (**13**), a ferulic acid derivative (**19**), a ferulic acid amide (**27**), dihydrocaffeoyl polyamine derivatives (**7**, **11**, **17**, and **30**), and compounds **37** and **39**. Compounds with high, positive loadings are present in lower amounts in L29 and L30 and include Tyr (**2**) and Trp (**8**). The loadings on the second PC suggest that the W2GBSS samples have higher quantities of a quercetin-based flavonoid (**15**), an anthocyanin (**24**), a hydroxycinnamate ester (**26**), and the unknown compounds **38** and **40**, whereas there are lower amounts of Tyr (**2**), Phe (**5**), caffeoyl putrescine (**6**), feruloyl putrescine (**9**), 3-*O*-caffeoyl quinate (**10**), and compound **31**. The compounds involved in the separation of SAMDC L26 and L27 are more difficult to identify, as their separation is less clear and spread on several PCs, but may include the phenolic polyamine conjugates **6** and **30**, the hydroxycinnamates **14** and **26**, and the uncharacterized compound labeled **31**.

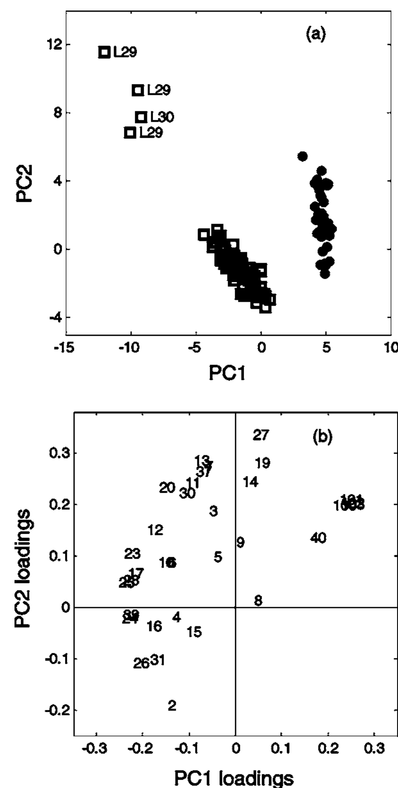

Figure 5. PCA of HPLC integrated intensities of Record (●) and Desirée (□) potatoes: (**a**) scores on PC1/2; (**b**) PC1 and PC2 loadings.

Univariate Analysis of HPLC Data. To look at the detail of the differences between lines, box-and-whisker plots were drawn for each of the 33 compounds separately. Note that for this, it was neither necessary nor desirable to replace missing values, and these were simply omitted. From this, a number of differences were noted either between the two varieties, for specific constructs (and in particular for the W2GBSS lines), or for specific lines (and again here in particular for L29 and L30). The compounds for which this was the case were in very good agreement with those suggested by the multivariate analysis, as all those listed by visual inspection of the box-and-whiskers plots had been listed in the PCA loadings. The box-and-whisker plots made it clear that, in addition to compounds measurable in only one of the varieties, Record and Desirée also differed clearly in their level of Tyr (**2**) and compound **16**

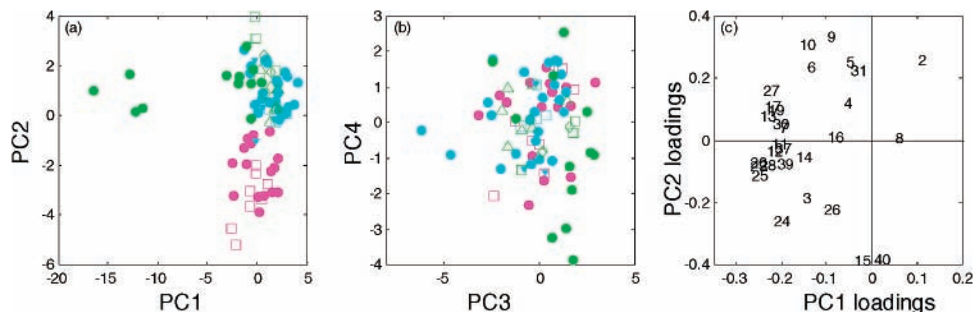

Figure 6. PCA of HPLC integrated intensities of Desirée potatoes: (**a**) scores on PC1/2; (**b**) scores on PC3/PC4 for the different constructs (W2GBSS-s, solid magenta circles; W2GBSS-ctrl ev, open magenta squares; MAL-a, solid blue circles; MAL-s, solid blue triangles; MAL-ctrl ev, open blue squares; SAMDC-a, solid green circles; SAMDC-ctrl tet rep, open green triangles; SAMDC-ctrl ev, open green diamonds; SAMDC-ctrl tc, open green squares); (**c**) PC1 and PC2 loadings.

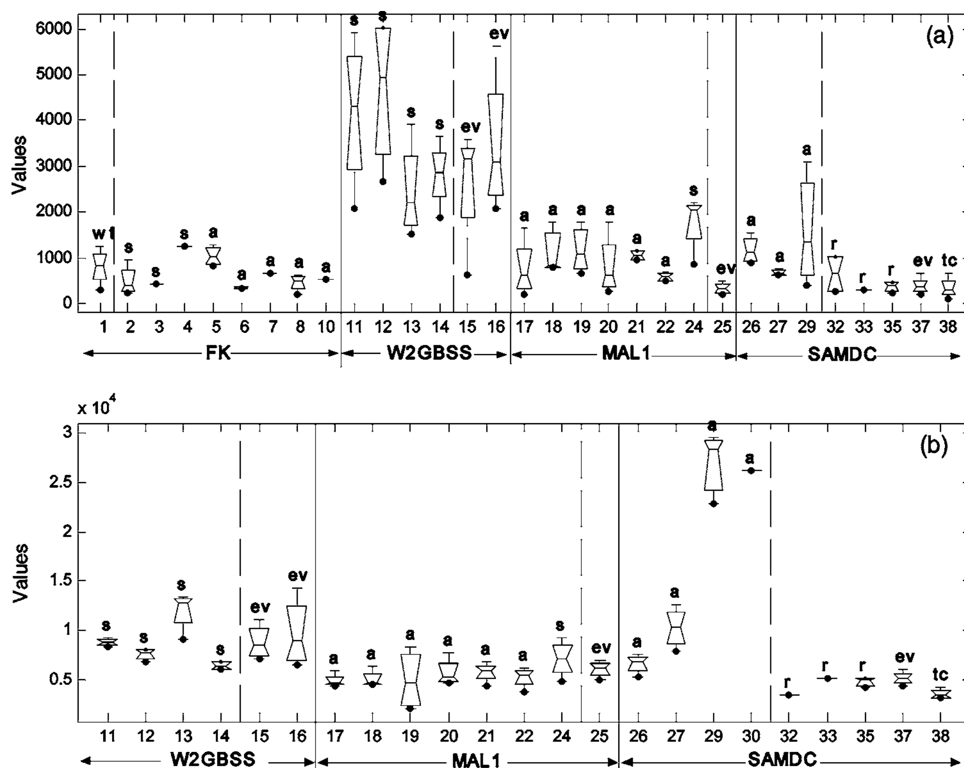

Figure 7. Box plot of the HPLC integrated intensities of potatoes for (**a**) quercetin-GlcGlcRha (**15**) and (**b**) pelargonidin 3-*O*-(coumaroyl-rutinose)-5-*O*-glucoside (**23**) (below detection limit in Record). The numbers on the *x*-axis are the line numbers.

(higher in Desirée), as well as the ferulic acid amide (**27**) and compound **40** (lower in Desirée). Within Desirée, the W2GBSS lines tended to have higher amounts of some flavonoids (**15**) and compound **40** and less feruloyl putrescine (**9**): this corroborates the PCA loadings. The differences noted for L29 and L30 also agreed with the PCA loadings.

ANOVA was applied to the HPLC results for the Desirée samples in the same way as described for the NMR data. The controls for each group were the same as indicated in **Table 3**. For L29 and L30 ∼20 of the 29 compounds measured by HPLC showed significant differences ($p < 0.05$) with respect to the EV control L37. Lines L11, L13 (W2GBSS), L22 (MAL1), and L26, L27 (SAMDC) each had 10−12 compounds in this category when compared with their controls, whereas all other lines typically had only 3 or 4 such compounds. Caffeoyl putrescine (**6**) and feruloyl putrescine (**9**) were the compounds that most frequently exhibited these differences in the W2GBSS and MAL1 groups. The same lines from the W2GBSS and MAL1 groups (L11, L13, and L22) were picked as most different from their controls by both HPLC and NMR, although there was little overlap in the compounds measured by the two techniques and the lines were much less obviously unusual than L29 and L30.

Two examples of the box plots are shown in **Figure 7**, but for space reasons it is not possible to present all of the results in this way. This demonstrates the usefulness of carrying out a multivariate analysis that is able to capture most of the information present in the data. However, a useful piece of information that is more readily available in the univariate analysis is the scale of the changes; here, even when most apparent, these usually did not exceed more than a 3−4-fold increase or decrease in mean value.

DISCUSSION

It might appear to be problematic that in two of the lines profiled by HPLC 20 of 29 compounds showed significant differences from the control. Even with this relatively modest number of compounds, the biological activity of most of the compounds has not been investigated, and some of them are still of unknown structure. These figures could be multiplied

in a more comprehensive metabolomics investigation. However, it should be remembered that the two SAMDC lines concerned showed a stunted phenotype and produced very small and elongated tubers (*21*). Plants from the SAMDC group would not have passed the primary "substantial equivalence" test based on agronomic performance and phenotype but were retained for this investigation so that a range of behaviors could be studied. It was pointed out that even for these two extreme lines the changes in levels of individual compounds compared with control were modest, and in the remaining two SAMDC lines the changes were even smaller, although these also had the same abnormal phenotype. Some of the compounds for which elevated levels were recorded in the GM samples (Pro, trigonelline, and choline) were perhaps indicative of a general response to osmotic stress. GM lines from within each of the FK, W2GBSS, and MAL1 groups showed relatively few compositional differences when compared with their controls. Within each group the different lines could not be distinguished phenotypically, but both profiling methods were sufficiently sensitive to pick one or two lines from each group that showed more differences than the rest. Some evidence was found for changes of composition when the empty vector controls associated with each group (which should nominally be equivalent) were compared. Metabolite profiling studies with additional empty vector and tissue culture only controls are required to test the hypothesis that somaclonal variation is a major source of compositional differences in GM potatoes. If nontargeted methods are to be made a part of the GM safety assessment process, extensive metabolomic data will also need to be collected for conventional varieties so that GM/control comparisons can be put in proper context. The fact that the largest differences seen in this study were found not between the GM potatoes and controls but between the two varieties involved only emphasizes this point.

ABBREVIATIONS USED

s, sense; a, antisense; wt, wild type; ev, empty vector; tc, tissue culture; ctrl, control; tet rep, tet repressor.

LITERATURE CITED

(1) Kuiper, H. A.; Kleter, G. A. The scientific basis for risk assessment and regulation of genetically modified foods. *Trends Food Sci. Technol.* **2003**, *14*, 277−293.

(2) Howlett, J.; Edwards, D. G.; Cockburn, A.; Hepburn, P.; Kleiner, J.; Knorr, D.; Kozianowski, G.; Müller, D.; Peijnenburg, A.; Perrin, I.; Pulsen, M.; Walker, R. The safety assessment of novel foods and concepts to determine their safety in use. *Int. J. Food Sci. Nutr.* **2003**, *54* (Suppl.), S1−S32.

(3) Rogan, G. J.; Bookout, J. T.; Duncan, D. R.; Fuchs, R. L.; Lavrik, P. B.; Love, S. L.; Mueth, M.; Olson, T.; Owens, E. D.; Raymond, P. J.; Zalewski, J. Compositional analysis of tubers from insect and virus resistant potato plants. *J. Agric. Food Chem.* **2000**, *48*, 5936−5945.

(4) OECD. Consensus document on compositional considerations for new varieties of potatoes: key food and feed nutrients, anti-nutrients and toxicants; 2002; http://www.olis.oecd.org/olis/2002doc.nsf/LinkTo/env-jm-mono(2002)5.

(5) 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.

(6) Firn, R. D.; Jones, C. G. Secondary metabolism and the risks of GMOs. *Nature* **1999**, *400*, 13−14.

(7) Le Gall, G.; Colquhoun, I. J.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E. Metabolite profiling of tomato (*Lycopersicon esculentum*) using ^{1}H NMR spectroscopy as a tool to detect potential unintended effects following a genetic modification. *J. Agric. Food Chem.* **2003**, *51*, 2447−2556.

(8) 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.

(9) Graham, T. L. A rapid, high resolution, high performance liquid chromatography profiling procedure for plant and microbial aromatic secondary metabolites. *Plant Physiol.* **1991**, *95*, 584−593.

(10) Fraser, P. D.; Pinto, M. E. S.; Holloway, D. E.; Bramley, P. M. Application of high-performance liquid chromatography with photodiode array detection to the metabolic profiling of plant isoprenoids *Plant J.* **2000**, *24*, 551−558.

(11) Tolstikov, V. V.; Fiehn, O. Analysis of highly polar compounds of plant origin: combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Anal. Biochem.* **2002**, *301*, 298−307.

(12) Roessner, U.; Wagner, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. Simultaneous analysis of metabolites in potato tuber by gas chromatography−mass spectrometry. *Plant J.* **2000**, *23*, 131−142.

(13) Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **2000**, *18*, 1157−1161.

(14) Aharoni, A.; de Vos, C. H. R.; Verhoeven, H. A.; Maliepaard, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B. Nontargeted metabolome analysis by use of Fourier Transform Ion Cyclotron Mass Spectrometry. *OMICS* **2002**, *6*, 217−234.

(15) Bevan, M. Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acids Res.* **1984**, *12*, 8711−8721.

(16) Becker, D.; Kemper, E.; Schell, J.; Masterson, R. New plant binary vectors with selectable markers located proximal to the left T-DNA border. *Plant Mol. Biol.* **1992**, *20*, 1195−1197.

(17) Davies, H. V. Commercial developments with transgenic potato. In *Fruit and Vegetable Biotechnology*; Valpuesta, V., Ed.; Woodhead Publishing: Cambridge, U.K., 2002; pp 222−249.

(18) van der Leij, F. R.; Visser, R. G. F.; Ponstein, A. S.; Jacobsen, E.; Feenstra, W. J. Sequence of the structural gene for granule-bound starch synthase of potato (*Solanum tuberosum* L.) and evidence for a single point deletion in the *amf* allele. *Mol. Gen. Genet.* **1991**, *28*, 240−248.

(19) Chambers, S. P.; Prior, S. E.; Barstow, D. A.; Minton, N. P. The pmTL nic-cloning vectors. I. Improved pUC polylinker regions to facilitate the use of sonicated DNA for nucleotide sequencing. *Gene* **1988**, *68*, 139−149.

(20) Taylor, M. A.; Ross, H. A.; McRae, D.; Stewart, D.; Roberts, I.; Duncan, G.; Wright, F.; Millam, S.; Davies, H. V. A potato α-glucosidase gene encodes a glycoprotein-processing α-glucosidase II-like activity and effects of down-regulation in transgenic plants. *Plant J.* **2000**, *24*, 305−316.

(21) Kumar, A.; Taylor, M. A.; Mad Arif, S. A.; Davies, H. V. Potato plants expressing antisense and sense *S*-adenosylmethionine decarboxylase (SAMDC) transgenes show altered levels of polyamines and ethylene: antisense plants display abnormal phenotypes. *Plant J.* **1996**, *9*, 147−158.

(22) Krzanowski, W. J. *Principles of Multivariate Analysis: A User's Perspective*; Oxford University Press: New York, 1988.

(23) Defernez, M.; Colquhoun, I. J. Factors affecting the robustness of metabolite fingerprinting using ^{1}H-NMR spectra. *Phytochemistry* **2003**, *62*, 1009−1017.

(24) Vogels, J. T. W. E.; Tas, A. C.; van den Berg, F.; van der Greef, J. A new method for classification of wines based on proton and carbon-13 NMR spectroscopy in combination with pattern recognition techniques. *Chem. Intell. Lab. Syst.* **1993**, *21*, 249−258.

(25) Vogels, J. T. W. E.; Terwel, L.; Tas, A. C.; van den Berg, F.; Dukel, F.; van der Greef, J. Detection of adulteration in orange juice by new screening method using proton NMR spectroscopy in combination with pattern recognition techniques. *J. Agric. Food Chem.* **1996**, *44*, 175−180.

(26) Nilsson, G. S.; Bergquist, K. E.; Nilsson, U.; Gorton, L. Determination of the degree of branching in normal and amylopectin type potato starch with ^{1}H-NMR spectroscopy. *Starch/Staerke* **1996**, *48*, 352−357.

(27) Davies, A. M. C. The free amino acids of tubers of potato varieties grown in England and Ireland. *Potato Res.* **1977**, *20*, 9−21.

(28) Leubner-Metzger, G.; Amrhein, N. The distribution of hydroxycinnamoylputrescines in different organs of *Solanum tuberosum* and other Solanaceous species. *Phytochemistry* **1993**, *32*, 551−556.

(29) Brandl, W.; Herrmann, K. Uber das Vorkommen der Chlorogensauren in der Kartoffel. *Z. Lebensm. Unters. Forsch.* **1984**, *178*, 192−194.

(30) Lewis, C. E.; Walker, J. R. L.; Lancaster, J. E.; Sutton, K. H. Determination of anthocyanins, flavonoids and phenolic acids in potatoes. I: coloured cultivars of *Solanum tuberosum* L. *J. Sci. Food Agric.* **1998**, *77*, 45−57.

(31) Giusti, M. M.; Rodriguez-Saona, L. E.; Griffin, D.; Wrolstad, R. E. Electrospray and tandem mass spectrometry as tools for anthocyanin characterization. *J. Agric. Food Chem*. **1999**, *47*, 4657−4664.

Received for review March 24, 2004. Revised manuscript received July 22, 2004. Accepted August 1, 2004. We acknowledge financial support from the EU 5th Framework Program (Project QLK1-1999-00765, ‘GMOCARE’).

JF049522E