DOI:10.1021/jf903104b

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

A Metabolomics Study of Cultivated Potato (*Solanum tuberosum*) Groups Andigena, Phureja, Stenotomum, and Tuberosum Using Gas Chromatography−Mass Spectrometry

Gary Dobson,*,† Tom Shepherd,† Susan R. Verrall,† Wynne D. Griffiths,† Gavin Ramsay,† James W. McNicol,‡ Howard V. Davies,† and Derek Stewart†

†Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, U.K. and ‡Biomathematics & Statistics Scotland, Dundee Unit, Dundee DD2 5DA, Scotland, U.K.

Phytochemical diversity was examined by gas chromatography−mass spectrometry in tubers of genotypes belonging to groups Andigena, Phureja, Stenotomum, and Tuberosum of the potato, *Solanum tuberosum*. Polar extracts (mainly amino acids, organic acids, sugars, and sugar alcohols) and nonpolar extracts (mainly fatty acids, fatty alcohols, and sterols) were examined. There was a large range in levels of metabolites, including those such as asparagine, fructose, and glucose, that are important to tuber quality, offering considerable scope for selecting germplasm for breeding programmes. There were significant differences in the levels of many metabolites among the groups. The metabolite profiles of genotypes belonging to Phureja and Stenotomum were similar and different from those of Tuberosum and the majority of Andigena genotypes. There was some agreement with the phylogeny of the groups in that Stenotomum is believed to be the ancestor of Phureja and they are both distinct from Tuberosum. Andigena genotypes could be partially distinguished according to geographical origin, Bolivian genotypes being particularly distinct from those from Ecuador. Biosynthetic links between metabolites were explored by performing pairwise correlations of all metabolites. The significance of some expected and unexpected strong correlations between many amino acids (e.g., between isoleucine, lysine, valine, and other amino acids) and between several nonpolar metabolites (e.g., between many fatty acids) is discussed. For polar metabolites, correlation analysis gave essentially similar results irrespective of whether the whole data set, only Andigena genotypes, or only Phureja genotypes were used. In contrast, for the nonpolar metabolites, Andigena only and Phureja only data sets resulted in weaker and stronger correlations, respectively, compared to the whole data set, and may suggest differences in the biochemistry of the two groups, although the interpretation should be viewed with some caution.

KEYWORDS: Gas chromatography−mass spectrometry; metabolic profiling; metabolomics; potato tuber; *Solanum tuberosum*.

INTRODUCTION

We are exploring the range of phytochemical diversity within potato (*Solanum tuberosum*) germplasm with a view to improving nutritional value and organoleptic properties such as texture, flavor, and taste. To this end, we have used metabolomics and volatile analysis approaches to further our understanding of the favorable organoleptic attributes of tubers of *S. tuberosum* group Phureja (*1*) compared to established cultivars of *S. tuberosum* group Tuberosum (*2*). Recently, we studied phytochemical diversity within a range cultivars and landraces belonging mainly to group Tuberosum using a gas chromatography−mass spectrometry (GC-MS) metabolomics approach (*3*).

GC-MS is a useful tool in metabolomics and is applicable to low molecular weight compounds, mainly those involved in primary metabolism. A GC-MS method, developed for analyzing polar metabolites (amino acids, aromatic amines, organic acids, sugars, and sugar alcohols) in fresh potato tubers (*4*), has been applied to studying the effects of environmental and genetic modifications in sucrose and starch metabolism (*4*−*7*) and for determining whether tubers from genetically modified plants were comparable to conventional cultivars (*8*). A GC-MS method for measuring polar and nonpolar (mainly fatty acids, alkanols and sterols) metabolites in freeze-dried (FD) tuber material has been validated and the limitations examined (*9*). In our studies on potato phytochemical diversity, we have found GC-MS to be a useful tool for highlighting differences between Phureja and Tuberosum (*2*) and among different Tuberosum cultivars and landraces (*3*).

Liquid chromatography−electrospray ionization−mass spectrometry (LC-ESI-MS) is applicable to higher molecular weight metabolites and is therefore complementary to GC-MS in being

*Corresponding author. Phone: +44 1382 562731. Fax: +44 1382 562426. E-mail: Gary.Dobson@scri.ac.uk.

 Published on Web 12/22/2009

able to detect many secondary metabolites (*10*, *11*). High mass accuracy can be achieved with the use of time-of -flight (TOF) detectors that make identification of unknown compounds feasible (*10*), and the use of hydrophilic interaction chromatography, in addition to more conventional reversed-phase chromatography, extends the range of detectable metabolites to highly polar metabolites such as oligosaccharides and sugar nucleotides. (*11*).

Within *S. tuberosum* there are four domesticated *S. tuberosum* groups, Andigena, Phureja, Stenotomum, and Tuberosum, that are cultivated in South America. In contrast to Tuberosum, Andigena, Phureja, and Stenotomum are all adapted to short-day flowering and tuberization (*12*) but Phureja is less strictly dependent on short days as it is adapted for growing in two or three cropping cycles a year. Like Tuberosum, Andigena is tetraploid, whereas Phureja and Stenotomum are diploid. Phureja is distinguished from Stenotomum in that the tubers of Phureja are less dormant and often sprout at harvest or shortly after (*12*). Our earlier phytochemical studies have focused on established European cultivars belonging to group Tuberosum (*2*, *3*), some Chilean landraces (*3*) and some Phureja genotypes (*2*). The four domesticated groups represent a greater genetic diversity of germplasm and may also be expected to exhibit a greater phytochemical diversity.

In the present study, we have used GC-MS to examine the polar and nonpolar metabolites in tubers from a range of genotypes from four cultivated *S. tuberosum* groups (Andigena, Phureja, Stenotomum, and Tuberosum) in the Commonwealth Potato Collection (CPC). The aims were to examine phytochemical diversity, determine differences among the groups and due to geographical origin, and finally to use the data to investigate biochemical linkages between metabolites.

MATERIALS AND METHODS

Plant Material. Seedlings of accessions from the tetrapoid *S. tuberosum* groups Tuberosum and Andigena and diploid *S. tuberosum* groups Phureja and Stenotomum were grown in a glasshouse at Dundee, Scotland, under natural daylength. The number of accessions from each group and geographical origin were as follows: Andigena [46 accessions; Peru (20), Bolivia (9), Ecuador (9), Colombia (3), Venezuela (2), Argentina (1), unknown origin (2)], Phureja [25 accessions; Colombia (16), Bolivia (3), Ecuador (1), Bolivia × Colombia (1), Colombia × Ecuador (1), Peru × Bolivia (1), Venezuela (1), unknown origin (1)], Stenotomum [13 accessions; Bolivia (4), Peru (4), Argentina × Bolivia (1), Bolivia × Peru (1), unknown origin (3)] and Tuberosum [6 accessions; Chile (5), Chile × Bolivia (1)]. Stenotomum and Andigena, unlike Phureja and Tuberosum, are both strict short-day plants (i.e., the tubers only set under short-day conditions), and if sown at the usual time in spring, the tubers will set later than those of Phureja and Tuberosum. Consequently, seeds were sown in June so that the natural differences in progression to tuber setting and maturity were compressed. Of the 90 accessions, two plants from 65 accessions and one plant from 25 accessions, giving a total of 155 plants, were grown to maturity in commercial peat-based compost in 15 cm pots. Five typical tubers were harvested from each plant in mid-November and stored at ambient temperature in a potato store at ca. 8−12 °C for three weeks.

Tubers were chopped into eighths and two opposite eighths from each of the five tubers of a single plant, representing a single replicate, were combined, FD, ground in a laboratory mill fitted with a 1 mm screen, and the resulting 155 powdered samples were stored in the dark at −20 °C until used for metabolite analysis.

Chemicals. Anhydrous pyridine, methoxylaminre hydrochloride, *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide (MSTFA), *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide (MSTFA), potassium hydrogen carbonate, sodium chloride, sulphuric acid, and standards of eicosane, hexadecane, methyl nonadecanoate, octatriacontane, ribitol, tetracosane, tetratriacontane, triacontane, tridecane, and undecane were purchased from Sigma-Aldrich Co. Ltd. (Poole, UK). Chloroform, isohexane, and methanol were of Distol grade and were supplied by Fischer Scientific UK (Loughborough, UK). Water was distilled in the laboratory.

Extraction and Derivatization of Polar and Nonpolar Metabolites from Freeze-Dried Powder. A detailed method has been previously published (*3*, *9*). Briefly, internal standards (IS) of ribitol (100 μL of 2 mg mL^{-1} in water) and methyl nonadecanoate (100 μL of 0.2 mg mL^{-1} in methanol) were added to powdered FD potato tuber (100 mg), which was shaken sequentially and additively for 30 min at 30 °C with each of methanol (3 mL), water (0.75 mL) and chloroform (6 mL). Finally, more water (1.5 mL) was added and the mixture was separated into polar and nonpolar fractions.

A portion (250 μL) of the polar fraction was oximated with methoxylamine hydrochloride in anhydrous pyridine (80 μL of 20 mg mL^{-1}) at 50 °C for 4 h and then silylated with MSTFA (80 μL) at 37 °C for 30 min. The nonpolar fraction, dissolved in chloroform (1 mL), was transesterified with 1% (v/v) methanolic sulphuric acid (2 mL) at 50 °C for 16 h and, after addition of aqueous sodium chloride (5 mL, 5% w/v), was extracted with chloroform (3 mL) which was subsequently washed with 1% (w/v) potassium hydrogen carbonate (3 mL). The dried chloroform extract was dissolved in chloroform (50 μL) and pyridine (10 μL) and silylated with MSTFA (80 μL) at 37 °C for 30 min. Subsamples (40 μL) of the derivatized polar and nonpolar fractions, separately added to an autosampler vial containing a mixture of *n*-alkanes (undecane, tridecane, hexadecane, eicosane, tetracosane, triacontane, tetratriacontane, and octatriacontane) to serve as retention index (R_t) markers, were diluted with pyridine (40 μL) and analyzed by GC-MS.

Analysis of Tuber Metabolites by Quadrupole GC-MS. The polar and nonpolar samples were analyzed similarly using a Thermo-Finnigan Trace DSQ GC-MS system. Polar and nonpolar samples were each analyzed in a random order over four consecutive days in four sequences each containing blank controls (subjected to extraction and derivatization but without sample) and two reference samples, derived from the same bulked Desiree and Phureja FD material and freshly prepared for each sequence of samples, as described earlier (*9*). Each reference sample was run nine times, in three blocks of three consecutive injections, throughout the run sequence. Samples were injected into a programmable temperature vaporizing (PTV) injector with a split of 40:1. The PTV, GC, and MS conditions were as described earlier (*9*), except that a 15 m × 0.25 mm i.d., 0.25 μm, Rtx-5MS column (Restek Corporation, Bellefonte, PA) was used. Data were acquired using the Xcalibur software package V. 1.4.

Data Analysis. Data analysis was performed as described previously (*3*, *9*). Specific ion(s) characteristic of each metabolite were selected for compound detection in processing methods created using Xcalibur. Selected ion chromatograms (SIC) were generated for each metabolite, and response ratios were automatically calculated relative to the IS, using the calculated SIC areas. Fructose and glucose methyloximes each produced two anomeric peaks, and the response ratios were summed. There were varying levels of unoximated fructose and glucose peaks that were accounted for by applying a correction factor (2.44 for fructose, 0.42 for glucose) to the response ratio for the major peak of each unoximated sugar, thus converting the values to those for oximated sugars. The values were then added to the sum of the values for the oximated peaks. The correction factors incorporated the differences between oximated and unoximated sugars in the degree of ionization and in the area of the selected ion (used to calculate the response ratios) as a proportion of the total ion count. It also took the other minor unoximated peaks into account.

Processed data were manually checked and corrected for incorrect integration or assignment of the position of the selected ion before being subject to further data analysis. Compounds were identified by analysis of standards and comparison with MS libraries. Processed data were subject to appropriate statistical treatment, including principal component analysis (PCA) of the correlation matrix, pairwise correlation analysis, and analysis of variance (ANOVA). PCA was applied to response ratio values of the combined polar and nonpolar metabolites data set of all genotypes. ANOVA was performed between groups for each metabolite. Metabolites were selected as showing significant differential expression among groups at the *p*-value corresponding to a 2% false discovery rate. Differentially expressed metabolites were grouped by hierarchical clustering according to similar pairwise standard error of difference (SED) patterns. Differences between means of two SEDs are significant at approximately 5%.

Table 1. List of Metabolites Used in the Analysis of Potato Tubers

compound class	metabolite
amino acid	alanine, β-alanine, γ-aminobutyric acid, asparagine, aspartic acid, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tyrosine, valine
organic acid	caffeic acid, citric acid, 2,3-dihydroxypropanoic acid, fumaric acid, 2-keto-1-gluconic acid,[a] malic acid, 2-piperidinecarboxylic acid (pipecolic acid), quinic acid, 2,3,4-trihydroxybutyric (threonic acid)
sugar[b]	carbohydrate1717, carbohydrate2363, fructose, glucaric or galactaric acid,[a] glucose, polysaccharide3062, polysaccharide3078, sucrose
sugar alcohol	inositol
other polar	*N*-ethyldiethanolamine,[a] phosphate.
unknown polar[b]	P1108, P1143, P1344, P1400, P1471, P1525, P1559, P1630, P1680, P1729, P1786
fatty acid[c]	*n*-14:0, *n*-15:0, *n*-16:0, *n*-17:0, *n*-18:0, *n*-20:0, *n*-21:0, *n*-22:0, *n*-23:0, *n*-24:0, *n*-25:0, *n*-26:0, *n*-28:0, *n*-29:0, n-30:0, *iso*-15:0, *iso*-17, 15:1, 16:1, 18:2(ω6), 18:3(ω3), 2-hydroxy-16:0, 2-hydroxy-24:0.
fatty alcohol[c]	*n*-21:0, *n*-22:0, *n*-23:0, *n*-24:0, *n*-26:0, *n*-27:0, *n*-28:0, *n*-29:0, *n*-30:0.
sterol	Δ5-avenasterol β-sitosterol, stigmasterol
other nonpolar	*n*-23:0 alkane, *cis*-ferulic acid, *trans*-ferulic acid, solanid-5-en-ol
unknown nonpolar[b]	N1785, N1868, N1914, N2268

[a] Metabolites tentatively identified by comparison of the mass spectra to library databases. All other named metabolites were identified by comparison of the mass spectra and retention times to those of standards. [b] Unidentified sugars and unknown polar (P) and non-polar (N) metabolites are designated with relative retention index (RRi) relative to a retention standard of straight-chain alkanes. [c] Fatty acids and fatty alcohols are designated either normal (*n*), or iso-branched (*iso*). The total number of carbons and number of double bonds is given, and the positions of double bonds in polyunsaturated fatty acids uses the ω annotation, e.g., 18:3(ω3) is an 18 carbon fatty acid with three double bonds in the 9, 12, and 15 positions from the carboxylic acid end of the molecule.

Pairwise correlations were performed on response ratios, and only those metabolites that gave correlation coefficients of 0.6 or greater with at least one other metabolite were included in plots. All statistical analyses were carried out using Genstat for Windows, 10th edition.

RESULTS AND DISCUSSION

Variability in the Levels of Metabolites among Genotypes. By considering the potato reference materials analyzed with each set of samples, there was some separation according to run sequence when the response ratio data for all metabolites were analyzed by PCA. Nineteen of the 68 polar metabolites and four out of 47 nonpolar metabolites were highly variable (% CV greater than 40%), and separation according to sequence was eliminated when these metabolites were removed from the data set. The majority of these metabolites were unidentified but included ascorbic acid, glutamine, α-glycerophosphate, homoserine, oxalic acid, oxoproline (formed from glutamic acid during the methyloximation step), Δ5,24(25)-stigmastadienol, and tryptophan. Subsequent analysis of the data excluded these metabolites. A list of all metabolites included in the study is given in **Table 1**.

There was considerable variation in the levels of the majority of metabolites among individual samples, providing substantial scope for selecting germplasm with enhanced or decreased levels of selected metabolites that may be incorporated into breeding programmes. The asparagine level is considered to be the limiting factor in the formation of potentially harmful acrylamide formed between free asparagine and reducing sugars in foods at high temperatures (*13*). The relative asparagine levels varied over a 100-fold among the genotypes. There was considerable variation in asparagine level within each *S. tuberosum* group; Tuberosum had the least and Phureja the most (**Figure 1**, cluster 1A). It should be noted that, as well as genotype, other factors including location and year of growth are known to affect amino acid levels in potatoes (*14*, *15*). It would appear that some amino acids are more variable than others. Proline appears to be consistently the most variable with location (*14*, *15*), and others (alanine, γ-aminobutyric acid, glutamine, histidine, tyrosine, and valine) have also been shown to vary appreciably (*14*). In one study (*14*), there was year-to-year variation in the quantities of all amino acids but the general amino acid composition was consistent, whereas in another study (*15*), aspartic and glutamic acids were the most variable.

High levels of reducing sugars in tubers is considered a negative quality because upon frying, as in the production of French fries, browning of the product occurs due to Maillard reactions between reducing sugars and amino acids (*16*). Reducing sugar levels are cultivar-dependent but generally decrease with increasing tuber maturity and increase during cold storage (*17*). The relative levels of fructose and glucose varied about 200- and 300-fold, respectively, in the genotypes studied and were significantly greatest in Tuberosum than the other groups (**Figure 1**, cluster 7).

Differences in Metabolite Levels among the Four *S. tuberosum* Groups. To determine whether there were any differences in metabolite levels between the four groups, the response ratios of the combined polar and nonpolar metabolites were analyzed by PCA. Principal component 1 (PC1) broadly represented total metabolite content (all loadings were approximately the same), and PC2 essentially contrasted total polar with total nonpolar metabolites. PC3 was important in separating the groups. In a plot of PC1 vs PC3, although there was some overlap among the groups, the majority of Phureja and Stenotomum genotypes formed a loose grouping that was separated from the majority of Andigena and Tuberosum genotypes that formed another grouping (**Figure 2**). An examination of the loadings for PC3 indicated especially that many fatty acids and some fatty alcohols may be important in the separations.

An ANOVA of the response ratios of all metabolites for the four *S. tuberosum* groups revealed that there were significant differences ($P < 0.05$) between 24 of the 43 nonpolar metabolites, including many fatty acids, and 17 of the 49 polar metabolites. Hierarchical clustering according to similar pairwise standard error of differences patterns revealed eight patterns using a cutoff of 93% similarity (**Figure 1**). In four clusters (clusters 1, 2, 4, and 5), Phureja and usually Stenotomum had greater levels of several metabolites, including many fatty acids and some fatty alcohols and amino acids, than Tuberosum. Similar to Tuberosum, Andigena tended to have the lowest levels of metabolites in clusters 2 and 5 but had intermediate levels in clusters 1A and B and 4. The decreased levels of some fatty acids in Tuberosum compared to Phureja had earlier been observed (*18*). In cluster 3, Phureja tended to have larger levels than all the other groups.

In cluster 7, Tuberosum had greater levels of some sugars than all other groups and the two metabolites in cluster 6 were increased compared to Andigena and Stenotomum, but not Phureja. In cluster 8, the levels of *n*-29:0 fatty acid and fatty

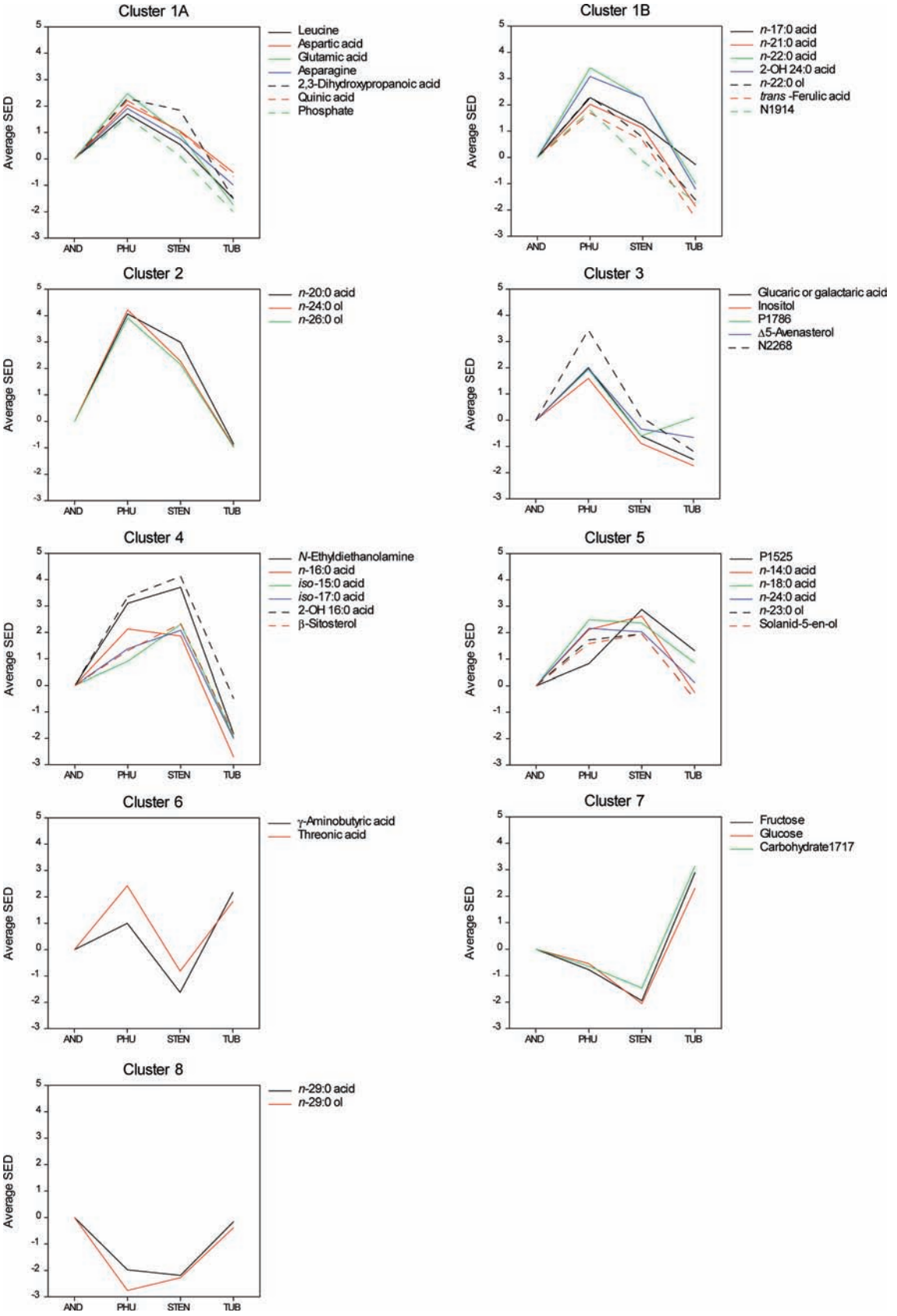

Figure 1. Pairwise standard error of difference (SED) plots from ANOVA of the response ratios of four *S. tuberosum* groups. Metabolites were grouped by similar patterns produced by hierarchical clustering with a cutoff of 93% similarity. Cluster 1 was split into polar (cluster 1A) and nonpolar (cluster 1B) metabolites, but these do not refer to subclusters produced by hierarchical clustering. AND, Andigena; PHU, Phureja; STEN, Stenotomum; TUB, Tuberosum.

alcohol were greater in Andigena and Tuberosum compared to Phureja and Stenotomum.

It is possible that some of the variation seen is not due to the different genetic backgrounds of the different groups but is more

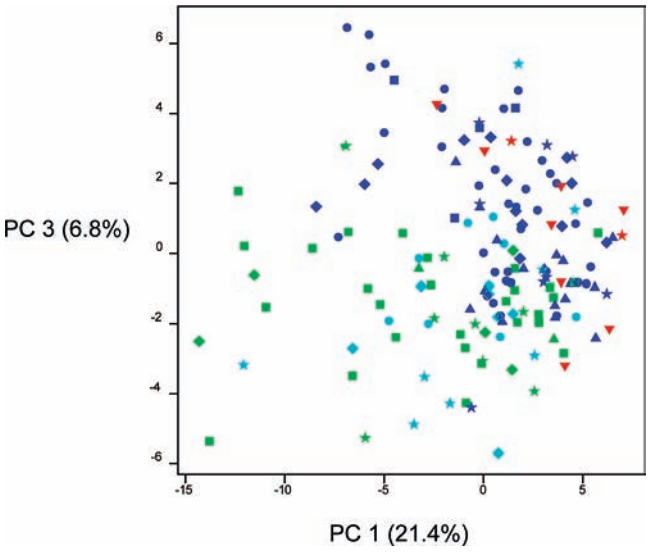

Figure 2. Principal component score plots of PC1 vs PC3 of response ratios of combined polar and nonpolar metabolites of all genotypes from the four *S. tuberosum* groups. Color indicates group: Andigena (dark blue), Phureja (green), Stenotomum (light blue), and Tuberosum (red). Symbol indicates geographical origin Bolivia (◆), Chile (▼), Colombia (■), Ecuador (▲), Peru (●), other or unknown origin (★).

influenced by their physiological status at sampling time. For example, reducing sugar (fructose and glucose) and sucrose levels generally decrease with increasing tuber maturity and the reducing sugars often increase at the end of the season (*17*, *19*). Phureja and Tuberosum would naturally be the earliest to tuberize (*20*), although measures were taken to minimize this effect (see Materials and Methods). It is possible that, at harvest, the tubers of these two groups were, to some degree, more mature than those from Andigena and Stenotomum. However, Phureja and Tuberosum did not align closely on the basis of metabolic profile (**Figures 1** and **2**), and the fructose and glucose levels were substantially different (**Figure 1**, cluster 7). Phureja and Stenotomum represent opposite extremes in terms of the dormancy of the harvested tubers, Phureja being the least physiologically dormant group (*20*), yet this difference is not evident in the metabolic profile data (**Figures 1** and **2**). It follows that the indications are that variations in metabolic profile are due to the genetic origins of the groups rather than their physiological state.

There is considerable evidence to suggest that in terms of the origins of the four groups, Stenotomum is believed to be the most primitive diploid species and has been proposed to be the ancestor of the diploid Phureja (*20*). Stenotomum, presumably through chromosome doubling, is also likely to be directly ancestral to tetraploid Andigena, a link recently supported by chloroplast and nuclear DNA markers (*21*). Chilean tetraploid Tuberosum are thought to have evolved under domestication from group Andigena, with some introgression from Southern landraces and wild species, and share cytoplasm, morphology, and phenology with European potatoes (*20*).

The similarity of Stenotomum and Phureja and their distinction from Tuberosum, is supported by the metabolite data. The data also supports the link between Chilean Tuberosum genotypes and Andigena.

Variation in Metabolite Profiles According to Geographical Origins of Genotypes. The various genotypes used in this study had diverse geographical origins within South American countries. The Phureja and Tuberosum genotypes originated mainly from Colombia and Chile, respectively, whereas Stenotomum genotypes originated mainly from Bolivia and Peru. A large proportion of Andigena genotypes, the largest group, were derived from Peru, and there were also substantial numbers from Bolivia and Ecuador. With the exception of Phureja, which is widespread but was not represented by genotypes from Ecuador and Peru, the genotypes within each group were representative of their geographical distribution in South America (*20*).

According to PCA, geographical origin, irrespective of group, was not a major influence on the separations (**Figure 2**). For example, Bolivian, and likewise Peruvian, genotypes belonging to more than one Tuberosum group do not form a separate cluster but are widely distributed in the PCA plot. However, within Andigena, but not the other groups, there was some separation of the genotypes according to geographical origin. The Bolivian and Ecuadorean genotypes especially formed two separate groupings. The Bolivian and many Peruvian genotypes had a positive score in PC3 and formed a loose grouping that tended to group with the majority of Tuberosum genotypes, whereas most Ecuadorean and some Peruvian genotypes had a negative score in PC3 and formed another cluster that grouped with some Phureja, Stenotomum, and Tuberosum genotypes.

An ANOVA of the response ratios of all metabolites for the four groups, but with Andigena split into three further groups according to geographical origin, revealed significant differences in 14 polar and 17 nonpolar metabolites involving at least one of the geographical subgroups. There were six clusters, using a cutoff of 88% similarity, by hierarchical clustering according to similar pairwise standard error of differences patterns (**Figure 3**). However, none of these metabolites were useful as markers for differentiating the three Andigena groups, i.e., the range of metabolite levels for the genotypes within any one group was not exclusive to that group.

In cluster 1, the Bolivian Andigena genotypes contained significantly greater levels of some amino acids and sugars than the other two Andigena groups (the difference in fructose levels between Bolivian and Peruvian Andigena was not significant), and the levels of these sugars and γ-aminobutyric acid were also greater in Tuberosum. The metabolites in cluster 2 were increased in Bolivian and Peruvian Andigena genotypes compared to Ecuadorean Andigena (the difference in *n*-29:0 fatty alcohol levels between Bolivian and Ecuadorian Andigena was not significant), and the fatty acid and alcohol were also increased in Tuberosum but decreased in Phureja and Stenotomum. The levels of metabolites in cluster 5 were greater in Bolivian Andigena than in Ecuadorean Andigena, and there were intermediate, but not significantly different, levels in Peruvian Andigena. These metabolites were also increased in Phureja and Stenotomum.

The levels of two fatty acids and two fatty alcohols in cluster 6A were greater (but not always significant) in Peruvian Andigena, together with Phureja and Stenotomum, than in Ecuadorean Andigena and usually Tuberosum. In cluster 6B, other fatty acids were increased in Phureja and Stenotomum compared to Andigena and Tuberosum, but there were no significant differences between the three Andigena groups.

In cluster 3, the metabolites tended to be at greater levels in Ecuadorean Andigena, together with Phureja and sometimes Stenotomum, than the other Andigena groups and Tuberosum. However, only for phosphate were the levels significantly increased.

The metabolite data therefore not only supports a link between Chilean Tuberosum and Andigena, but additionally there is an indication that the Tuberosum are closer to Bolivian and some Peruvian Andigena than to other Peruvian and Ecuadorean Andigena. Thus the supposed route of Tuberosum under domestication along the spine of the Andes through Bolivia to Chile is

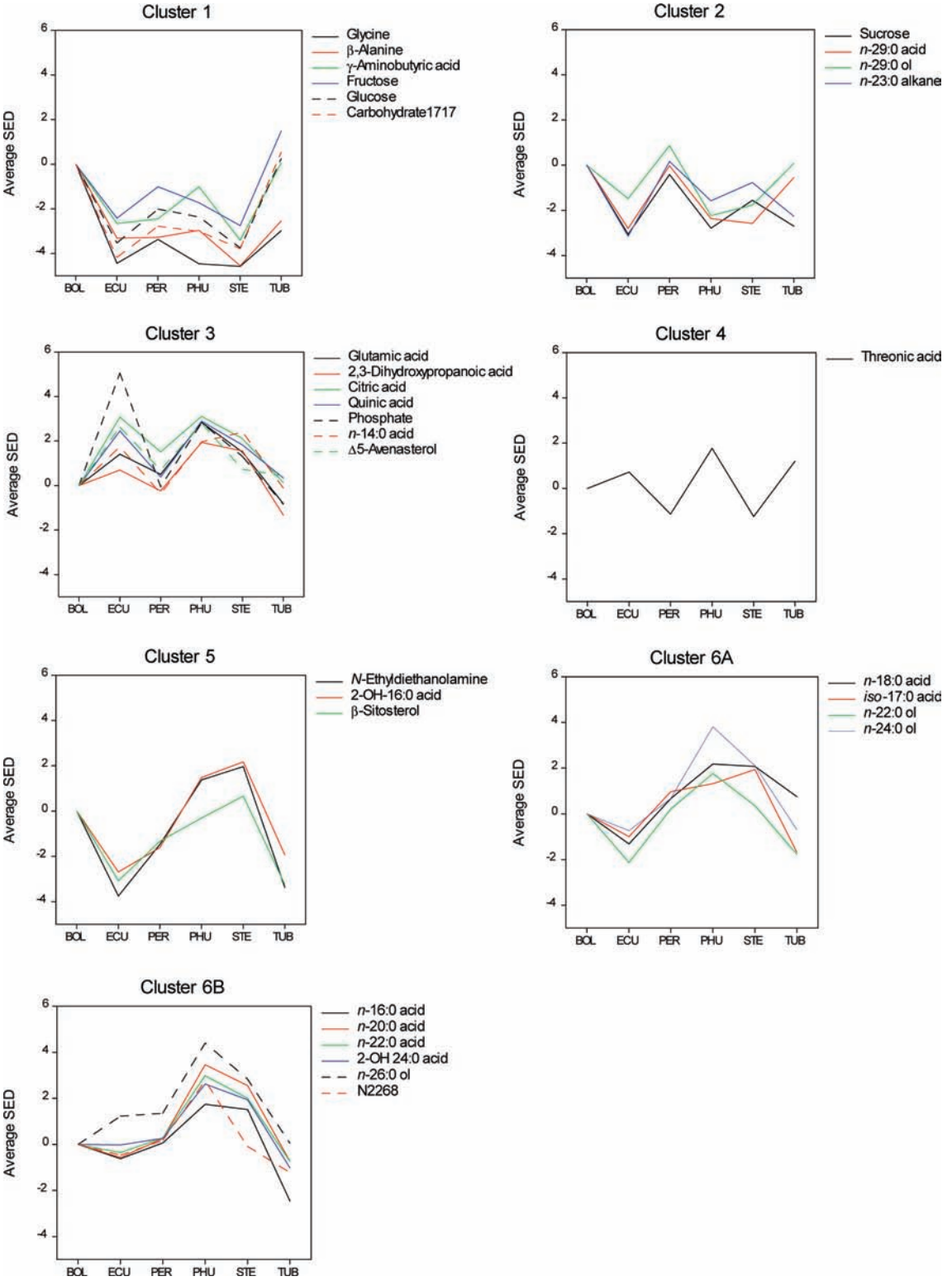

Figure 3. Pairwise standard error of difference (SED) plots from ANOVA of the response ratios of four *S. tuberosum* groups with Andigena in three subgroups according to geographical origin. Metabolites were grouped by similar patterns produced by hierarchical clustering with a cutoff of 88% similarity. Only metabolites in which there were significant differences ($P < 0.05$) involving at least one of the Andigena subgroups were included. Cluster 6 was split into clusters 6A and 6B, but these do not refer to subclusters produced by hierarchical clustering. Key to abbreviations as in **Figure 1**, except STE, Stenotomum. BOL, Bolivia; ECU, Ecuador; PER, Peru refer to geographical origins of Andigena genotypes.

compatible with the patterns of metabolomic diversity uncovered in this study.

Correlation of Metabolite Levels. The degree of correlation of the levels of metabolites between similar sample types can be used to aid the understanding of biosynthetic links between metabolites (*22*). Using the whole data set from the response ratios of 92 polar and nonpolar metabolites, and including all samples, a total of 4186 pairwise correlation coefficient values were obtained. Metabolites were considered to be highly correlated if the coefficient had a value of 0.7 or greater (*3*, *6*), and on this basis

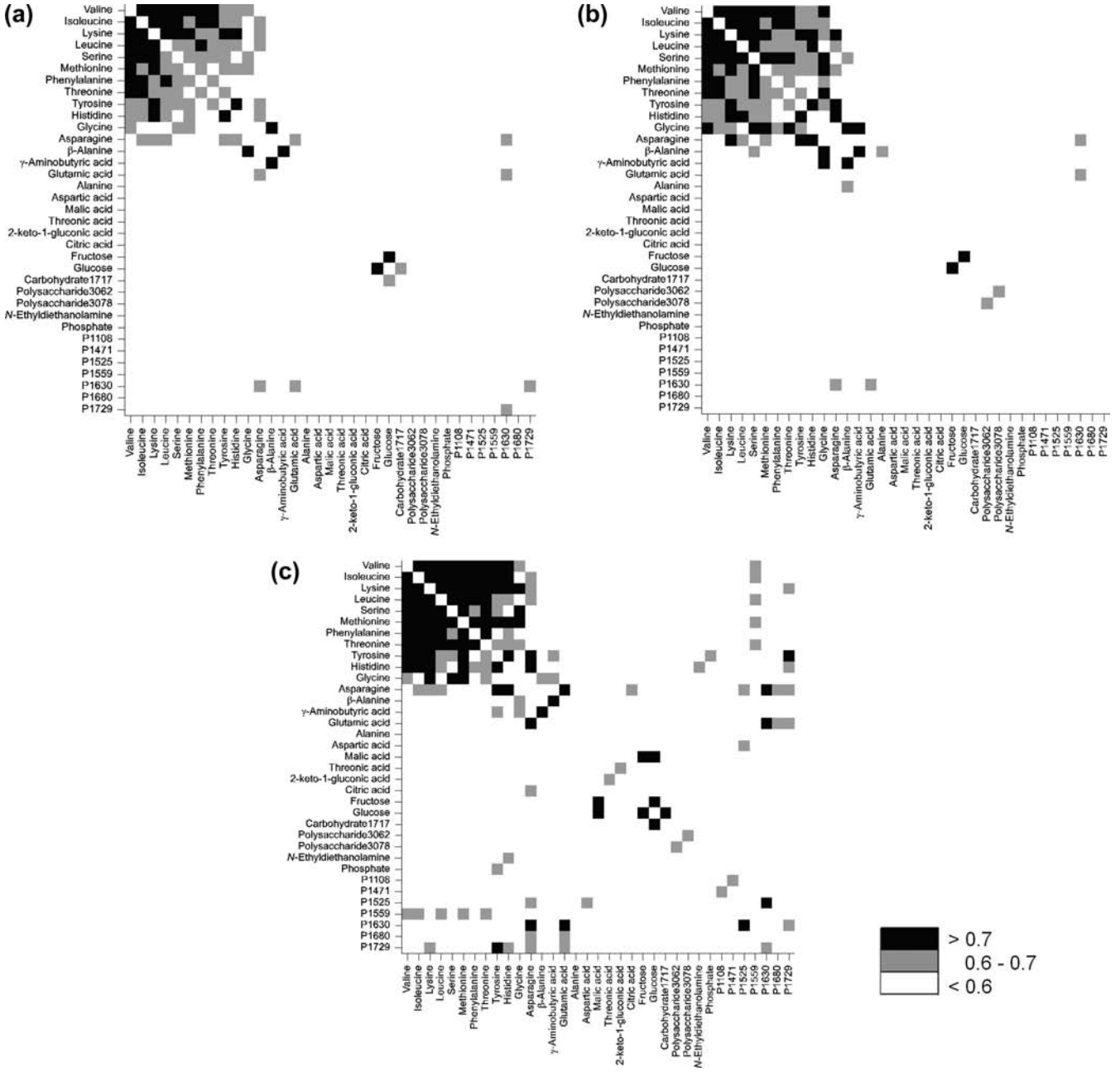

Figure 4. Correlation matrices of polar metabolites from (**a**) all four *S. tuberosum* groups, (**b**) Andigena only, and (c) Phureja only. Only those metabolites that had correlation coefficients of 0.6 or greater with at least one other metabolite in at least one of the three data sets were included.

there were 60 strongly positively correlated pairs of metabolites, 22 between polar metabolites (**Figure 4a**) and 38 between nonpolar metabolites (**Figure 5a**). There were no correlations between polar and nonpolar metabolites, and there were no strong negative correlations. Considering the biochemical remoteness of the different classes of polar metabolites to the nonpolar metabolite classes included in this study, it is not surprising that there was not a correlation between polar and nonpolar metabolites.

Within the polar metabolites, 21 correlations involved 13 amino acids, and the remaining correlation was between fructose and glucose, not surprisingly considering both sugars are derived from sucrose. There were a further 30 correlations between amino acids that had values between 0.6 and 0.7. A high degree of correlation between amino acids was earlier observed for a set of potato cultivars and landraces belonging almost exclusively to group Tuberosum (*3*). In the present study, the correlations, although strong, were not as great as in the earlier study and may reflect the greater genetic diversity of genotypes in the present study, i.e., belonging to four *S. tuberosum* groups. For example, some expected strong correlations (e.g., between phenylalanine and tyrosine, and β-alanine and alanine), observed in the earlier study (*3*), were not present in this study, although the values were between 0.5 and 0.6. However, some unexpected correlations (e.g., between serine and lysine, phenylalanine and isoleucine, β-alanine and γ-aminobutyric acid), also observed in the earlier study (*3*) and in other studies on potato lines (*6*, *22*), were apparent. Also similar to the earlier study (*3*), β-alanine, γ-aminobutyric acid, and to a greater extent, aspartic acid, glutamic acid, and proline, were only weakly correlated to most other amino acids. The most notable difference compared to the earlier study (*3*) was the absence of strong correlations of alanine and asparagine to other amino acids, although asparagine had correlations of between 0.6 and 0.7 with several amino acids.

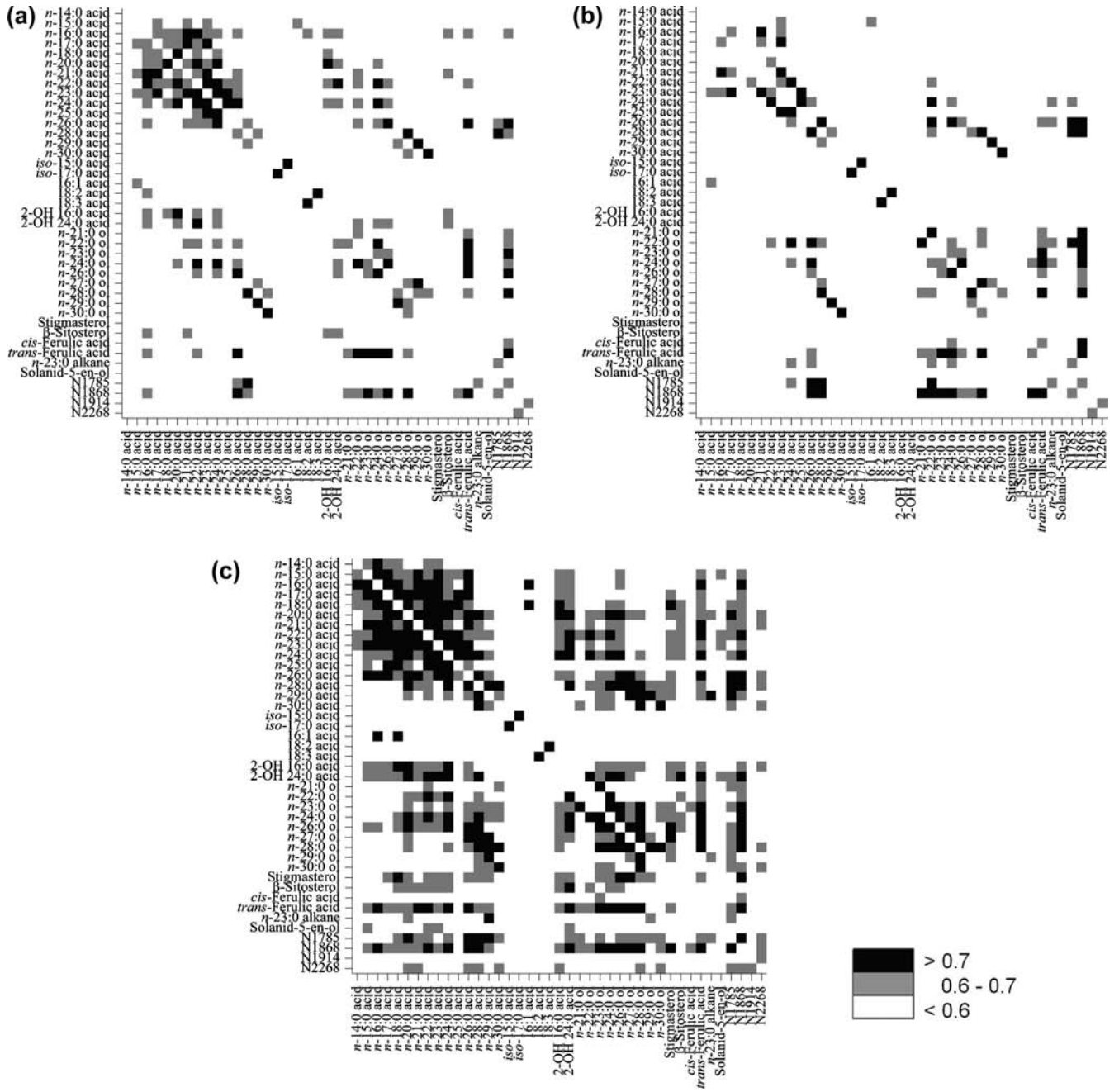

Figure 5. Correlation matrices of nonpolar metabolites from (**a**) all four *S. tuberosum* groups, (**b**) Andigena only, and (**c**) Phureja only. Only those metabolites that had correlation coefficients of 0.6 or greater with at least one other metabolite in at least one of the three data sets were included.

As in the previous study (*3*), the strong correlation among amino acids supports, to a large extent, the mechanism of general amino acid control in plants (*23*). The high correlation of "minor amino acids" (defined as branched amino acids isoleucine, leucine and valine, and aromatic amino acids phenylalanine and tyrosine) observed by Noctor et al. (*24*) was also largely supported.

When correlation analysis was performed using only the Andigena genotype data (**Figure 4b**), the results were generally similar to that obtained using the whole data set, the major difference being the greater correlation of glycine with other amino acids. With Phureja genotypes only (**Figure 4c**), again the results were similar, although correlations between amino acids tended to be stronger. In addition, malic acid was strongly correlated to fructose and glucose, and there were some increased correlations mainly involving unidentified metabolites. Correlations were not performed with Stenotomum only and Tuberosum only data sets because of the low numbers of genotypes represented.

The small increase in degree of correlations observed when single groups of *S. tuberosum* were analyzed, compared to all four groups, may be explained by a narrowing of the genetic diversity of the data set and presumably a corresponding increase in the biochemical similarity, as previously observed for a study of Tuberosum cultivars and landraces (*3*). In the earlier study (*3*), increased correlations between some amino acids were observed when the landraces were removed from the whole data set. In that case, proline and aspartic acid were mainly affected but in the present study, glycine, for example, was affected.

Within the nonpolar metabolites, with a few exceptions, there were strong correlations between saturated fatty acids ranging from C_{16} to C_{26} (**Figure 5a**). As noted in the earlier study (*3*), there were correlations not only between fatty acids of even numbers of

carbons and between those of odd numbers but also between those of each type despite each being derived from different starting units. In contrast to the earlier study, there were correlations between fatty acids of C_{18} or less with those of C_{20} to C_{24}. Perhaps this is not surprising because, although the sites of biosynthesis and enzyme systems differ for the 2 groups of fatty acids, C_{18} is the precursor of longer even chain length fatty acids (*25*). The strong correlation of 18:2(*n*-6) and 18:3(*n*-3), also observed in the earlier study (*3*), reflects their adjacent biosynthetic positions. The two hydroxy fatty acids (2-OH 16:0 and 2-OH 24:0) were correlated to their respective nonhydroxy saturated precursors (*n*-16:0 and *n*-18:0) and to other even carbon fatty acids.

Strong correlations of even carbon fatty alcohols with the fatty acid of the same chain length, and often with fatty acids of other even carbon chain lengths, can be rationalized because fatty alcohols are formed via the aldehyde from the corresponding acyl-CoA of the same chain length (*25*). There were also some strong correlations between different fatty alcohols, but these were not as extensive as for the fatty acids, as they tended to be restricted over narrow chain length distributions. Unexpectedly, there were strong correlations between fatty alcohols and *trans*-ferulic acid.

In contrast to the polar metabolites, when correlations were performed with Andigena genotypes only (**Figure 5b**) and with Phureja genotypes only (**Figure 5c**), the patterns of correlation between metabolites were considerably different from that obtained for all genotypes and were remarkably different between the two groups. For Andigena, correlations were considerably weaker, and for Phureja they were considerably stronger, than for the whole data set. The correlations for Andigena were weaker among saturated straight-chain fatty acids and between these acids and hydroxy fatty acids. For Phureja, the converse was apparent; there were virtually all possible correlations among saturated straight-chain fatty acids from C_{15} to C_{26} and between these acids and the two hydroxy acids. Likewise, the extent of correlation between fatty acids and alcohols, and among fatty alcohols, was greater. Additionally, there were reasonably strong correlations between sterols (stigmasterol and β-sitosterol) and saturated straight-chain and hydroxy fatty acids as earlier observed, to a lesser extent, for Tuberosum cultivars (*3*).

The dissimilarity in the correlation patterns of nonpolar metabolites between Andigena and Phureja may reflect differences in the biochemistry of the two groups, supported by the greater levels of many nonpolar metabolites in Phureja, whereas the biochemistry of the polar metabolites may be relatively consistent between the groups. These biochemical differences may have a functional basis, as Phureja is adapted for the lower-altitude moister habitats and the foliage of Phureja is distinguishable from other taxa in that it is generally more shiny (*20*). Alternatively, the relative lack of correlation in Andigena might reflect the greater genetic diversity within this group (Andigena genotypes were geographically more widespread than Phureja which was represented mainly by Colombian genotypes) and may be a pitfall in correlating data from different genotypes. Indeed, other studies on correlating metabolites from potatoes have used either only one genotype with several replicates grown under uniform controlled conditions but resulting in random fluctuations in metabolite profile (*22*) or highly related genetically modified lines (*6*). Using the former approach with a single representative genotype from each group may be one approach to address differences in biochemistry between the four groups. This approach may be of particular benefit to our efforts in trying to understand why Phureja is different and organoleptically generally preferable to Tuberosum (*1*).

In summary, GC-MS has been shown to be a useful tool for exploring phytochemical diversity in tubers of *Solanum tuberosum* groups Andigena, Phureja, Stenotomum, and Tuberosum. Among individual genotypes, there was a large range in the levels of polar metabolites, including asparagine, fructose, and glucose that are important to quality, and nonpolar metabolites, offering substantial scope for selecting germplasm that may be incorporated into breeding programmes. However, as has been demonstrated for amino acids (*14*, *15*), metabolite levels would be expected to vary from year to year and among different locations as well as different genotypes. An important factor for selecting genotypes for breeding programmes is whether the ranking of genotypes with respect to the levels of a particular metabolite is consistent from year to year and among different locations. Published data of this type is sparse for potatoes, but it would appear that for amino acids there may be some year to year differences in ranking (*14*), which may impact on the usefulness of the data for breeding regimes.

Of course, there is a limit to the type of metabolite that can be analyzed by GC-MS, being restricted to lower molecular weight and mainly primary metabolites. To expand the range to include those of higher molecular weight secondary metabolites, it would be useful to apply LC-ESI-MS technologies (*10*, *11*).

Using PCA, it was shown that, on the basis of the metabolic profile, Phureja and Stenotomum were similar to one another and different from Tuberosum and the majority of Andigena genotypes. There were significant differences in the levels of many metabolites between the four groups. The separation of Tuberosum should be viewed with some caution as only few Tuberosum genotypes were included in this study, although their Chilean origin was representative of their distribution in South America. It is interesting that the relatedness of the groups on the basis of metabolic profile loosely agrees with the phylogeny of the groups in that Stenotomum is believed to be the ancestor of Phureja and they are both distinct from Tuberosum (*20*). Andigena genotypes could be partially separated by metabolite profile according to geographical origin. In particular, those from Bolivia were distinct from those from Ecuador.

The plants in this study were from diverse South American origins, but they were grown in a glasshouse in the United Kingdom. This might raise questions about the relevance of the data and whether the tubers reached their full metabolic potential. Glasshouse-raised potatoes will perform differently from field-grown material. However, in the case of diverse germplasm from a number of different ecogeographical zones and adapted to different daylength regimes, ex situ culture in a uniform and suitable environment provides the best way to make meaningful comparisons of the biosynthetic potential of the material. By growing plants later in the season, using natural daylight to compress differences in tuberization response due to daylength, all plants developed tubers as normal and they all had sufficient time to fill and mature the tubers initiated, resulting in tubers at similar stages of maturity at harvesting. Although the details of the metabolic profiles could vary if the plants were grown in their respective original environments, the objective of this work was to study the genetic potential of the material. The possibility that differences in the physiological status of the different groups at sampling might have influenced the metabolic profile has been discussed, but the indications are that variations in metabolic profiles are largely due to genetic variation.

There was a strong correlation between many amino acids as observed in our earlier study using a range of mainly group Tuberosum cultivars (*3*). It is interesting that some unexpected correlations between amino acids were apparent in both studies. Also common to both studies, some amino acids, such as

β-alanine, were not strongly correlated to other amino acids. The correlation of nonpolar metabolites also showed many similarities, and some differences, to our earlier study (*3*). The correlations between polar metabolites was essentially similar when the analysis was performed using only Phureja genotypes and only Andigena genotypes, but this was not the case for nonpolar metabolites; compared to the whole data set, Andigena and Phureja gave substantially weaker and stronger correlations, respectively. This may suggest differences in the biochemistry of the two groups, or alternatively may indicate a pitfall in using different genotypes for such studies, and therefore the interpretation must be viewed with caution.

ABBREVIATIONS USED

ANOVA, analysis of variance; CPC, Commonwealth Potato Collection; FD, freeze-dried; GC-MS, gas chromatography−mass spectrometry; IS, internal standard; MSTFA, *N*-methyl-*N*-(trimethylsilyl)-trifluoroacetamide; LC-ESI-MS, liquid chromatography−electrospray ionization−mass spectrometry; PCA, principal component analysis; PTV, programmable temperature vaporizing; RRi, relative retention index; R_T, retention index; SIC, selected ion chromatogram.

Supporting Information Available: Response ratios of polar and nonpolar metabolites in all genotypes from the four *S. tuberosum* groups. Loadings plots for principal component analysis of response ratios of combined polar and nonpolar metabolites of all genotypes from the 4 *S. tuberosum* groups (XLS). This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) De Maine, M. J.; Lees, A. K.; Muir, D. D.; Bradshaw J. E.; Mackay, G. R. Long-day adapted phureja as a resource for potato breeding and genetic research. In *Potato, Global Research and Development*; Khurana, S. M. P., Shekhawat, G. S., Singh, B. P., Pandey, S. K., Eds.; Indian Potato Association: Shimla, India, 2000; Vol. 1, pp 134−137.

(2) Dobson, G.; Shepherd, T.; Marshall, R.; Verrall, S. R.; Conner, S.; Griffiths, D. W.; McNicol, J. W.; Stewart, D.; Davies, H. V. Application of metabolite and flavour profiling to studies of biodiversity in *Solanum* species. In *Concepts in Plant Metabolomics*; Nikolau, B. J., Wurtele, E. S., Eds.; Springer: Dordrecht, The Netherlands, 2007; pp 259−268.

(3) Dobson, G.; Shepherd, T.; Verrall, S. R.; Conner, S.; McNicol, J. W.; Ramsay, G.; Shepherd, L. V. T.; Davies, H. V.; Stewart, D. Phytochemical diversity in tubers of potato cultivars and landraces using a GC-MS metabolomics approach. *J. Agric. Food Chem.* **2008**, *56*, 10280–10291.

(4) 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.

(5) Davies, H. V.; Shepherd, L. V. T.; Burrell, M. M.; Carrari, F.; Urbanczyk-Wochniak, E.; Leisse, A.; Hancock, R. D.; Taylor, M.; Viola, R.; Ross, H.; McRae, D.; Willmitzer, L; Fernie, A. R. Modulation of fructokinase activity of potato (*Solanum tuberosum*) results in substantial shifts in tuber metabolism. *Plant Cell Physiol.* **2005**, *46*, 1103–1115.

(6) Roessner, U.; Luedemann, A.; Brust, D.; Fiehn, O.; Linke, T.; Willmitzer, L.; Fernie, A. R. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **2001**, *13*, 11–29.

(7) Szopa, J.; Wrobel, M.; Matysiak-Kata, I.; Swiedrych, A. The metabolic profile of the 14-3-3 repressed transgenic potato tubers. *Plant Sci.* **2001**, *161*, 1075–1082.

(8) Catchpole, G. S.; Beckmann, M.; Enot, D. P.; Mondhe, M.; Zywicki, B.; Taylor, J.; Hardy, N.; Smith, A.; King, R. D.; Kell, D. B.; Fiehn, O; Draper, J. Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. *Proc. Nat. Acad. Sci. U.S.A.* **2005**, *102*, 14458–14462.

(9) Shepherd, T.; Dobson, G.; Verrall, S. R.; Conner, S.; Griffiths, D. W.; McNicol, J. W.; Davies, H. V.; Stewart, D. Potato metabolomics by GC-MS: what are the limiting factors? *Metabolomics* **2007**, *3*, 475–488.

(10) von Roepenack-Lahaye, E.; Degenkolb, T.; Zerjeski, M.; Franz, M.; Roth, U.; Wessjohann, L.; Schmidt, J.; Scheel, D.; Clemens, S. Profiling of *Arabidopsis* secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *Plant Physiol.* **2004**, *134*, 548–559.

(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) Huaman, Z.; Spooner, D. M. Reclassification of landrace populations of cultivated potatoes (*Solanum* Sect. *Petota*). *Am. J. Bot.* **2002**, *89*, 947–965.

(13) Matsuura-Endo, C.; Ohara-Takada, A.; Chuda, Y.; Ono, H.; Yada, H.; Yoshida, M.; Kobayashi, A.; Tsuda, S.; Takigawa, S.; Noda, T.; Yamauchi, H.; Mori, M. Effects of storage temperature on the contents of sugars and free amino acids in tubers from different potato cultivars and acrylamide in chips. *Biosci. Biotechnol. Biochem.* **2006**, *70*, 1173–1180.

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

(15) Talley, E. A.; Fitzpatrick, T. J.; Porter, W. L. Chemical composition of potatoes. VIII. Effect of variety, location and year of growth on the content of nitrogen compounds. *Amer. Potato J.* **1970**, *47*, 231–244.

(16) Shallenberger, R. S.; Smith, O.; Treadway, R. H. Role of sugars in the browning reaction in potato chips. *J. Agric. Food Chem.* **1959**, *7*, 274–277.

(17) Richardson, D. L.; Davies, H. V.; Ross, H. A. Potato tuber sugar content during development and storage (10 °C): possible predictors of storage potential and the role of sucrose in storage hexose accumulation. *Potato Res.* **1990**, *33*, 241–245.

(18) Dobson, G.; Griffiths, D. W.; Davies, H. V.; McNicol, J. W. Comparison of fatty acid and polar lipid contents of tubers from two potato species, *Solanum tuberosum* and *Solanum phureja*. *J. Agric. Food Chem.* **2004**, *52*, 6306–6314.

(19) Kolbe, H.; Stephan-Beckmann, S. Development, growth and chemical composition of the potato crop (*Solanum tuberosum* L.). II. Tuber and whole plant. *Potato Res.* **1997**, *40*, 135–153.

(20) Hawkes, J. G., *The Potato. Evolution, Biodiversity and Genetic Resources*; Belhaven Press: London, U.K.,1990.

(21) Sukhotu, T.; Hosaka, K. Origin and evolution of Andigena potatoes revealed by chloroplast and nuclear DNA markers. *Genome* **2006**, *49*, 636–647.

(22) Steuer, R.; Kurths, J.; Fiehn, O.; Weckwerth, W. Interpreting correlations in metabolomic networks. *Biochem. Soc. Trans.* **2003**, *31*, 1476–1478.

(23) Halford, N. G.; Hey, S.; Jhurreea, D.; Laurie, S.; McKibbin, R. S.; Zhang, Y.; Paul, M. J. Highly conserved protein kinases involved in the regulation of carbon and amino acid metabolism. *J. Exp. Bot.* **2004**, *55*, 35–42.

(24) Noctor, G.; Novitskaya, L.; Lea, P. J.; Foyer, C. H. Co-ordination of leaf minor amino acid contents in crop species: significance and interpretation. *J. Exp. Bot.* **2002**, *53*, 939–945.

(25) Shepherd T. Wax pathways. In *Encyclopedia of Applied Plant Sciences*; Thomas, B., Murphy, D. J., Murray, B. G., Eds.; Elsevier: Oxford, U.K., 2003; pp 1204−1225.

Received for review September 3, 2009. Revised manuscript received November 24, 2009. Accepted November 25, 2009. We acknowledge financial support of the Rural and Environment Research and Analysis Directorate of the Scottish Government.