

## GENOMICS ARTICLE

# Metabolic Profiling Allows Comprehensive Phenotyping of Genetically or Environmentally Modified Plant Systems

Ute Roessner,<sup>a</sup> Alexander Luedemann,<sup>a</sup> Doreen Brust,<sup>b</sup> Oliver Fiehn,<sup>a</sup> Thomas Linke,<sup>b</sup> Lothar Willmitzer,<sup>a</sup> and Alisdair R. Fernie<sup>a,1</sup>

<sup>a</sup>Max-Planck-Institut für Molekulare Pflanzenphysiologie, Am Mühlenberg 1, 14476 Golm, Germany

<sup>b</sup>Institut für Informatik, Universität Potsdam, Am Neuen Palais 10, 14469 Potsdam, Germany

Metabolic profiling using gas chromatography–mass spectrometry technologies is a technique whose potential in the field of functional genomics is largely untapped. To demonstrate the general usefulness of this technique, we applied to diverse plant genotypes a recently developed profiling protocol that allows detection of a wide range of hydrophilic metabolites within a single chromatographic run. For this purpose, we chose four independent potato genotypes characterized by modifications in sucrose metabolism. Using data-mining tools, including hierarchical cluster analysis and principle component analysis, we were able to assign clusters to the individual plant systems and to determine relative distances between these clusters. Extraction analysis allowed identification of the most important components of these clusters. Furthermore, correlation analysis revealed close linkages between a broad spectrum of metabolites. In a second, complementary approach, we subjected wild-type potato tissue to environmental manipulations. The metabolic profiles from these experiments were compared with the data sets obtained for the transgenic systems, thus illustrating the potential of metabolic profiling in assessing how a genetic modification can be phenocopied by environmental conditions. In summary, these data demonstrate the use of metabolic profiling in conjunction with data-mining tools as a technique for the comprehensive characterization of a plant genotype.

## INTRODUCTION

Enormous progress has been made over the last few years in the development of tools to create and characterize genetic diversity in plant systems. Transgenic knock-out populations, transposon insertions, chemical gene machines, and highly efficient ways to genotype single nucleotide polymorphisms within large populations have paved the way to a much broader base of diversity than imagined a few years ago (Aarts et al., 1993; Schaefer and Zryd, 1997; Strepp et al., 1998; Cho et al., 1999; Zhu et al., 1999). Furthermore, these developments have occurred in tandem with the elucidation of complete genomes and the rapid development of multiparallel technologies designed to access and describe the properties of biological systems (Celis et al., 2000). Most prominent among these new technologies has been the establishment of protocols for determining the expression levels of many thousands of genes in parallel, mRNA profiling. This is achieved by a process of mass hybridization reactions that use arrays of either expressed se-

quence tag or oligonucleotide collections representing large portions of the entire genome of the system in question (Lockhart et al., 1996; Ruan et al., 1998; Terryn et al., 1999; Aharoni et al., 2000; Richmond and Somerville, 2000). A second, albeit currently less advanced technology concerns the detection and quantification of the protein complement, or proteome, of a system (Shevchenko et al., 1996; Santoni et al., 1998; Chang et al., 2000).

Much attention has been focused on developing mRNA profiling and proteomic approaches, whereas the development of multiparallel techniques allowing analysis of the levels of low molecular weight compounds has been largely overlooked. This is true not only in plant sciences but across all biological disciplines, and thus this field is still in its infancy (Trethewey et al., 1999a). There are a few examples of this approach being applied to medical analyses (e.g., Duez et al., 1996; Matsumoto and Kuhara, 1996; Ning et al., 1996); however, only a handful of reports detail its application to plant systems (e.g., Adams et al., 1999; Katona et al., 1999).

We recently developed a method allowing the nonbiased, simultaneous, and rapid determination of metabolites in plants, using potato tubers or Arabidopsis as the experimental

<sup>1</sup>To whom correspondence should be addressed. E-mail Fernie@mpimp-golm.mpg.de; fax 49-331-5678408.

system. This technology, which combines gas chromatographic separation of compounds with a subsequent mass spectrometric identification, allows the simultaneous detection of >100 compounds within a single analysis (Fiehn et al., 2000; Roessner et al., 2000). In this article, the application of metabolic profiling to a variety of different genetically manipulated systems is described. We used various transgenic potato lines variously modified in sucrose metabolism as a first example for two main reasons. First, these lines have been extensively characterized previously by the use of classical biochemical approaches—a prerequisite that was important to authenticate data obtained from metabolic-profiling studies. Second, we specifically chose these examples because the applied genetic modifications targeted the same metabolic locus, that of sucrose degradation. This approach was taken to gain insight into the resolving power of metabolic profiling and to test its capacity to distinguish very similar situations.

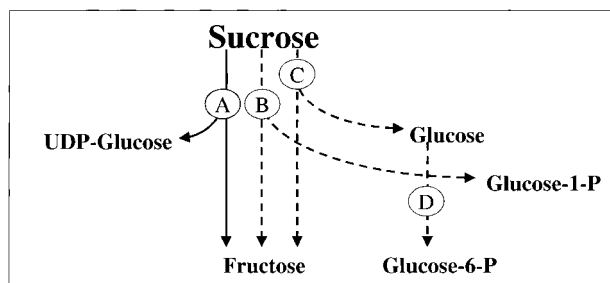
In addition to presenting the results of our analyses and interpreting their physiological implications, we also describe the application of data-mining tools to the data set obtained. These tools include hierarchical clustering and principal component analysis, detection of the metabolites determining the clustering behavior of the grouped plants, and a comprehensive analysis of the correlations between all metabolites of the various plants studied. Finally, we extended this analysis to include environmental manipulations of wild-type tissue in an attempt to produce phenocopies of the applied genetic manipulations. The data shown demonstrate that the application of data-mining tools to metabolic-profiling analysis allows insight into the relatedness of cer-

tain genetic situations. Moreover, correlation analysis allows the confirmation of established hypotheses concerning metabolic interactions within these systems. We believe these data further illustrate the use of metabolic profiling as an additional tool in multiparallel system analysis and as such demonstrate its importance for functional genomics.

## RESULTS

### Confirmation of Transgene Expression and Primary Metabolic Characterization

We grew transgenic potato plants that had been altered in tuber sucrose catabolism, as explained in Figure 1, in parallel under identical greenhouse conditions and then harvested samples from developing tubers. We chose the following transgenic lines for this study because the primary metabolic changes in these lines are well documented: INV-30, INV-33, and INV-42 (Sonnewald et al., 1997; Riedel, 1999); GK3-41, GK3-29, and GK3-38 (Trethewey et al., 1998); and SP-2, SP-11, and SP-29 (Trethewey et al., 2001). The introduced enzyme activities observed in extracts from these plants were similar to those we have previously reported (data not shown). Having confirmed that the plants do indeed express heterologous enzymes, we decided to verify that the primary changes within the potato tubers were in accordance with those previously determined (Figures 2A to 2D; Sonnewald et al., 1997; Trethewey et al., 1998, 2001; Riedel, 1999); notably, glucose levels were not increased in lines INV-42 and GK3-38. We then determined the levels of the other possible products of sucrose catabolism—UDP-Glc, glucose-6-P, glucose-1-P, and fructose-6-P (Figures 2E to 2H). The levels of all these compounds were in close agreement to those determined previously (Trethewey et al., 1998, 2001; Riedel, 1999), and as such they documented the suitability of these lines for further study.

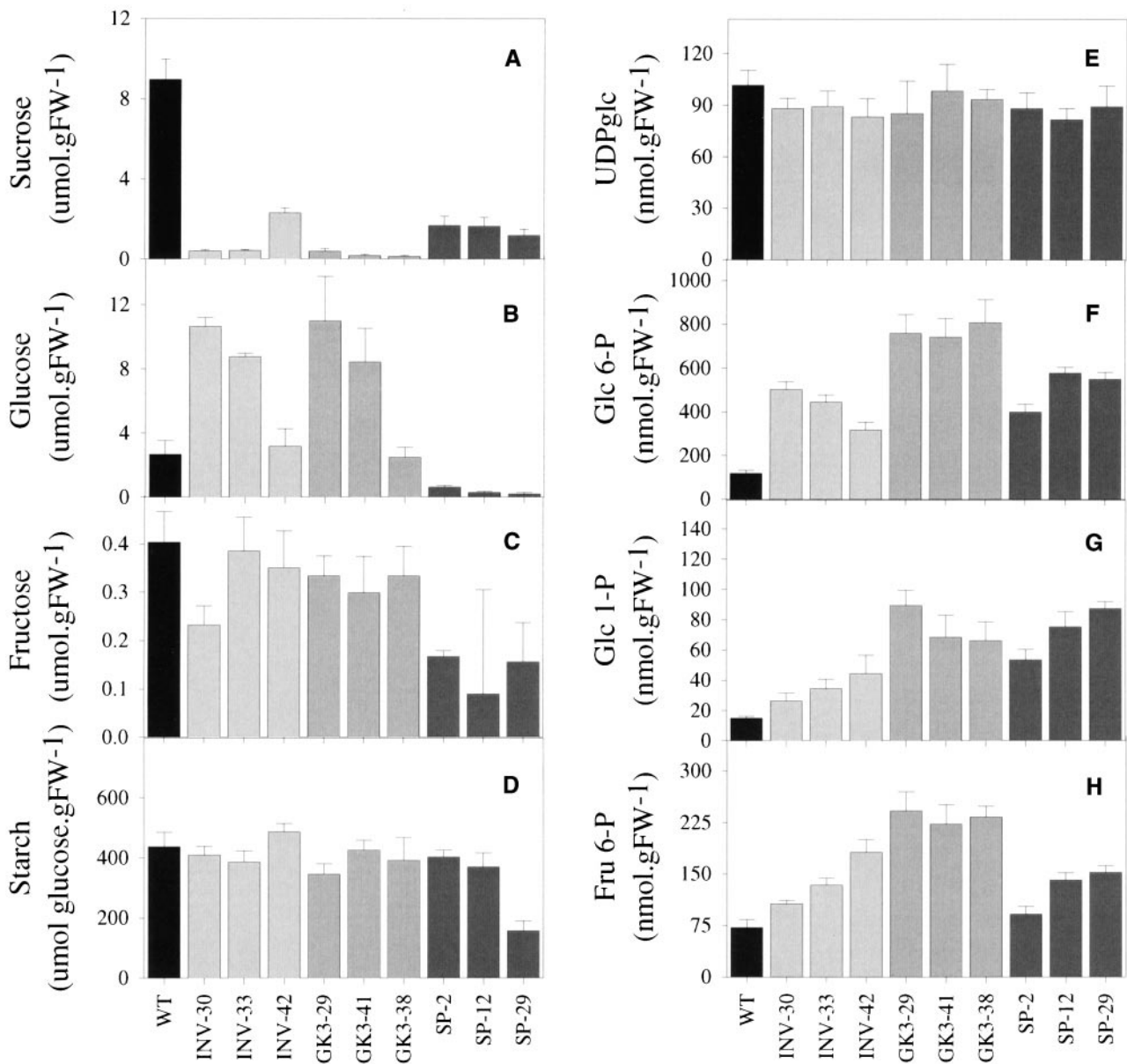


**Figure 1.** Substrates and Products of Endogenous and Introduced Reactions of Sucrose Catabolism within the Transgenic Potato Lines Studied.

The solid line (A) represents the plant's endogenous sucrose synthase, whereas the broken lines (B to D) represent the reactions catalyzed by the expression of a bacterial sucrose phosphorylase (SP lines), a yeast invertase (INV lines), and a bacterial glucokinase (expressed in combination with the yeast invertase; GK3 lines), respectively. Additional cofactors required for the reactions are UDP for sucrose synthase, Pi for sucrose phosphorylase, and ATP for glucokinase.

### Comparison of Relative Metabolite Levels within the Transgenic Tubers

Having confirmed that the transgenic lines were suitable for further experimentation, we extracted replicate samples from the same plants used for the preliminary characterization and then separated and characterized the detectable hydrophilic metabolite complement using a recently established gas chromatography–mass spectrometry (GC-MS) protocol (Roessner et al., 2000). Because of the large sample size of this experiment, we extracted a separate set of wild-type samples per transgenic line, despite the fact that all plants were grown in parallel under identical conditions. This allowed us independent references for each individual machine run. Results from this analysis are presented in



**Figure 2.** Sugar, Starch, and Sugar Phosphate Content of Transgenic Tubers.

Potato plants were grown in the greenhouse in 3.5-liter pots. Developing tubers were taken from plants harvested in the spring after 10 weeks of growth and while the plants were still fully green. Sucrose (A), glucose (B), fructose (C), and starch content (D) were determined in extracts from six individual plants per line. UDP-Glc (E), glucose-6-P (F), glucose-1-P (G), and fructose-6-P (H) were determined in extracts from four individual plants per line. All data are presented in  $\mu\text{mol.g}^{-1}$  fresh weight and represent the mean  $\pm$  SE. FW, fresh weight.

Table 1; the data set contains 88 metabolites (61 of which were defined with respect to their chemical nature), including sugars, sugar alcohols, amino acids, organic acids, and several miscellaneous compounds.

The majority of the compounds detected were found to al-

ter within the transgenic lines, in agreement with the data obtained using conventional spectrophotometric or HPLC methods (Figure 2; Trethewey et al., 1998, 1999b, 2001; Riedel, 1999). When taking mean values into consideration, we found that some interesting trends emerged (Table 1). In most

**Table 1.** Metabolite Levels in Wild-Type Developing Potato Tubers and in Tubers of Transgenic Potato Plants

metabolites	WT	SE %	I 30	SE %	I 33	SE %	I 42	SE %	WT	SE %	GK 38	SE %
1 alanine	1.00 ± 0.32		1.70 ± 0.17		2.43 ± 0.33		2.06 ± 0.35		1.00 ± 0.16		3.31 ± 0.15	
2 arginine	1.00 ± 0.20		1.66 ± 0.38		1.04 ± 0.24		1.99 ± 0.28		1.00 ± 0.20		2.43 ± 0.39	
3 asparagine	1.00 ± 0.19		1.31 ± 0.14		1.06 ± 0.09		1.57 ± 0.20		1.00 ± 0.05		1.21 ± 0.07	
4 aspartate	1.00 ± 0.11		1.20 ± 0.08		0.91 ± 0.09		1.04 ± 0.07		1.00 ± 0.03		1.18 ± 0.11	
5 b-alanine	1.00 ± 0.16		0.84 ± 0.05		1.17 ± 0.13		1.11 ± 0.16		1.00 ± 0.09		1.35 ± 0.12	
6 cysteine	1.00 ± 0.16		2.17 ± 0.22		2.33 ± 0.41		1.37 ± 0.27		1.00 ± 0.20		1.13 ± 0.08	
7 GABA	1.00 ± 0.10		1.20 ± 0.08		1.22 ± 0.11		1.40 ± 0.13		1.00 ± 0.13		0.79 ± 0.05	
8 glutamate	1.00 ± 0.07		1.22 ± 0.07		1.09 ± 0.07		1.16 ± 0.09		1.00 ± 0.04		1.02 ± 0.12	
9 glutamine	1.00 ± 0.26		0.39 ± 0.17		0.76 ± 0.20		1.17 ± 0.22		1.00 ± 0.16		0.65 ± 0.14	
10 glycine	1.00 ± 0.19		0.87 ± 0.07		0.93 ± 0.13		0.88 ± 0.10		1.00 ± 0.06		1.52 ± 0.12	
11 histidine	n.d.		n.d.		n.d.		n.d.		n.d.		565.33 ± 162.73	
12 homocysteine	n.d.		n.d.		n.d.		n.d.		n.d.		n.d.	
13 homoglutamine	n.d.		n.d.		n.d.		n.d.		n.d.		n.d.	
14 homoserine	1.00 ± 0.12		0.54 ± 0.08		0.90 ± 0.08		1.18 ± 0.16		1.00 ± 0.06		0.46 ± 0.02	
15 isoleucine	1.00 ± 0.19		1.00 ± 0.14		0.98 ± 0.14		1.07 ± 0.05		1.00 ± 0.23		0.93 ± 0.19	
16 leucine	1.00 ± 0.32		1.22 ± 0.23		1.11 ± 0.19		1.03 ± 0.06		1.00 ± 0.44		1.27 ± 0.22	
17 lysine	1.00 ± 0.16		0.50 ± 0.11		0.51 ± 0.08		0.94 ± 0.06		1.00 ± 0.25		0.70 ± 0.14	
18 methionine	1.00 ± 0.13		0.84 ± 0.10		0.84 ± 0.11		1.26 ± 0.11		1.00 ± 0.14		0.89 ± 0.13	
19 norleucine	1.00 ± 0.12		0.80 ± 0.13		0.60 ± 0.08		0.51 ± 0.04		1.00 ± 0.24		0.77 ± 0.14	
20 norvaline	1.00 ± 0.14		0.66 ± 0.07		0.57 ± 0.06		0.46 ± 0.05		1.00 ± 0.12		0.62 ± 0.08	
21 ornithine	1.00 ± 0.30		0.96 ± 0.22		1.17 ± 0.23		2.04 ± 0.32		1.00 ± 0.13		1.90 ± 0.28	
22 5-oxoproline	1.00 ± 0.18		0.78 ± 0.11		0.84 ± 0.19		0.99 ± 0.15		1.00 ± 0.06		1.02 ± 0.09	
23 phenylalanine	1.00 ± 0.20		1.73 ± 0.16		1.18 ± 0.16		1.17 ± 0.09		1.00 ± 0.13		1.89 ± 0.15	
24 proline	1.00 ± 0.18		0.69 ± 0.07		1.05 ± 0.09		1.20 ± 0.12		1.00 ± 0.08		1.53 ± 0.18	
25 serine	1.00 ± 0.20		2.84 ± 0.19		2.80 ± 0.22		2.44 ± 0.22		1.00 ± 0.08		4.02 ± 0.46	
26 threonine	1.00 ± 0.21		0.89 ± 0.10		1.05 ± 0.11		1.19 ± 0.10		1.00 ± 0.09		1.26 ± 0.16	
27 tryptophan	n.d.		450.83 ± 215.72		244.00 ± 28.56		390.50 ± 56.50		1.00 ± 0.45		17.48 ± 4.26	
28 tyrosine	1.00 ± 0.28		1.94 ± 0.32		0.87 ± 0.19		1.34 ± 0.16		1.00 ± 0.43		5.04 ± 0.66	
29 valine	1.00 ± 0.14		0.84 ± 0.08		1.01 ± 0.09		1.01 ± 0.08		1.00 ± 0.07		0.87 ± 0.10	
1 ascorbate	1.00 ± 0.31		0.68 ± 0.24		0.76 ± 0.16		1.46 ± 0.09		1.00 ± 0.11		2.39 ± 0.50	
2 citrate	1.00 ± 0.03		1.07 ± 0.05		1.12 ± 0.06		1.03 ± 0.03		1.00 ± 0.03		0.94 ± 0.07	
3 fumarate	1.00 ± 0.31		0.30 ± 0.03		0.24 ± 0.02		0.23 ± 0.03		1.00 ± 0.12		0.36 ± 0.05	
4 glucuronate	n.d.		n.d.		n.d.		n.d.		1.00 ± 0.11		0.03 ± 0.00	
5 glutarate	n.d.		74.00 ± 5.70		38.63 ± 4.24		24.00 ± 1.75		n.d.		n.d.	
6 glycerate	1.00 ± 0.08		6.64 ± 0.44		5.69 ± 0.59		3.54 ± 0.74		1.00 ± 0.08		6.33 ± 0.52	
7 isocitrate	1.00 ± 0.10		0.73 ± 0.08		1.04 ± 0.19		1.46 ± 0.13		1.00 ± 0.07		0.42 ± 0.07	
8 malate	1.00 ± 0.19		1.92 ± 0.21		1.98 ± 0.30		0.64 ± 0.12		1.00 ± 0.06		1.85 ± 0.17	
9 oxalate	n.d.		n.d.		n.d.		n.d.		1.00 ± 0.27		11.26 ± 1.37	
10 quinate	1.00 ± 0.12		1.35 ± 0.09		1.27 ± 0.12		1.51 ± 0.12		1.00 ± 0.10		0.78 ± 0.03	
11 shikimate	1.00 ± 0.17		2.70 ± 0.21		2.32 ± 0.58		1.50 ± 0.28		1.00 ± 0.08		0.64 ± 0.17	
12 succinate	1.00 ± 0.21		2.34 ± 0.19		2.14 ± 0.58		1.63 ± 0.25		1.00 ± 0.18		2.53 ± 0.39	
13 threonate	1.00 ± 0.13		1.38 ± 0.08		1.16 ± 0.11		0.91 ± 0.10		1.00 ± 0.05		1.24 ± 0.14	
1 fructose	1.00 ± 0.44		1.82 ± 0.52		5.67 ± 3.80		0.81 ± 0.13		1.00 ± 0.13		6.01 ± 2.55	
2 galactose	1.00 ± 0.35		0.60 ± 0.17		0.63 ± 0.16		0.25 ± 0.09		1.00 ± 0.16		0.17 ± 0.03	
3 glucose	1.00 ± 0.41		4.23 ± 0.57		3.56 ± 0.62		0.75 ± 0.47		1.00 ± 0.17		2.20 ± 1.33	
4 glycerol	n.d.		n.d.		n.d.		n.d.		1.00 ± 0.12		0.02 ± 0.00	
5 inositol	1.00 ± 0.11		0.18 ± 0.01		0.28 ± 0.04		0.43 ± 0.06		1.00 ± 0.20		0.40 ± 0.06	
6 isomaltose	n.d.		17.00 ± 2.05		n.d.		n.d.		n.d.		n.d.	
7 maltitol	n.d.		710.33 ± 105.64		384.00 ± 37.95		n.d.		n.d.		134.00 ± 40.13	
8 maltose	n.d.		9256.83 ± 1107.27		3834.00 ± 621.48		n.d.		n.d.		1979.00 ± 1053.58	
9 mannitol	1.00 ± 0.06		3.22 ± 0.23		2.75 ± 0.38		1.12 ± 0.13		1.00 ± 0.04		6.00 ± 0.53	
10 mannose	1.00 ± 0.38		7.02 ± 0.88		4.24 ± 0.74		0.85 ± 0.18		1.00 ± 0.11		1.19 ± 0.66	
11 sucrose	1.00 ± 0.25		0.13 ± 0.02		0.28 ± 0.06		0.27 ± 0.02		1.00 ± 0.09		0.04 ± 0.01	
12 trehalose	n.d.		223.33 ± 32.78		114.17 ± 17.36		n.d.		n.d.		32.33 ± 7.34	
1 fructose-6-P	1.00 ± 0.08		22.26 ± 2.77		17.84 ± 2.70		5.52 ± 1.30		1.00 ± 0.10		40.66 ± 10.85	
2 glucose-6-P	1.00 ± 0.21		17.82 ± 2.12		15.04 ± 2.21		4.90 ± 1.00		1.00 ± 0.08		30.52 ± 7.62	
3 phosphoethanolamine	1.00 ± 0.16		1.02 ± 0.08		1.13 ± 0.15		0.74 ± 0.05		1.00 ± 0.16		2.97 ± 0.62	
4 3-PGA	n.d.		89.88 ± 14.41		61.19 ± 11.90		39.11 ± 11.53		n.d.		75.67 ± 21.68	
5 6-P-gluconate	n.d.		479.33 ± 95.00		347.17 ± 53.43		213.00 ± 47.87		n.d.		177.50 ± 25.80	
6 phosphate	1.00 ± 0.02		0.94 ± 0.04		0.96 ± 0.07		0.83 ± 0.05		1.00 ± 0.03		0.88 ± 0.07	
spermidine	1.00 ± 0.11		1.88 ± 0.09		1.53 ± 0.09		1.69 ± 0.14		1.00 ± 0.08		1.33 ± 0.06	
1 UNKNOWN SP	n.d.		n.d.		n.d.		n.d.		n.d.		n.d.	
2 PT01	1.00 ± 0.05		0.74 ± 0.05		0.85 ± 0.04		0.61 ± 0.04		1.00 ± 0.05		0.73 ± 0.10	
3 PT02	1.00 ± 0.10		1.33 ± 0.14		1.11 ± 0.12		1.25 ± 0.13		1.00 ± 0.22		5.54 ± 0.59	
4 PT04	1.00 ± 0.07		0.88 ± 0.09		1.21 ± 0.12		0.70 ± 0.04		1.00 ± 0.10		0.82 ± 0.11	
5 PT05	1.00 ± 0.14		1.14 ± 0.08		0.92 ± 0.07		0.77 ± 0.04		1.00 ± 0.05		1.16 ± 0.14	
6 PT06	1.00 ± 0.13		1.13 ± 0.11		1.05 ± 0.08		1.48 ± 0.18		1.00 ± 0.09		1.23 ± 0.12	
7 PT07	1.00 ± 0.35		0.54 ± 0.18		0.93 ± 0.26		1.74 ± 0.42		1.00 ± 0.19		2.12 ± 0.57	
8 PT08	1.00 ± 0.05		0.76 ± 0.03		0.76 ± 0.05		0.95 ± 0.06		1.00 ± 0.10		0.73 ± 0.14	
9 PT09	1.00 ± 0.14		1.73 ± 0.10		1.26 ± 0.12		1.18 ± 0.06		1.00 ± 0.19		1.94 ± 0.11	
10 PT10	1.00 ± 0.23		0.54 ± 0.07		0.83 ± 0.14		1.00 ± 0.19		1.00 ± 0.14		1.13 ± 0.21	
11 PT11	1.00 ± 0.12		1.82 ± 0.08		1.55 ± 0.20		1.15 ± 0.13		1.00 ± 0.04		1.47 ± 0.05	
12 PT12	1.00 ± 0.24		1.23 ± 0.25		1.56 ± 0.30		0.73 ± 0.09		1.00 ± 0.19		0.87 ± 0.15	
13 PT14	1.00 ± 0.05		1.67 ± 0.10		1.80 ± 0.11		1.38 ± 0.07		1.00 ± 0.09		1.32 ± 0.12	
14 PT15	1.00 ± 0.18		1.48 ± 0.16		1.66 ± 0.20		2.90 ± 0.43		1.00 ± 0.10		2.02 ± 0.22	
15 PT16	1.00 ± 0.19		8.35 ± 0.61		6.65 ± 0.68		2.05 ± 1.04		1.00 ± 0.33		2.90 ± 1.36	
16 PT17	1.00 ± 0.22		1.41 ± 0.17		2.26 ± 0.60		2.65 ± 0.87		1.00 ± 0.14		2.65 ± 0.38	
17 PT18	1.00 ± 0.11		0.80 ± 0.06		0.90 ± 0.09		0.64 ± 0.09		1.00 ± 0.13		0.77 ± 0.16	
18 PT19	1.00 ± 0.41		9.51 ± 1.17		6.92 ± 1.43		0.89 ± 0.57		1.00 ± 0.27		2.47 ± 1.54	
19 PT20	1.00 ± 0.02		1.39 ± 0.13		1.34 ± 0.13		0.94 ± 0.08		1.00 ± 0.07		1.03 ± 0.04	
20 PT21	1.00 ± 0.20		2.62 ± 0.33		2.17 ± 0.32		0.90 ± 0.12		1.00 ± 0.05		2.62 ± 0.32	
21 PT23	1.00 ± 0.19		1.94 ± 0.19		1.35 ± 0.15		1.52 ± 0.12		1.00 ± 0.32		3.32 ± 0.31	
22 PT24	1.00 ± 0.17		2.15 ± 0.24		2.21 ± 0.18		2.98 ± 0.54		1.00 ± 0.05		1.78 ± 0.09	
23 PT25	1.00 ± 0.04		1.50 ± 0.08		1.22 ± 0.06		1.05 ± 0.06		1.00 ± 0.07		1.40 ± 0.05	
24 PT26	1.00 ± 0.08		0.64 ± 0.05		0.71 ± 0.06		0.54 ± 0.05		1.00 ± 0.06		0.66 ± 0.09	
25 PT27	1.00 ± 0.06		1.68 ± 0.06		1.37 ± 0.09		0.97 ± 0.05		1.00 ± 0.03		1.09 ± 0.03	
26 PT32	1.00 ± 0.09		1.12 ± 0.08		1.38 ± 0.15		1.01 ± 0.05		1.00 ± 0.06		1.04 ± 0.12	
27 PT33	1.00 ± 0.15		0.35 ± 0.04		0.47 ± 0.02		0.53 ± 0.05		1.00 ± 0.12		0.20 ± 0.03	

Continued

Table 1. (continued).

	GK 29	SE %	GK 41	SE %	WT	SE %	SP 2	SE %	SP 12	SE %	SP 29	SE %
1 alanine	3.00 ± 0.17		3.70 ± 0.43		1.00 ± 0.18		2.04 ± 0.14		1.73 ± 0.19		2.13 ± 0.06	
2 arginine	<b>3.04 ± 0.38</b>		<b>2.02 ± 0.26</b>		1.00 ± 0.16		<b>2.71 ± 0.29</b>		<b>2.64 ± 0.44</b>		<b>2.73 ± 0.44</b>	
3 asparagine	1.05 ± 0.05		1.01 ± 0.09		1.00 ± 0.33		1.28 ± 0.51		1.12 ± 0.28		1.14 ± 0.43	
4 aspartate	0.80 ± 0.08		0.89 ± 0.13		1.00 ± 0.05		<b>0.76 ± 0.04</b>		0.89 ± 0.03		<b>0.81 ± 0.04</b>	
5 b-alanine	1.26 ± 0.15		1.18 ± 0.07		1.00 ± 0.15		0.89 ± 0.10		1.03 ± 0.21		0.97 ± 0.08	
6 cysteine	1.79 ± 0.33		2.21 ± 0.51		1.00 ± 0.15		<b>2.33 ± 0.42</b>		<b>2.62 ± 0.31</b>		<b>1.61 ± 0.13</b>	
7 GABA	1.10 ± 0.13		1.04 ± 0.14		1.00 ± 0.04		1.04 ± 0.12		0.93 ± 0.12		1.02 ± 0.08	
8 glutamate	<b>0.69 ± 0.05</b>		<b>0.57 ± 0.12</b>		1.00 ± 0.03		0.90 ± 0.07		0.96 ± 0.06		<b>0.84 ± 0.04</b>	
9 glutamine	1.08 ± 0.16		1.03 ± 0.14		1.00 ± 0.28		0.82 ± 0.27		0.76 ± 0.18		0.85 ± 0.19	
10 glycine	<b>1.36 ± 0.12</b>		<b>1.88 ± 0.12</b>		1.00 ± 0.12		1.00 ± 0.11		1.00 ± 0.16		1.08 ± 0.06	
11 histidine	<b>751.83 ± 266.69</b>		n.d.		1.00 ± 0.32		1.41 ± 0.35		2.01 ± 0.32		1.52 ± 0.32	
12 homocysteine	n.d.		n.d.		n.d.		n.d.		<b>10.00 ± 1.70</b>		<b>5.00 ± 0.88</b>	
13 homoglutamine	n.d.		n.d.		n.d.		n.d.		<b>17.50 ± 5.03</b>		<b>11.50 ± 3.34</b>	
14 homoserine	<b>0.37 ± 0.02</b>		<b>0.01 ± 0.00</b>		1.00 ± 0.24		<b>0.01 ± 0.00</b>		<b>0.01 ± 0.00</b>		<b>0.34 ± 0.03</b>	
15 isoleucine	0.82 ± 0.09		<b>2.09 ± 0.19</b>		1.00 ± 0.20		1.15 ± 0.31		1.44 ± 0.23		1.58 ± 0.29	
16 leucine	1.27 ± 0.14		<b>3.35 ± 0.34</b>		1.00 ± 0.37		1.62 ± 0.55		1.94 ± 0.44		2.45 ± 0.68	
17 lysine	0.64 ± 0.05		0.83 ± 0.10		1.00 ± 0.26		1.31 ± 0.32		1.75 ± 0.27		<b>2.04 ± 0.35</b>	
18 methionine	0.91 ± 0.08		1.14 ± 0.06		1.00 ± 0.15		1.09 ± 0.15		1.21 ± 0.09		1.19 ± 0.11	
19 norleucine	0.83 ± 0.16		1.64 ± 0.24		1.00 ± 0.14		1.05 ± 0.33		1.44 ± 0.26		1.49 ± 0.27	
20 norvaline	0.64 ± 0.13		0.75 ± 0.11		1.00 ± 0.07		0.90 ± 0.11		0.92 ± 0.07		0.91 ± 0.08	
21 ornithine	<b>2.22 ± 0.29</b>		<b>3.15 ± 0.24</b>		1.00 ± 0.36		1.52 ± 0.35		2.79 ± 0.93		<b>3.53 ± 0.82</b>	
24 5-oxoproline	0.97 ± 0.07		<b>2.03 ± 0.26</b>		1.00 ± 0.04		<b>4.83 ± 0.30</b>		<b>5.27 ± 0.30</b>		<b>4.65 ± 0.87</b>	
22 phenylalanine	<b>1.67 ± 0.25</b>		<b>2.88 ± 0.08</b>		1.00 ± 0.16		1.46 ± 0.21		1.56 ± 0.22		<b>1.57 ± 0.17</b>	
23 proline	0.91 ± 0.07		0.98 ± 0.14		1.00 ± 0.13		0.83 ± 0.05		<b>0.63 ± 0.05</b>		<b>0.66 ± 0.07</b>	
25 serine	<b>2.81 ± 0.20</b>		<b>3.04 ± 0.18</b>		1.00 ± 0.13		<b>2.33 ± 0.15</b>		<b>1.98 ± 0.14</b>		<b>2.13 ± 0.39</b>	
26 threonine	1.14 ± 0.09		1.45 ± 0.07		1.00 ± 0.11		0.98 ± 0.10		1.12 ± 0.13		1.23 ± 0.13	
27 tryptophan	<b>12.79 ± 2.24</b>		<b>14.56 ± 5.25</b>		1.00 ± 0.39		<b>360.03 ± 103.64</b>		<b>488.21 ± 136.51</b>		<b>380.82 ± 113.52</b>	
28 tyrosine	<b>3.60 ± 0.56</b>		<b>5.55 ± 0.52</b>		1.00 ± 0.26		<b>2.36 ± 0.24</b>		<b>2.50 ± 0.28</b>		<b>2.56 ± 0.22</b>	
29 valine	0.80 ± 0.06		<b>1.29 ± 0.10</b>		1.00 ± 0.07		0.96 ± 0.08		0.95 ± 0.06		0.97 ± 0.08	
1 ascorbate	<b>2.42 ± 0.46</b>		0.91 ± 0.26		1.00 ± 0.33		<b>2.53 ± 0.38</b>		<b>2.91 ± 0.28</b>		2.07 ± 0.35	
2 citrate	1.00 ± 0.03		<b>0.69 ± 0.08</b>		1.00 ± 0.03		<b>0.62 ± 0.13</b>		1.07 ± 0.02		<b>1.31 ± 0.04</b>	
3 fumarate	<b>0.37 ± 0.07</b>		<b>0.52 ± 0.07</b>		1.00 ± 0.15		<b>0.27 ± 0.02</b>		<b>0.32 ± 0.03</b>		<b>0.22 ± 0.01</b>	
4 glucuronate	<b>1.60 ± 0.11</b>		<b>0.03 ± 0.00</b>		n.d.		n.d.		n.d.		n.d.	
5 glutarate	n.d.		n.d.		n.d.		n.d.		n.d.		n.d.	
6 glycerate	<b>7.27 ± 0.69</b>		<b>12.06 ± 1.56</b>		n.d.		<b>1255.17 ± 147.32</b>		<b>1087.50 ± 167.15</b>		<b>772.33 ± 129.86</b>	
7 isocitrate	<b>0.30 ± 0.08</b>		<b>0.16 ± 0.03</b>		1.00 ± 0.08		0.84 ± 0.05		0.93 ± 0.10		<b>0.70 ± 0.05</b>	
8 malate	<b>1.51 ± 0.10</b>		<b>2.15 ± 0.28</b>		1.00 ± 0.08		0.97 ± 0.08		0.71 ± 0.13		0.76 ± 0.11	
9 oxalate	<b>13.56 ± 1.87</b>		<b>9.04 ± 1.02</b>		1.00 ± 0.16		<b>0.46 ± 0.11</b>		<b>3.81 ± 0.37</b>		<b>2.63 ± 0.43</b>	
10 quinate	<b>0.65 ± 0.07</b>		<b>0.62 ± 0.07</b>		1.00 ± 0.05		0.98 ± 0.06		1.04 ± 0.06		1.19 ± 0.15	
11 shikimate	1.21 ± 0.17		0.95 ± 0.08		1.00 ± 0.12		0.71 ± 0.15		0.80 ± 0.11		0.67 ± 0.13	
12 succinate	<b>3.94 ± 0.56</b>		<b>3.23 ± 0.69</b>		1.00 ± 0.20		<b>1.72 ± 0.20</b>		0.95 ± 0.08		1.05 ± 0.07	
13 threonate	1.79 ± 0.34		<b>1.91 ± 0.16</b>		1.00 ± 0.04		0.90 ± 0.07		<b>0.70 ± 0.07</b>		<b>0.74 ± 0.07</b>	
1 fructose	<b>9.72 ± 2.01</b>		3.13 ± 1.07		1.00 ± 0.15		<b>2.54 ± 0.42</b>		0.58 ± 0.13		0.85 ± 0.24	
2 galactose	<b>0.43 ± 0.09</b>		<b>0.01 ± 0.00</b>		1.00 ± 0.46		2.09 ± 0.50		0.13 ± 0.03		0.22 ± 0.04	
3 glucose	<b>11.70 ± 1.88</b>		<b>9.59 ± 2.54</b>		1.00 ± 0.29		<b>0.16 ± 0.04</b>		<b>0.01 ± 0.00</b>		<b>0.02 ± 0.00</b>	
4 glycerol	<b>0.02 ± 0.00</b>		<b>0.02 ± 0.00</b>		1.00 ± 0.08		0.92 ± 0.02		<b>0.77 ± 0.05</b>		0.73 ± 0.08	
5 inositol	<b>0.18 ± 0.01</b>		<b>0.19 ± 0.05</b>		1.00 ± 0.29		0.36 ± 0.03		<b>0.20 ± 0.02</b>		<b>0.14 ± 0.01</b>	
6 isomaltose	n.d.		n.d.		n.d.		n.d.		n.d.		n.d.	
7 maltitol	<b>331.17 ± 43.33</b>		<b>640.83 ± 39.29</b>		n.d.		n.d.		n.d.		n.d.	
8 maltose	<b>6215.00 ± 952.51</b>		<b>8136.00 ± 1458.91</b>		n.d.		n.d.		n.d.		n.d.	
9 mannitol	<b>4.03 ± 0.20</b>		<b>3.73 ± 0.38</b>		1.00 ± 0.07		2.29 ± 0.82		<b>3.64 ± 0.55</b>		<b>2.45 ± 0.11</b>	
10 mannose	<b>5.50 ± 1.15</b>		<b>7.45 ± 1.44</b>		1.00 ± 0.23		<b>0.26 ± 0.05</b>		<b>0.03 ± 0.00</b>		<b>0.03 ± 0.00</b>	
11 sucrose	<b>0.04 ± 0.00</b>		<b>0.03 ± 0.00</b>		1.00 ± 0.10		<b>0.13 ± 0.02</b>		<b>0.06 ± 0.01</b>		<b>0.09 ± 0.02</b>	
12 trehalose	<b>149.00 ± 24.80</b>		<b>146.00 ± 16.74</b>		n.d.		n.d.		n.d.		n.d.	
1 fructose-6-P	<b>29.18 ± 1.26</b>		<b>11.77 ± 1.33</b>		1.00 ± 0.08		<b>8.55 ± 1.06</b>		<b>4.38 ± 0.48</b>		<b>3.56 ± 0.33</b>	
2 glucose-6-P	<b>23.49 ± 0.87</b>		<b>12.28 ± 0.68</b>		1.00 ± 0.07		<b>14.36 ± 1.72</b>		<b>7.46 ± 0.79</b>		<b>7.18 ± 0.65</b>	
3 phosphoethanolamine	<b>1.53 ± 0.13</b>		1.57 ± 0.35		1.00 ± 0.12		1.15 ± 0.10		1.35 ± 0.14		0.95 ± 0.07	
4 3-PGA	<b>47.17 ± 8.84</b>		n.d.		1.00 ± 0.04		<b>2.65 ± 0.18</b>		<b>3.82 ± 0.30</b>		<b>2.12 ± 0.13</b>	
5 6-P-glucuronate	<b>192.83 ± 41.72</b>		n.d.		n.d.		n.d.		<b>321.17 ± 31.60</b>		<b>197.33 ± 40.23</b>	
6 phosphate	0.84 ± 0.06		<b>0.57 ± 0.09</b>		1.00 ± 0.03		<b>0.42 ± 0.06</b>		1.04 ± 0.03		1.07 ± 0.01	
spermidine	<b>1.48 ± 0.04</b>		0.77 ± 0.08		1.00 ± 0.01		1.07 ± 0.06		0.90 ± 0.08		<b>0.68 ± 0.04</b>	
1 UNKNOWN SuPho	n.d.		n.d.		n.d.		<b>8405.83 ± 662.91</b>		<b>6114.50 ± 566.20</b>		<b>2987.17 ± 677.50</b>	
2 PT01	<b>0.63 ± 0.02</b>		<b>0.68 ± 0.05</b>		1.00 ± 0.07		<b>0.71 ± 0.03</b>		<b>0.43 ± 0.04</b>		<b>0.43 ± 0.03</b>	
3 PT02	<b>6.65 ± 0.85</b>		<b>6.00 ± 0.54</b>		1.00 ± 0.11		<b>3.38 ± 0.44</b>		<b>4.16 ± 0.49</b>		<b>9.17 ± 0.80</b>	
4 PT04	<b>0.74 ± 0.04</b>		1.47 ± 0.19		1.00 ± 0.06		<b>0.82 ± 0.04</b>		<b>0.55 ± 0.02</b>		<b>0.49 ± 0.03</b>	
5 PT05	<b>0.70 ± 0.06</b>		0.81 ± 0.16		1.00 ± 0.14		<b>0.34 ± 0.04</b>		<b>0.53 ± 0.09</b>		<b>0.41 ± 0.03</b>	
6 PT06	1.05 ± 0.07		<b>2.35 ± 0.27</b>		1.00 ± 0.09		0.90 ± 0.09		1.22 ± 0.12		0.91 ± 0.10	
7 PT07	<b>2.39 ± 0.50</b>		<b>4.92 ± 0.32</b>		1.00 ± 0.22		1.17 ± 0.18		<b>2.86 ± 0.71</b>		<b>8.66 ± 1.67</b>	
8 PT08	<b>0.54 ± 0.03</b>		<b>0.35 ± 0.03</b>		1.00 ± 0.06		<b>0.01 ± 0.00</b>		<b>0.01 ± 0.00</b>		<b>0.01 ± 0.00</b>	
9 PT09	<b>1.83 ± 0.22</b>		<b>3.92 ± 0.32</b>		1.00 ± 0.12		<b>1.69 ± 0.26</b>		<b>1.86 ± 0.26</b>		<b>2.08 ± 0.22</b>	
10 PT10	1.00 ± 0.14		<b>1.84 ± 0.17</b>		1.00 ± 0.08		<b>0.54 ± 0.06</b>		<b>0.55 ± 0.07</b>		<b>0.53 ± 0.07</b>	
11 PT11	<b>1.89 ± 0.08</b>		<b>1.62 ± 0.14</b>		1.00 ± 0.06		<b>1.59 ± 0.07</b>		<b>1.56 ± 0.07</b>		<b>1.32 ± 0.09</b>	
12 PT12	1.39 ± 0.11		0.77 ± 0.12		1.00 ± 0.15		<b>0.51 ± 0.08</b>		<b>0.56 ± 0.05</b>		<b>0.35 ± 0.03</b>	
13 PT14	<b>1.43 ± 0.08</b>		<b>2.20 ± 0.22</b>		1.00 ± 0.09		<b>1.30 ± 0.06</b>		1.23 ± 0.10		1.11 ± 0.10	
14 PT15	<b>2.39 ± 0.25</b>		<b>7.21 ± 0.86</b>		1.00 ± 0.12		<b>1.68 ± 0.17</b>		<b>5.07 ± 0.84</b>		<b>11.91 ± 1.78</b>	
15 PT16	<b>20.10 ± 2.56</b>		10.59 ± 4.41		1.00 ± 0.28		0.53 ± 0.16		0.01 ± 0.00		0.01 ± 0.00	
16 PT17	<b>5.45 ± 1.10</b>		<b>9.24 ± 1.92</b>		1.00 ± 0.24		0.52 ± 0.10		0.99 ± 0.13		<b>0.36 ± 0.06</b>	
17 PT18	<b>0.58 ± 0.04</b>		<b>0.44 ± 0.02</b>		1.00 ± 0.06		<b>0.65 ± 0.04</b>		<b>0.42 ± 0.04</b>		<b>0.41 ± 0.04</b>	
18 PT19	<b>13.54 ± 2.94</b>		<b>25.79 ± 4.93</b>		1.00 ± 0.14		<b>0.29 ± 0.02</b>		<b>0.14 ± 0.01</b>		<b>0.14 ± 0.03</b>	
19 PT20	<b>1.32 ± 0.09</b>		1.16 ± 0.06		1.00 ± 0.03		1.14 ± 0.06		1.14 ± 0.09		0.96 ± 0.11	
20 PT21	<b>2.17 ± 0.12</b>		<b>2.74 ± 0.21</b>		1.00 ± 0.09		<b>2.49 ± 0.24</b>		<b>1.48 ± 0.11</b>		1.21 ± 0.10	
21 PT23	<b>2.45 ± 0.31</b>		<b>7.29 ± 0.80</b>		1.00 ± 0.21		<b>3.26 ± 0.46</b>		<b>3.42 ± 0.55</b>		<b>3.52 ± 0.40</b>	
22 PT24	<b>2.32 ± 0.24</b>		<b>2.45 ± 0.53</b>		1.00 ± 0.08		<b>1.74 ± 0.14</b>		<b>5.58 ± 0.43</b>		<b>4.67 ± 0.45</b>	
23 PT25	<b>1.74 ± 0.24</b>		1.18 ± 0.08		1.00 ± 0.06		<b>1.33 ± 0.05</b>		1.34 ± 0.14		0.91 ± 0.05	
24 PT26	<b>0.53 ± 0.04</b>		<b>0.42 ± 0.02</b>		1.00 ± 0.05		0.91 ± 0.04		<b>0.66 ± 0.06</b>		<b>0.66 ± 0.04</b>	
25 PT27	<b>1.38 ± 0.08</b>		0.92 ± 0.09		1.00 ± 0.04		0.92 ± 0.03		<b>0.79 ± 0.08</b>		<b>0.59 ± 0.04</b>	
26 PT32	0.99 ± 0.07		0.95 ± 0.06		1.00 ± 0.05		0.98 ± 0.07		1.07 ± 0.08		0.93 ± 0.03	
27 PT33	<b>0.32 ± 0.04</b>		<b>0.23 ± 0.04</b>		1.00 ± 0.07		<b>0.35 ± 0.03</b>		<b>0.29 ± 0.02</b>		<b>0.21 ± 0.01</b>	

Metabolites were determined using the same samples from developing tubers as those used to measure starch, sugars, and sugar phosphates, as presented in Figure 2. Data are normalized to the mean response calculated for the wild type (WT) of each measured batch. (So that measured batches could be compared, individual wild-type values were normalized in the same way.) Values presented are the mean ±SE of six independent determinants. Those that are significantly different from the wild type are identified in boldface. n.d. indicates compounds that were not determined in a particular set of chromatograms.

instances, metabolite levels in the transgenics increased, which was consistent with the increased respiratory flux observed in these transgenics (Trethewey et al., 1998, 2001). Strikingly, but perhaps not surprisingly, the metabolite levels within several pathways tended to increase in tandem. Such concerted increases were exemplified by the amino acids, that is, the aromatic amino acids (phenylalanine, tryptophan, and tyrosine), all of which derive from shikimate, and also those deriving from 3-PGA (cysteine and serine). A further example was the coupled increase between arginine and ornithine; however, in this instance, the trend was not absolute. In contrast, when the data set for intermediates of the tricarboxylic acid cycle was considered, the novel finding that changes in the individual metabolites do not correlate became apparent. Succinate and malate levels generally increased, whereas the levels of citrate, isocitrate, and fumarate generally decreased. Because these trends are observable in the mean values, it is clear that a nonbiased correlation analysis that takes into account every single value within an independent sample may prove more revealing.

#### Appearance of Novel Metabolites in Chromatograms from Transgenic Tubers

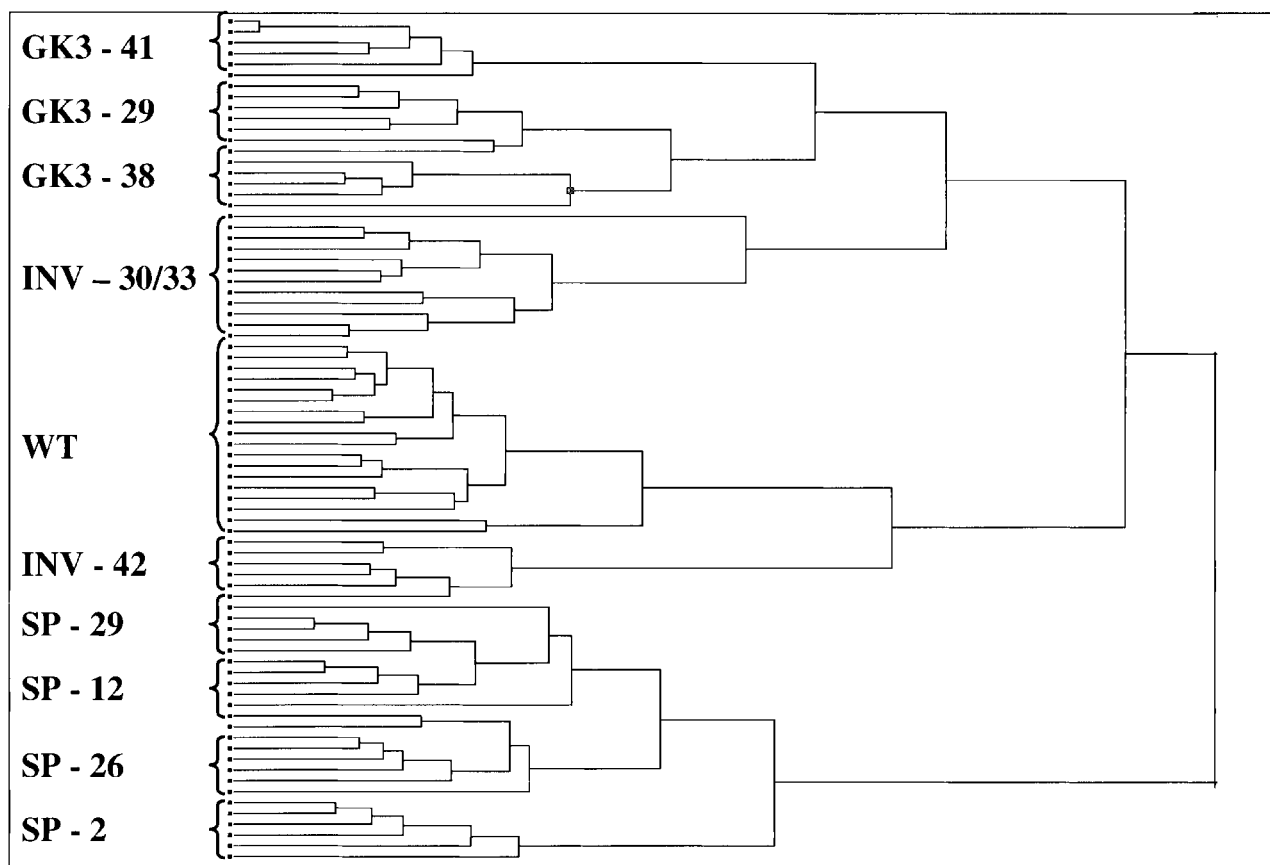
On first glance at the chromatograms, it became clear that several compounds were present in the transgenic lines that had not been detected in the chromatograms of wild-type tissue (data not shown). On inspection of the calibrated data, this observation was confirmed: nine of the 88 metabolites presented in Table 1 were below the level of detection in wild-type tubers. Some of these metabolites were observed in all of the transgenic types studied, whereas others were only present for a certain transgenic manipulation or even for a single transgenic line. In the first category are glutarate and 6-phosphogluconate. In contrast, maltose, trehalose, and maltitol become detectable in only INV and GK lines, whereas homoglutamine and homocysteine become detectable exclusively within SP lines. Isomaltose was only detected in line INV-30. In addition, an unknown peak, PT00, which is bigger in magnitude than that of sucrose, appears in the SP lines. It is conceivable that this peak is the result of a side reaction catalyzed by sucrose phosphorylase, and the mass spectrum suggests that it contains a glucosyl residue. This observation is consistent with the findings of Kitao and co-workers, who performed detailed characterization of the side reactions of sucrose phosphorylase from *Leuconostoc mesenteroides* (Kitao and Sekine, 1994a, 1994b) and found that this enzyme was also capable of transferring glycosyl residues to a wide range of acceptors. Unfortunately, when we analyzed the following commercially available compounds (found to be products or constituents of products of the sucrose phosphorylase in *L. mesenteroides*)—arbutin, catechin-glucosides, kojic acid, kojibiose, and nigerose—none of them co-eluted with the

unknown peak. The final elucidation of the exact chemical structure therefore requires further study.

#### Hierarchical Cluster Analysis and Principle Component Analysis of Steady State Metabolite Concentrations in the Transgenic Tubers

It is clear from the preceding paragraphs that analysis of such a large data set is a daunting task. It is even more so when the genetic diversity in question is centered around primary carbohydrate metabolism and the number of changes in steady state metabolite pool size is as large as that observed here (Figure 2 and Table 1). For this reason, we decided to apply bioinformatic tools to our data set. Given that there is a fair degree of natural variation between samples for many of the metabolites in question, we chose to plot all individual chromatograms rather than the mean values presented in Table 1 to assess whether individual transformants and/or transgenics exhibited similar behavior with respect to their total metabolic profile. When we applied hierarchical cluster analysis (HCA) to our data set, as shown in Figure 3, we found that all 18 wild-type samples clustered as a single distinct group; likewise, all the GK3 and SP lines clustered by both the nature of the transformation and the magnitude of the introduced activity. In contrast, the INV lines did not cluster in the same manner; rather, INV-30 and INV-33 formed a single cluster that was closer to the GK3 lines than to the wild type, whereas INV-42 was closest to the wild type. The fact that INV-42 is the line closest to the wild type is interesting in that it is also the line that exhibits the lowest invertase activity.

We then took a second, complementary approach of applying principle component analysis (PCA) to our data set. PCA uses an  $n$ -dimensional vector approach to separate samples on the basis of the cumulative correlation of all metabolite data and then identifies the vector that yields the greatest separation between samples. The results from the chosen vector were then displayed in two dimensions (Figure 4A). Once again, wild-type tubers constituted a single cluster, and INV-42 samples clustered independently of all other lines. Furthermore, the SP lines formed a distinct cluster, and differences between individual samples of the SP lines appeared to be related to the activity of the introduced enzyme. However, contrary to the observations made using HCA, the GK lines and INV-33 and INV-30 were not too distinct, and considerable overlap between INV-30 and all the GK3 lines (to which it served as a parental line) exists. That said, within each cluster of transgenics, subclusters that represent individual transgenic plants can be easily recognized. When taken together, the results from both types of cluster analysis are in close agreement, with the exception of the resolution of INV-30, INV-33, and GK3 lines, and indicate that despite the fact that sucrose phosphorylase acts on the same target molecule as does invertase, plants ex-



**Figure 3.** Dendrogram Obtained after HCA of the Metabolic Profiles of the Analyzed Transgenic Systems.

The distances between these populations were calculated as described in Methods, using the normalized data of the single measurements from which the means presented in Table 1 are derived. In addition, data from a further line, SP-26, are included that were not presented previously. Wherever possible, individual branches are grouped in brackets for ease of reading. WT, wild type.

pressing sucrose phosphorylase have a clearly distinct metabolic profile.

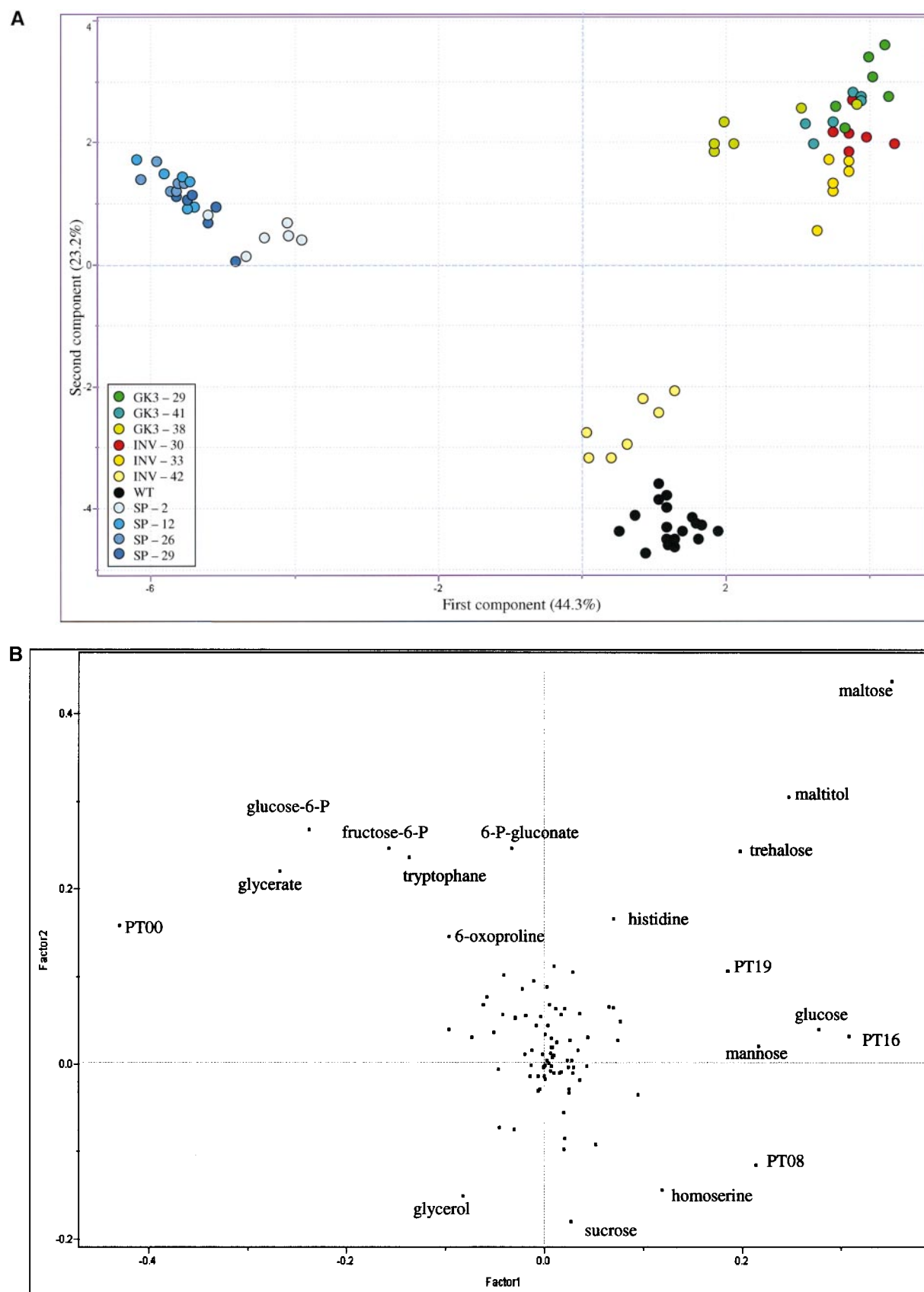
#### Assessment of the Metabolites Exerting the Largest Influence on Cluster Formation

Using the vector-based approach of PCA, it is possible to distinguish the compounds that exhibit the greatest variance within a population and thereby distinguish the contribution of these compounds to the formation of distinct clusters. When the steady state levels of metabolites within these four genotypes were compared, the main contributors to the cluster formation were determined to be sugars or closely related compounds, including the aforementioned PT00, maltose, maltitol, trehalose, glucose and mannose, glycerate, both glucose-6-P and fructose-6-P, and unknowns PT08 and PT16 (Figure 4B). Given that the novel unknown PT00, which was detectable only in the SP lines, is a

major contributor to PCA, we also calculated a vector in which this component was omitted from the PCA. This omission resulted in no changes in the clusters formed by either HCA or PCA. Further studies performed that removed all novel components of the metabolite profiles also did not result in gross changes in the clustering patterns produced by either method of component analysis (data not shown). These data demonstrate that our approach indeed takes into account the entire spectrum of metabolites detailed; it does not merely compare the behavior of a single, or a few, metabolite(s) within these lines, and thus, it provides validation of these analyses.

#### Correlation of Metabolite Levels and Analysis of Dependencies

These methods exemplify how interpretations can be made on the basis of the entire metabolic complement of a system.



**Figure 4.** PCA of the Metabolic Profiles of the Analyzed Transgenic Systems.



Specifically, they show how genetically distinct systems can be identified and how the most important component(s) of this phenotype can be determined. Although this technique is clearly very powerful, we decided also to look at the trends within the individual metabolites by plotting the level of every metabolite in individual samples of the wild type and various transgenic lines against every other metabolite within that sample. We analyzed a total of 3872 such plots. The observed dependencies could be classified into three major groups: those exhibiting no dependency (i.e., scatter), those exhibiting linear correlation, and those exhibiting a more complex correlation. As would perhaps be expected, most plots were of metabolite levels that were independent of each other; however, several interesting results came to light during this analysis. (A complete list of metabolites exhibiting correlations with a coefficient  $>0.7$  is presented on our web page at <http://www.mpimp-golm.mpg.de/willmitzer/metabolic-profiling-e.html>.) For discussion of these data, we have presented a few examples of some of the trends (Figure 5). An obvious example of a linear correlation between metabolites is that observed for fructose-6-P and glucose-6-P (Figure 5A), which are separated by only a single enzyme, phosphoglucose isomerase, which catalyzes a near-equilibrium reaction. This correlation holds for all lines, confirming that the potato tuber has a very high capacity for equilibrating these two metabolites. However, this relationship was also observed in situations in which metabolites are not consecutive within a pathway, for example, between leucine and isoleucine (Figure 5B). When we consider that the pathways for the biosynthesis of these amino acids share the same terminal enzyme activity, branched chain amino acid transaminase, and the same cofactor, glutamate, the reason for the close relationship between the metabolites becomes apparent. Methionine and lysine (Figure 5C) display a nonlinear correlation, which is most pronounced in GK3 and SP lines; this correlation seems to be in agreement with the relationship that one would predict from proposed models of feedback regulation (Bartlem et al., 2000; see Discussion for detailed explanation). Perhaps even more exciting are cases in which the relationship between metabolites *a* and *b* is different in different genotypes. One example of this is shown in Figure 5D, in which the glycine level is plotted against sucrose. However, this

example is trivial because in each case, the genetic modification introduced is targeted at sucrose, and therefore a different linear regression would be expected between transgenic lines and the wild type. A more informative example is provided in Figure 5E, in which PT07 is plotted against PT15; here, it can be clearly seen that in GK3 and SP lines, the metabolites show a different dependency than they do in wild-type and INV lines. This type of analysis may also prove crucial in identifying unknown compounds, because in several instances (e.g., as shown in Figure 5F, in which unknown PT19 is plotted against mannose), the observation that the level of an unknown exhibits strong positive correlation with the increase in level of a known metabolite provides hints about the biosynthesis or subsequent metabolism of that compound.

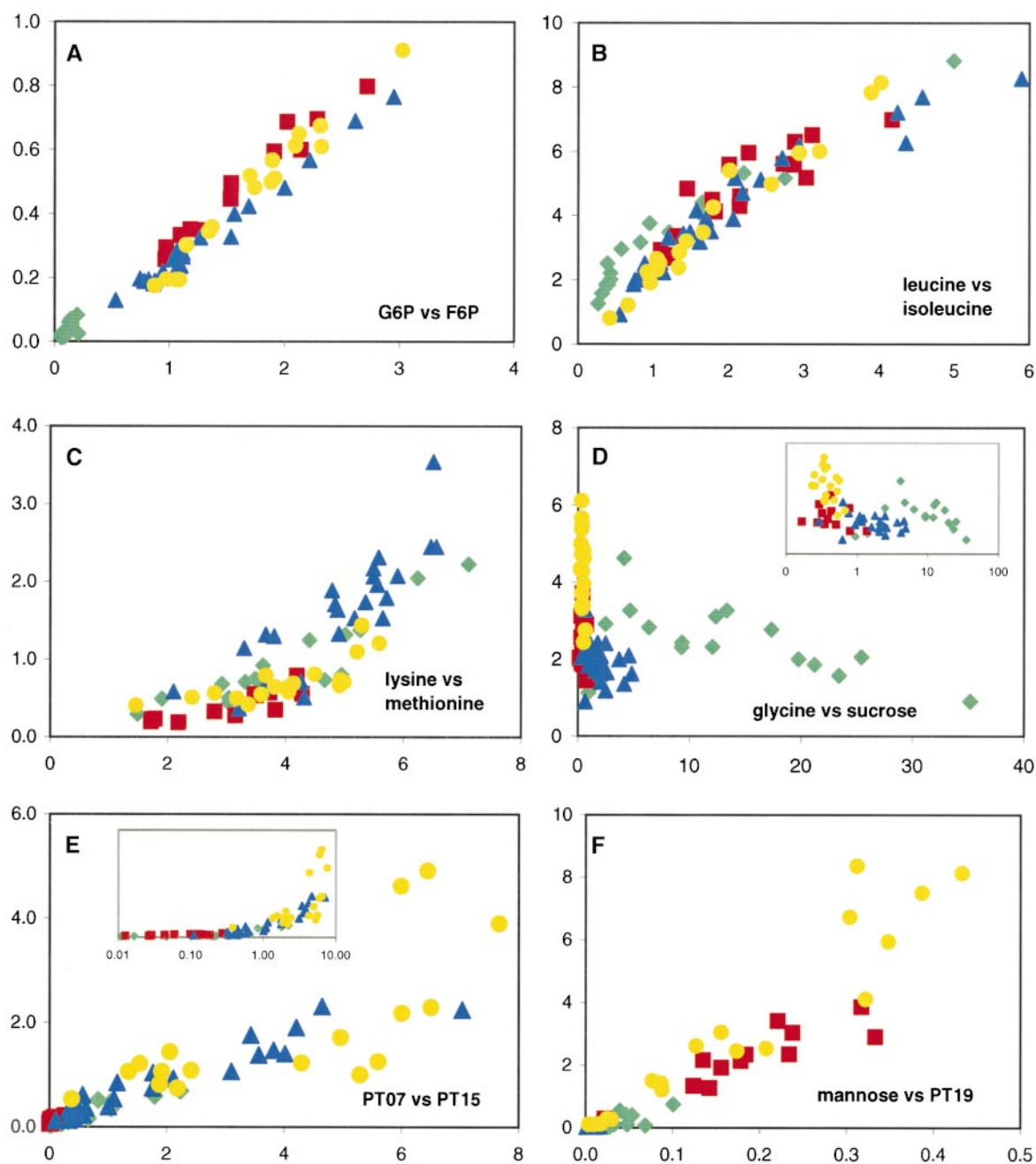
#### Effect of Environmental Perturbation on the Steady State Metabolite Levels in Wild-Type Tuber Discs

As a further example of the use of metabolic profiling, we investigated the metabolite levels of wild-type tuber discs incubated for 2 hr in 0, 20, 50, 100, 200, or 500 mM glucose, corresponding to cellular glucose levels of 1.7, 8.0, 9.4, 18.1, 30.2, and 68.2  $\mu\text{M g fresh weight}^{-1}$ , respectively (mean,  $n = 4$ ), as measured by GC-MS. The levels of more than half of the 86 compounds we measured were found to decrease on incubation in comparison with the nonincubated controls, irrespective of the presence or absence of glucose (experimental data are available at <http://www.mpimp-golm.mpg.de/willmitzer/metabolic-profiling-e.html>). Despite the large changes caused by incubation, when the levels of compounds in discs incubated in glucose were contrasted with the levels of those found in samples incubated in buffer (10 mM Mes-KOH, pH 6.5) alone, a different picture emerged. Only samples that were incubated in 200 and 500 mM glucose exhibited significant differences. The exceptions to this include malate and glucose-6-P, whose levels significantly increased on incubation in 100, 200, or 500 mM glucose and mannitol, and of course glucose, whose levels increased after incubation in glucose at all concentrations tested. Again, several of the unknown compounds displayed patterns of change similar to those for

**Figure 4.** (continued).

**(A)** The distances between these populations were calculated as described in Methods, using the log-transformed, normalized data of the single measurements from which the means presented in Table 1 are derived. In addition, data from a further line, SP-26, are included that were not presented previously. PCA vectors span a 10-dimensional space to give best sample separation, with each point representing a linear combination of all the metabolites from an individual sample. Vectors 1 and 2 were chosen for best visualization of differences between genotypes and include 67.5% of the information derived from metabolic variances. WT, wild type.

**(B)** The contribution of individual metabolites to the PCA vector calculation by linear combination. The closer to the origin, the smaller the influence a given metabolite has on the linear combination. The most important metabolites for separation of the differently treated samples are labeled.



**Figure 5.** Correlation between Metabolite Levels of the Analyzed Transgenic Systems.

Correlations between the relative response ratios of each of the 88 metabolites with those of all other metabolites were assessed, and several interesting trends were observed. Examples of these trends are shown. Symbols are as follows: green diamonds, wild type; red squares, INV; yellow circles, GK3; blue triangles, SP.

(A) Glucose-6-P (G6P) versus fructose-6-P (F6P).

(B) Leucine versus isoleucine.

(C) Lysine versus methionine.

(D) Glycine versus sucrose. The insert shows sucrose values plotted on a logarithmic scale.

(E) PT07 versus PT15. The insert shows PT07 values plotted on a logarithmic scale.

(F) Mannose versus PT19.

compounds for which we know the chemical nature. This is in itself interesting, but it may also indicate chemical similarity between the correlating metabolites and thus may help in identifying the unknown metabolites.

### HCA and PCA of the Metabolic Complement of Glucose-Incubated Samples

Applying cluster analysis to the data from glucose incubation (i.e., using wild-type steady state metabolite concentrations) revealed interesting results. HCA showed that the wild-type tuber discs incubated in buffer alone had the most similar metabolite complement to the steady state wild-type levels (Figure 6). Furthermore, the glucose-fed samples formed a distinct cluster that was more similar to the wild-type steady state complement than to that of any of the transgenics, and the metabolic profile of the discs fed with 500 mM glucose was distinct from the profile of the discs fed with lower concentrations of glucose. The relationship between the transgenic lines shown in Figure 6 is different from that shown in Figure 4; however, this is an inherent feature of this type of component analysis, because a new hierarchy is established.

PCA revealed very similar trends (Figure 7). Furthermore, when we used this method of clustering, the buffer-incubated samples were indistinct from the wild-type steady state levels. In addition, the 20 to 200 mM glucose-fed samples form an independent cluster, as do the 500 mM glucose-fed samples, and these clusters are closer to the wild-type steady state cluster than to any of those of the transgenics. Moreover, when glucose-fed samples and their respective controls were clustered independently of the transgenic lines, the same clustering pattern was formed (data not shown). Figure 8 reveals that the compounds that exhibited the greatest variance when the metabolic profiles of the glucose-fed samples were considered alone were asparagine, glucose, maltose, proline, tryptophan, PT10, PT14, PT16, and PT19. In contrast, the most important components for the clustering when both genetically and environmentally modified systems were compared were mainly sugars and sugar derivatives, including glucose, maltose, mannose, maltitol, trehalose, 6-phosphogluconate, both of the hexose-6-phosphates, and the unknown PT00 (Figure 9). For both analyses, we reevaluated the data sets, this time omitting the novel compounds; however, the observed clustering was remarkably similar in both instances (data not shown).

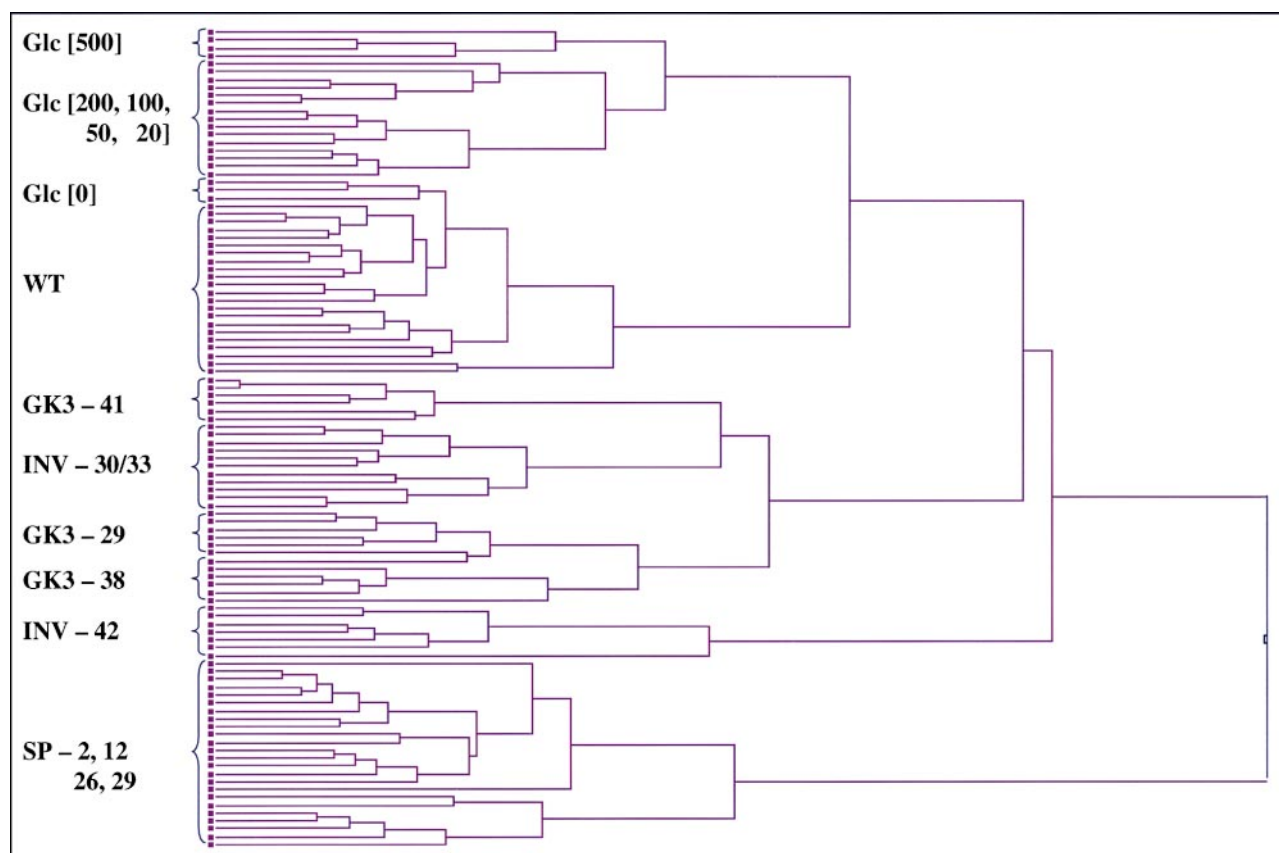
When the glucose-feeding region of the PCA was expanded (Figure 10A) and the individual incubations were highlighted, a clear trend of increased distance from the wild type (steady state) occurred with increased concentration of glucose in the incubation medium. One of the primary aims of this experiment was to attempt to phenocopy the metabolic complement of the transgenics by using environmental manipulation. This was clearly not possible in this instance

because all situations formed distinct clusters. Thus, we decided to broaden our analyses by comparing the metabolic profile of transgenic potatoes expressing invertase at an apoplastic location that we had previously measured (Roessner et al., 2000) with that of the glucose-fed samples by using both methods of cluster analysis. We were fascinated to find that PCA showed that the apoplastic invertase samples co-clustered with the glucose-fed samples (Figure 10B); furthermore, both HCA and an in-depth search of the changes in the chromatograms confirmed this result (data not shown). To our knowledge, this result in which a genetically modified system has been phenocopied by modification of environmental conditions is novel. When the data from the apoplastic invertase-expressing potatoes were considered alongside the data from the other transgenics and the wild-type and glucose-fed samples, the most important components for the clustering turned out to be identical to those when only the INV, GK3, and SP lines were compared with the wild-type and glucose-fed samples, except that fructose became marginally more important (see Figure 9). Once again, we reevaluated the cluster analysis by omitting novel compounds to determine whether the clustering pattern observed was reflective of the entire metabolic complement or whether the control of cluster formation was vested in merely a few compounds. As in other cases reported here, the clustering pattern that resulted from this PCA was essentially the same as the one obtained when all metabolites were considered.

## DISCUSSION

This study investigated the potential of metabolic profiling, using GC-MS for phenotyping and comprehensive characterization of plant systems. We selected the particular transgenic lines because they have already been fairly well characterized at the metabolite level, and they differ only slightly in the biochemical activity that is always targeted at enhancing sucrose cleavage. The similarity of the data obtained using the GC-MS protocol compared with that previously obtained using conventional HPLC and spectrophotometric methods (Trethewey et al., 1998, 1999b, 2001; Riedel, 1999) validates the authenticity of the measurements obtained using this protocol, and therefore confirms the protocol's suitability for use in this study.

Although a major purpose of this study was to combine multiparallel metabolite analysis with bioinformatic tools for data analysis, the comprehensive analysis achieved by metabolite profiling alone allowed some important conclusions to be made. For example, the fact that so many amino acids increased in the lines investigated in this study is in itself fascinating. There are two possible mechanisms for these increases: either an increased synthesis in source tissues and an increased transport of amino acids to the tubers or an elevated rate of amino acid biosynthesis within the



**Figure 6.** Dendrogram Obtained after HCA of the Metabolic Profiles of both Genetically and Environmentally Modified Systems.

The vector for the HCA described in Figure 3 was recalculated to include the metabolic profiles achieved after incubation of wild-type (WT) potato tuber tissue in a range of glucose concentrations. Thus, the full data set used was the individual measurements of samples from all transgenic lines as well as individual measurements from glucose-fed wild-type tissue (from which the mean data, presented on our web page at <http://www.mpimp-golm.mpg.de/willmitzer/metabolic-profiling-e.html>, were derived). Wherever possible, individual branches are grouped in brackets for ease of reading. Note that the relationship between the various transgenic lines is different from that observed in Figure 4. This is an inherent feature of this form of cluster analysis because a new hierarchy is established.

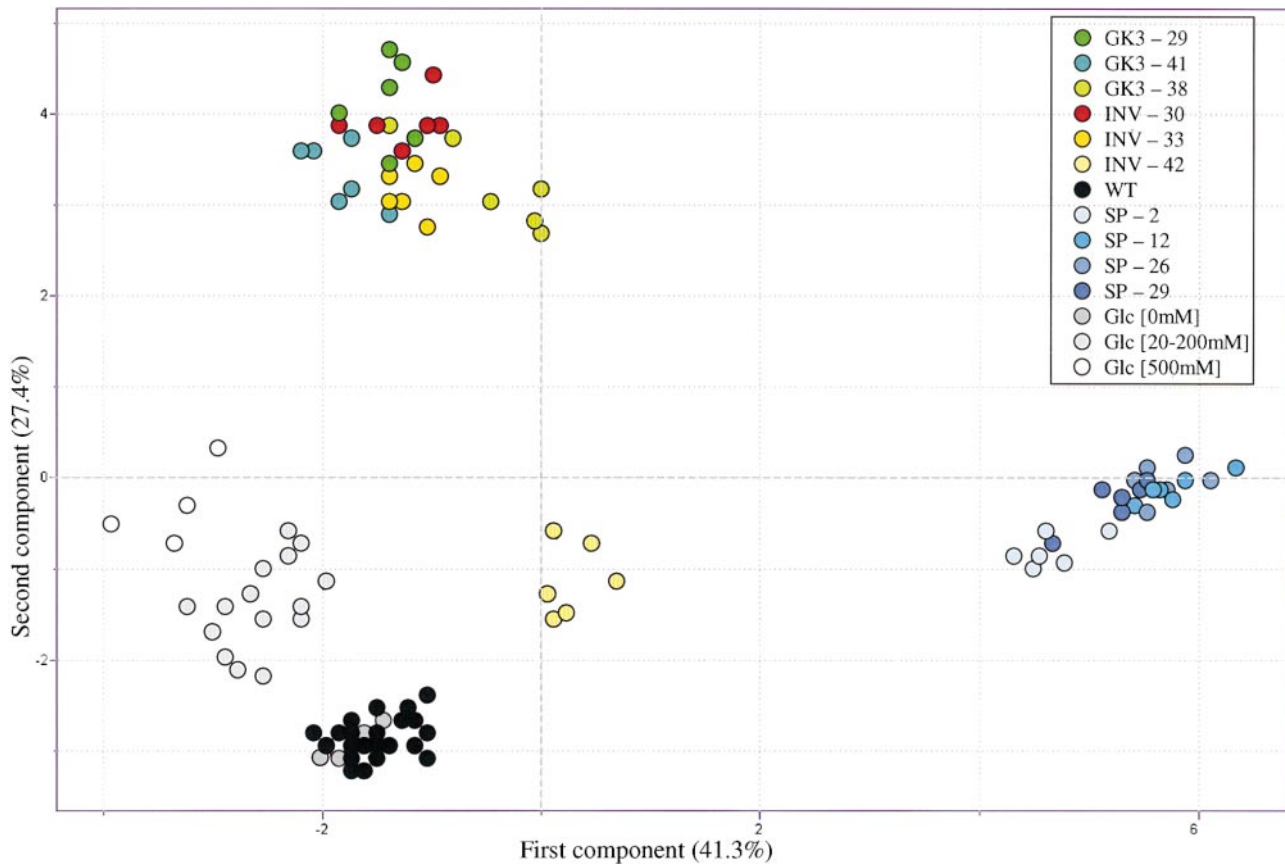
potato tuber. GC-MS analysis revealed that the amino acid levels in the leaves of the transgenic plants did not change (data not shown). These data are what would be expected given the use of a tuber-specific promoter for transgene expression, and they indicate that the later hypothesis is the more likely. Amino acid biosynthesis in potato tubers in particular and in storage tissues in general is poorly understood. Although recently several genes for amino acid biosynthesis have been cloned from the potato tuber (Muday and Herrmann, 1992; Riedel et al., 1999; Casazza et al., 2000; Maimann et al., 2000), it is not known whether tubers possess the necessary machinery to synthesize all amino acids. The data presented in this study, although indirect, provide the first evidence that the potato tuber is likely to contain the required machinery to produce all amino acids de novo. This example illustrates clearly the power of met-

abolic profiling in functional genomics in that compounds are identified that imply the presence and influence of gene products involved in their synthesis. A further example is that of ascorbate because little is known about the location of synthesis of this vitamin (Smirnoff and Wheeler, 2000). The data presented here indicate that ascorbate can also be synthesized de novo within the tuber: ascorbate is increased in the tubers in several of the transgenic lines studied but not in their leaves (data not shown). Ascorbate also increased after incubation of wild-type tuber tissue in glucose, so ascorbate synthesis is possible in the tuber, at least under conditions in which glucose is plentiful. These examples therefore return our attention to searching for genes. Furthermore, we anticipate that once the chemical nature of the unknowns is established, clear new targets for gene discovery will be identified.

A further example of the type of conclusions that can be drawn from such a broad-based profiling method involves differences in metabolic profiles that can be assigned on the basis of difference in the genetic manipulation imposed. Interestingly, several compounds, namely, maltose, trehalose, isomaltose, maltitol, malate, PT16, PT19, and PT20, increase starkly only in the INV and GK lines that also exhibit elevated glucose. This observation fits with recent results suggesting the operation of sugar-sensing mechanisms within plants (Jang et al., 1997; Smeekens, 2000). However, although these changes can be correlated directly to glucose levels, they are limited to only a few metabolites; thus, these data seem to argue against a major signaling role for glucose within the tuber system. These findings are therefore in agreement with previous studies in which we directly modified the levels of glucose and of glucose-using en-

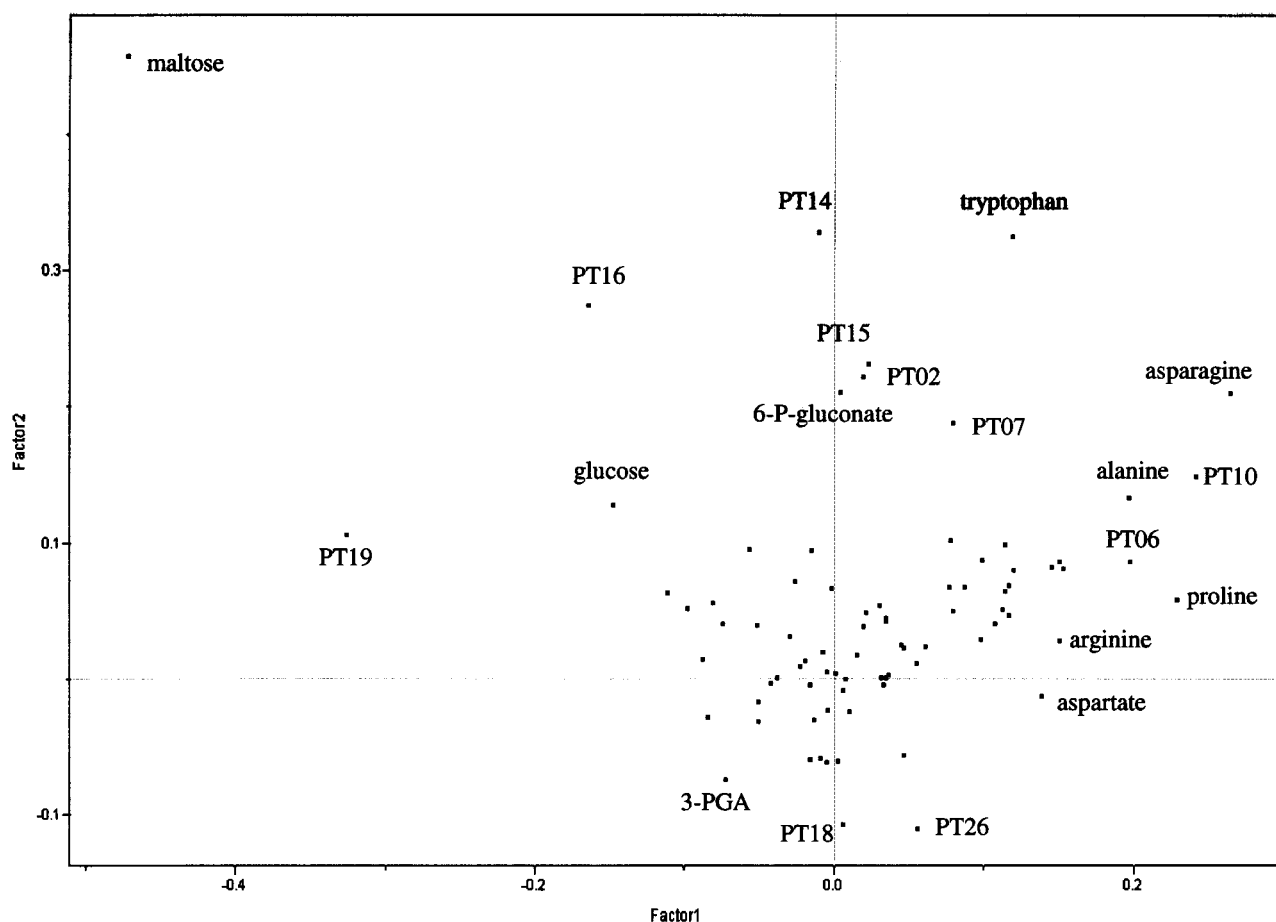
zymes by using a transgenic approach (Veramendi et al., 1999; Fernie et al., 2000).

The above examples clearly illustrate the power of a non-biased metabolic screen to help us draw conclusions from our data that have both breadth and novelty. However, analysis at the level of single metabolites is an ominous task, particularly when ~11,000 data points must be assessed. We therefore applied bioinformatic tools for data mining to our results. The four initial genotypes analyzed (wild type, INV, GK3, and SP potato lines) had distinct metabolic profiles, despite the fact that the target of the genetic manipulation was the same in each instance. Both methods of cluster analysis independently led to the same interpretation and gave a high level of resolution between the genotypes. Moreover, reanalysis of the data sets when the most influential contributing metabolites of the individual clusters were



**Figure 7.** PCA of the Metabolic Profiles of Both Genetically and Environmentally Modified Systems.

The distances between these populations were calculated as described in Methods by using the log-transformed, normalized data of the single measurements from which the means presented on our web page (<http://www.mpimp-golm.mpg.de/willmitzer/metabolic-profiling-e.html>) were derived. PCA vectors span a 10-dimensional space to give best sample separation, with each point representing a linear combination of all the metabolites from an individual sample. Vectors 1 and 2 were chosen for best visualization of differences between genotypes and include 68.7% of the information derived from metabolic variances. WT, wild type.



**Figure 8.** Assessment of the Metabolites Exerting the Largest Cluster Formation When Only Glucose-Fed Samples Are Considered.

Shown is the contribution of individual metabolites to the PCA vector calculation by linear combination. The closer to the origin, the smaller the influence a given metabolite has on the linear combination. The most important metabolites with respect to the separation of the differently treated samples are labeled.

removed yielded very similar clusters in all instances. We believe that this allows us to have a high level of confidence in interpreting these clusters because it demonstrates that the cluster formation is not dependent on merely a few metabolites or even a single metabolite. That the combination of metabolic profiling and cluster analysis allows resolution of very similar situations suggests that it is of general use for phenotyping diverse genetically or environmentally modified plant systems.

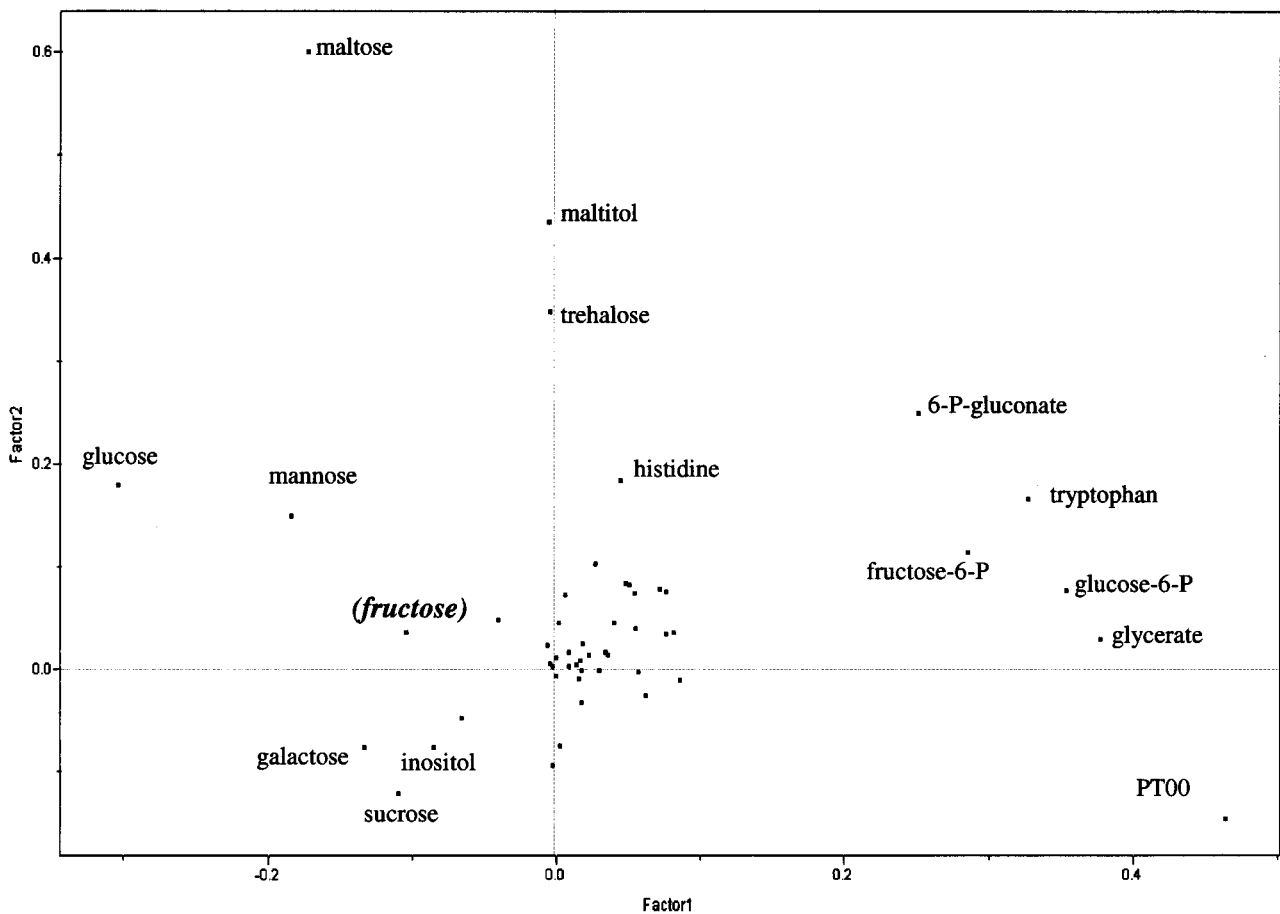
A further advantage of using a multiparallel approach is that because all metabolites are analyzed within a single extract, relationships between the levels of the various metabolites can be determined. By plotting all possible correlations, we were able to assess which metabolite concentrations were strongly linked. Although some of these linkages have been reported previously for plants (e.g., Hatzfeld and Stitt, 1990; Fernie et al., 2001a), and some are

probably trivial, these analyses also gave insight into regulation of metabolism within the tuber. The hyperbolic nature of the curve obtained when lysine was plotted against methionine fits models of feedforward and feedback regulation of the aspartate family biosynthetic pathway in *Arabidopsis*, as outlined by Bartlem et al. (2000). These authors suggest that under conditions of high flux, the methionine-threonine branch point is tightly regulated such that when methionine accumulates, its feedback inhibits expression of cystathione  $\gamma$ -synthase mRNA, but feedforward activates the competing branch point catalyzed by threonine synthase. If such control were operating in the potato tuber, it would follow that methionine would only accumulate to a threshold level, whereas lysine would continue to increase with increasing flux through the pathway. Thus, a hyperbolic relationship between these metabolites indicates that the biosynthetic pathway of the potato tuber aspartate family is regulated in

a manner analogous to that of *Arabidopsis*. That these plots can indicate metabolic regulation at a certain locus is very exciting, because the screening of, for example, mutant populations for individuals lines that do not fit these relationships have the potential to allow identification of component genes of regulatory factors at these loci. A further function of these plots is that the high degree of correlation between the unknown compounds and those for which the chemical nature is established may aid in the identification of these compounds and hence to improvements in the efficacy of our protocol.

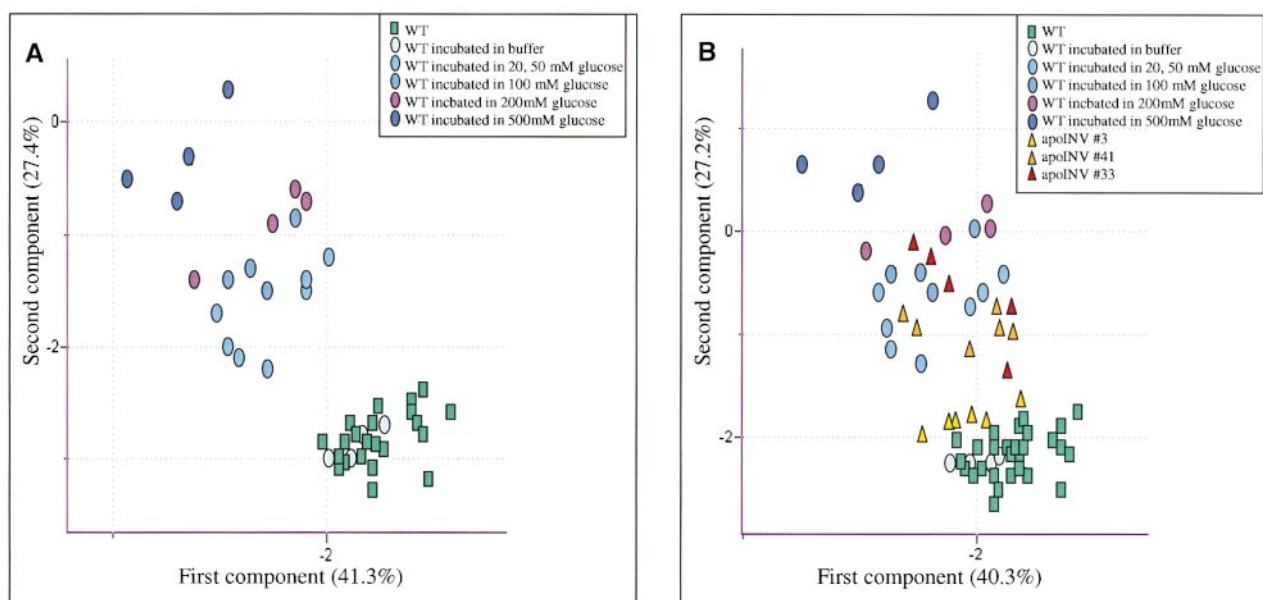
As a first example of the power of metabolic profiling, we demonstrated its use in identifying phenocopies of certain genetic modifications—an approach that obviously will be

useful in functional genomics. For this purpose, we incubated tuber discs in various concentrations of glucose and determined their subsequent metabolic profile. This manipulation led to metabolic profiles that formed distinct clusters from the transgenic lines we initially chose for our study (INV, GK3, and SP). However, when the metabolic profiles of the glucose-incubated samples were compared with transgenics we had profiled previously (Roessner et al., 2000), we were able to phenocopy one of them—potato plants expressing a yeast invertase within the tuber at an apoplastic location. A possible explanation for the phenocopying of these situations is the presence of a factor on the plasma membrane that has been implicated to sense the carbohydrate status of the cell wall space and mediate effects on



**Figure 9.** Assessment of the Metabolites Exerting the Largest Cluster Formation When Both Genetically and Environmentally Modified Systems Are Considered.

Shown is the contribution of individual metabolites to the PCA vector calculation by linear combination. The closer to the origin, the smaller the influence a given metabolite has on the linear combination. The most important metabolites with respect to the separation of the differently treated samples are labeled. The metabolite that appears in brackets, fructose, only has this degree of influence when glucose feeding profiles are considered.



**Figure 10.** PCA of the Metabolic Profiles of Genetically and Environmentally Modified Systems in Combination with the Metabolic Profiles of Previously Profiled Transgenic Plants.

**(A)** Expansion of the glucose-feeding cluster region presented in Figure 7.

**(B)** Relationship between apoplastic invertase-expressing tubers and glucose-fed wild-type tubers.

The distances between these populations were calculated as described in Methods by using the log-transformed, normalized data from both environmentally and genetically modified systems. PCA vectors span a 10-dimensional space to give best sample separation, with each point representing a linear combination of all the metabolites from an individual sample. Vectors 1 and 2 were chosen for best visualization of differences between genotypes and/or different environmental conditions and include 67.5% of the information derived from metabolic variances. WT, wild type.

cellular metabolism (Lalonde et al., 1999; Fernie et al., 2001b, 2000). The fact that certain environmental conditions can phenocopy genetic modifications, even when many parameters are considered, proves the general utility of this approach.

## Conclusion and Perspectives

The work presented in this article demonstrates that metabolic profiling coupled with bioinformatic tools represent an additional exciting approach to the analysis of complexity within plant systems. We believe that the data herein illustrate that our protocol allows the phenotyping of diverse plant systems and gives multiple insights into regulation and relationships of metabolite levels within plant cells. The exact number of chemical compounds present in the combined plant kingdoms is unknown; however, estimates range from 90,000 to 200,000 different molecules, with a single species such as *Arabidopsis* having a complexity in excess of 5000 compounds. However, a large proportion of this enormous diversity results from compounds of second-

ary metabolism. It is therefore obvious that the approach described here, in which we essentially limit ourselves to ~80 compounds, does not represent the end of this development. Unpublished data from our group suggest that applying different data extraction algorithms (using peak deconvolution software) to the original chromatograms increases the number of distinct compounds detected by a factor of three. A further extension of the metabolic-profiling approach is in the development of similar, automated technologies for the nonvolatile or highly fragile compounds; liquid chromatography coupled to MS represents one such approach. The biggest hurdle is probably determination of the exact chemical structure of the individual compounds seen. Here, a multitude of approaches such as MS, nuclear magnetic resonance, and other techniques will be useful. It is our belief that only the combination of many analytical techniques will allow a full description of the metabolome status of an organism and thus create a third level of multi-parallel approaches. When taken together with RNA and protein analyses, the metabolic complement will allow a full picture of the complexity of the biological entity under study.



## METHODS

### Plant Materials

*Solanum tuberosum* cv Desiree was obtained from Saatzucht Lange AG (Bad Schwartau, Germany). The generation and selection of the transgenic lines used here have been described previously by Sonnewald et al. (1997) and Trethewey et al. (1998, 2001). Plants were maintained in tissue culture with a 16-hr-light/8-hr-dark regime on Murashige and Skoog (1962) medium that contained 2% sucrose. In the greenhouse, plants from all transgenic lines and wild-type controls were grown in parallel under the same light regime with a minimum of 250  $\mu\text{mol photons m}^{-2} \text{sec}^{-1}$  at 22°C. In this article, the term “developing tubers” is used for tubers (>10 g fresh weight) harvested from healthy 10-week-old plants.

### Chemicals

All chemicals and pure standard substances were purchased from either Sigma-Aldrich Chemie GmbH (Deisenhofen, Germany) or Merck KGaA (Darmstadt, Germany).

### Confirmation of Preliminary Biochemical Characteristics of Transgenic Lines

Extraction and assaying of invertase and glucokinase were performed according to Trethewey et al. (1998), and those of sucrose phosphorylase were performed following the protocol of Trethewey et al. (2001). Carbohydrate levels were determined exactly as described in Morrell and ap Rees (1986), whereas phosphorylated intermediates were measured according to protocols described in Fernie et al. (2001a). Recoveries of metabolites in the trichloroacetic acid extracts have been documented previously (e.g., Trethewey et al., 1998; Veramendi et al., 1999; Fernie et al., 2001a).

### Extraction, Derivatization, and Analysis of Potato Tuber Metabolites Using Gas Chromatography–Mass Spectrometry

Potato tuber tissue (100 mg) was extracted in 1400  $\mu\text{L}$  of methanol, as described by Roessner et al. (2000); 50  $\mu\text{L}$  of internal standard (2 mg of ribitol in  $\text{mL}^{-1}$  water) was added for quantification. The mixture was extracted for 15 min at 70°C, mixed vigorously with 1 volume of water, centrifuged at 2200g, and subsequently reduced to dryness in vacuo. The residue was redissolved and derivatized for 90 min at 30°C (in 80  $\mu\text{L}$  of 20 mg  $\text{mL}^{-1}$  methoxyamine hydrochloride in pyridine) followed by a 30-min treatment at 37°C (with 80  $\mu\text{L}$  of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide). Forty microliters of a retention time standard mixture (3.7% [w/v] heptanoic acid, 3.7% [w/v] nonanoic acid, 3.7% [w/v] undecanoic acid, 3.7% [w/v] tridecanoic acid, 3.7% [w/v] pentadecanoic acid, 7.4% [w/v] nonadecanoic acid, 7.4% [w/v] tricosanoic acid, 22.2% [w/v] heptacosanoic acid, and 55.5% [w/v] hentriacontanoic acid dissolved in 10 mg  $\text{mL}^{-1}$  tetrahydrofuran) was added before trimethylsilylation. Sample volumes of 1  $\mu\text{L}$  were then injected with a split ratio of 25:1, using a hot needle technique.

The gas chromatography–mass spectrometry (GC-MS) system was composed of an AS 2000 autosampler, a GC 8000 gas chromatograph, and a Voyager quadrole mass spectrometer (Thermo-

Quest, Manchester, UK). The mass spectrometer was tuned according to the manufacturer's recommendations, using tris-(perfluorobutyl)-amine (CF43). GC was performed on a 30-m SPB-50 column with 0.25- $\mu\text{m}$  film thickness (Superlco, Bellfonte, CA). The injection temperature was set at 230°C, the interface at 250°C, and the ion source adjusted to 200°C. Helium was used as the carrier gas at a flow rate of 1  $\text{mL min}^{-1}$ . The analysis was performed under the following temperature program: 5 min of isothermal heating at 70°C, followed by a 5°C  $\text{min}^{-1}$  oven temperature ramp to 310°C, and a final 1 min of heating at 310°C. The system was then temperature equilibrated for 6 min at 70°C before injection of the next sample. Mass spectra were recorded at 2 scan  $\text{sec}^{-1}$  with a scanning range of 50 to 600  $\text{m/z}$ . Both chromatograms and mass spectra were evaluated using the MASSLAB program (ThermoQuest). A retention time and mass spectral library for automatic peak quantification of metabolite derivatives were implemented within the MASSLAB method format. Substances were identified by comparison with authentic standards, as described in Roessner et al. (2000). The recovery of small representative amounts of each metabolite through the extraction, derivatization, storage, and quantification procedures has been documented previously (Roessner et al., 2000). Data sets measured at different times are not directly comparable because of varying tuning parameters of the GC-MS machine over time; we therefore normalized the data by using the wild-type control of each measured batch as a reference. To include all the specific ions used for quantification of the metabolites (Roessner et al., 2000), we averaged all response numbers for the wild-type control and divided all data from a measured batch by the calculated factor.

### Glucose Incubation of Potato Tuber Slices

Glucose incubations were performed essentially as described by Geiger et al. (1998). Discs were cut directly from developing tubers from nonsenescent wild-type plants and washed three times in 10 mM Mes-KOH. They were then placed in 100-mL flasks (eight discs per flask) containing 5 mL of incubation medium (10 mM Mes-KOH, pH 6.5), supplemented with 0, 20, 50, 100, 200, or 500 mM glucose, and incubated with shaking (at 150 rpm) for 2 hr. An aliquot of the incubation media then was immediately frozen in liquid  $\text{N}_2$  for subsequent analysis. Samples were washed three times in 10 mM Mes-KOH, pH 6.5, before they were dried and frozen in liquid  $\text{N}_2$  for subsequent analysis. Analysis of the tuber extracts was performed as described above, except that the glucose level of the sample was quantified by calibration, as described previously (Fernie et al., 2001b; Roessner et al., 2000).

### Cluster Analysis

Hierarchical cluster analysis (HCA) and principle component analysis (PCA) were performed with the S-Plus system, as detailed by Venables and Ripley (1999). For an independent confirmation of the results obtained by this method, we also used the informatic program Pirouette 2.6 (Infometrix, Woodinville, WA). HCA allows the presentation of cluster results in a dendrogram, where the similarity of two samples can be determined from the value on the distance axis at which they join in a single cluster (the smaller the distance, the more similar the sample). All HCAs described in this article were transformed by log 10 to allow better comparison of large and small numbers. We used the Euclidean distance to calculate the matrix of all samples. The

complete linkage method was then used in the assignment of clusters. For PCA, the  $n$ -dimensional data set was transformed into a second  $n$ -dimensional data set in which what was designated as the most important information of the original data set was stored in the first few dimensions. These transformations allowed the reduction of the original data set to only the most important dimensions, hence allowing more distinct cluster formation. The results of these analyses were then presented as a two-dimensional graphical display of the data in which a single sample is represented by a point in three-dimensional space.

### Statistical Analysis

If two observations are described in the text as different, this means that their difference was determined to be statistically significant by the performance of  $t$  tests using the algorithm incorporated into Microsoft Excel 7.0. (Microsoft, Seattle, WA).

### ACKNOWLEDGMENTS

We thank Drs. Joachim Kopka and Richard Trethewey for helpful discussions. We are also grateful to Anna Lytovchenko for determination of carbohydrate contents. T.L. and D.B. acknowledge the support of the Ministerium für Wissenschaft, Forschung und Kultur des Landes Brandenburg.

Received October 20, 2000; accepted December 5, 2000.

### REFERENCES

- Aarts, M.G.M., Dirkse, W.G., Stiekema, W.J., and Pereira, A. (1993). Transposon tagging of a male sterility gene in *Arabidopsis*. *Nature* **366**, 715–717.
- Adams, M.A., Chen, Z.L., Landman, P., and Colmer, T.D. (1999). Simultaneous determination by capillary gas chromatography of organic acids, sugars, and sugar alcohols in plant tissue extracts as their trimethylsilyl derivatives. *Anal. Biochem.* **266**, 77–84.
- Aharoni, A., et al. (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* **12**, 647–661.
- Bartlem, D., Lambein, I., Okamoto, T., Itaya, A., Uda, Y., Kijama, F., Tamaki, E., Nambara, E., and Naito, S. (2000). Mutation in the threonine synthase gene results in an over-accumulation of soluble methionine in *Arabidopsis*. *Plant Physiol.* **123**, 101–110.
- Casazza, A.P., Basner, A., Hofgen, R., and Hesse, H. (2000). Expression of threonine synthase from *Solanum tuberosum* L. is not metabolically regulated by photosynthesis-related signals or by nitrogenous compounds. *Plant Sci.* **157**, 43–50.
- Celis, J.E., Kruhoffer, M., Gromova, I., Frederiksen, C., Ostergaard, M., Thykjaer, T., Gromov, P., Yu, J., Palsdottir, H., Magnusson, N., and Ornoft, T.F. (2000). Gene expression profiling: Monitoring transcription and translation products using DNA microarrays and proteomics. *FEBS Lett.* **480**, 2–16.
- Chang, W.W.P., Huang, L., Shen, M., Webster, C., Berlingame, A.L., and Roberts, J.K.M. (2000). Patterns of protein synthesis and tolerance of anoxia in root tips of maize seedlings acclimated to a low-oxygen environment, and identification of proteins by mass spectrometry. *Plant Physiol.* **122**, 295–317.
- Cho, R.J., et al. (1999). Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat. Genet.* **23**, 203–207.
- Duez, P., Kumps, A., and Mardens, Y. (1996). GC-MS profiling of urinary organic acids evaluated as a quantitative method. *Clin. Chem.* **42**, 1609–1615.
- Fernie, A.R., Riesmeier, J.W., Martiny, A., Ramalingam, S., Willmitzer, L., and Trethewey, R.N. (2000). Consequences of the expression of a bacterial glucokinase, both in combination with and independently of a yeast-derived invertase, in potato tubers. *Aust. J. Plant Physiol.* **27**, 827–833.
- Fernie, A.R., Roscher, A., Ratcliffe, R.G., and Kruger, N.J. (2001a). Fructose 2,6-bisphosphate activates pyrophosphate:fructose-6-phosphate 1-phosphotransferase and increases triose phosphate to hexose phosphate cycling in heterotrophic cells. *Planta*, in press.
- Fernie, A.R., Roessner, U., and Geigenberger, P. (2001b). The sucrose analog palatinose leads to a stimulation of sucrose degradation and starch synthesis when supplied to discs of growing potato tubers (*Solanum tuberosum*). *Plant Physiol.*, in press.
- Fiehn, O., Kopka, J., Dörmann, P., Altmann, T., Trethewey, R.N., and Willmitzer, L. (2000). Metabolite profiling for plant functional genomics. *Nat. Biotechnol.* **18**, 1157–1161.
- Geiger, M., Stitt, M., and Geigenberger, P. (1998). Metabolism in slices from growing potato tubers responds differently to addition of glucose and sucrose. *Planta* **206**, 234–244.
- Hatzfeld, W.D., and Stitt, M. (1990). A study of the rate of recycling of triose phosphates in heterotrophic *Chenopodium rubrum* cells, potato tubers and maize endosperm. *Planta* **180**, 198–204.
- Jang, J.-C., León, P., Zhou, L., and Sheen, J. (1997). Hexokinase as a sugar sensor in higher plants. *Plant Cell* **9**, 5–19.
- Katona, Z.F., Sass, P., and Nölnar-Perl, I. (1999). Simultaneous determination of sugars, sugar alcohols, acids and amino acids in apricots by gas chromatography–mass spectrometry. *Anal. Biochem.* **266**, 77–84.
- Kitao, S., and Sekine, H. (1994a).  $\alpha$ -D-Glucosyl transfer to phenolic compounds by sucrose phosphorylase from *Leuconostoc mesenteroides* and production of  $\alpha$ -arbutin. *Biosci. Biotechnol. Biochem.* **58**, 38–42.
- Kitao, S., and Sekine, H. (1994b). Syntheses of two kojic acid glucosides with sucrose phosphorylase from *Leuconostoc mesenteroides*. *Biosci. Biotechnol. Biochem.* **58**, 419–420.
- Lalonde, S., Boles, E., Hellmann, H., Barker, L., Patrick, J.W., Frommer, W.B., and Ward, J.M. (1999). The dual function of sugar carriers: Transport and sugar sensing. *Plant Cell* **11**, 707–726.
- Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. (1996). Expression monitoring by hybridisation to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680.
- Maimann, S., Wagner, C., Kreft, O., Zeh, M., Willmitzer, L., Höfgen, R., and Hesse, H. (2000). Transgenic potato plants reveal the

- indispensable role of cystathionine  $\beta$ -lyase in plant growth and development. *Plant J.* **23**, 747–756.
- Matsumoto, I., and Kuhara, T.** (1996). A new chemical diagnostic method for inborn errors of metabolism by mass spectrometry – Rapid, practical, and simultaneous urinary metabolite analysis. *Mass Spectrom. Rev.* **15**, 43–57.
- Morrell, S., and ap Rees, T.** (1986). Sugar metabolism in developing tubers of *Solanum tuberosum*. *Phytochemistry* **25**, 1579–1585.
- Muday, G.K., and Herrmann, K.M.** (1992). Wounding induces one of the 2-isoenzymes of 3-deoxy-D-arabino-heptulosonate 7 phosphate synthase in *Solanum tuberosum* L. *Plant Physiol.* **98**, 496–500.
- Murashige, T., and Skoog, F.** (1962). A revised medium for rapid growth and bioassays with tobacco tissue cultures. *Physiol. Plant.* **15**, 473–497.
- Ning, C., Kuhara, T., Inoue, Y., Zhang, C.H., Matsumoto, M., Shinka, T., Furumoto, T., Yokota, K., and Matsumoto, I.** (1996). Gas chromatography mass spectrometric metabolic profiling of patients with fatal infantile mitochondrial myopathy with De Toni-Tanconi-Debre syndrome. *Acta Paediatr. Jpn.* **38**, 661–666.
- Richmond, T., and Somerville, S.** (2000). Chasing the dream: Plant EST microarrays. *Curr. Opin. Plant Biol.* **3**, 108–116.
- Riedel, K.** (1999). Regulation of Starch Metabolism in Potato Tubers. Ph.D. Dissertation (Berlin: Freie Universität).
- Riedel, K., Mangelsdorf, C., Streber, W., Willmitzer, L., Hofgen, R., and Hesse, H.** (1999). Cloning and characterization of cystathionine gamma-synthase from *Solanum tuberosum* L. *Plant Biol.* **1**, 638–644.
- Roessner, U., Wagner, C., Kopka, J., Trethewey, R.N., and Willmitzer, L.** (2000). Simultaneous analysis of metabolites in potato tuber by gas chromatography–mass spectrometry. *Plant J.* **23**, 131–142.
- Ruan, Y., Gilmore, J., and Conner, T.** (1998). Towards *Arabidopsis* genome analysis: Monitoring expression profiles of 1400 genes using cDNA microarrays. *Plant J.* **15**, 821–833.
- Santoni, V., et al.** (1998). Use of proteome strategy for tagging proteins present at the plasma membrane. *Plant J.* **16**, 633–641.
- Schaefer, D.G., and Zryd, J.P.** (1997). Efficient gene targeting in the moss *Physcomitrella patens*. *Plant J.* **11**, 1195–1206.
- Shevchenko, A., Jensen, O.N., Podtelejnikov, A.V., Sagliocco, F., Wilm, M., Vorn, O., Mortensen, P., Shevchenko, A., Boucherie, H., and Mann, M.** (1996). Linking genome and proteome by mass spectrometry: Large scale identification of yeast proteins from two-dimensional gels. *Proc. Natl. Acad. Sci. USA* **93**, 14440–14445.
- Smeeckens, S.** (2000). Sugar-induced signal transduction in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **51**, 49–81.
- Smirnoff, N., and Wheeler, G.L.** (2000). Ascorbic acid in plants: Biosynthesis and function. *Crit. Rev. Plant Sci.* **19**, 267–290.
- Sonnenwald, U., Hajiraezaei, M.-R., Kossmann, J., Heyer, A., Trethewey, R.N., and Willmitzer, L.** (1997). Expression of a yeast invertase in the apoplast of potato tubers increases tuber size. *Nat. Biotechnol.* **15**, 794–797.
- Strepp, R., Scholz, S., Kruse, S., Speth, V., and Reski, R.** (1998). Plant nuclear gene knockout reveals a role in plastid division for the homolog of the bacterial cell division protein FtsZ, an ancestral tubulin. *Proc. Natl. Acad. Sci. USA* **95**, 4368–4376.
- Terryn, N., Rouze, P., and Van Montagu, M.** (1999). Plant genomics. *FEBS Lett.* **452**, 3–6.
- Trethewey, R.N., Geigenberger, P., Riedel, K., Hajiraezaei, M.R., Sonnenwald, U., Stitt, M., Riesmeier, J.W., and Willmitzer, L.** (1998). Combined expression of glucokinase and invertase in potato tubers leads to a dramatic reduction in starch accumulation and a stimulation of glycolysis. *Plant J.* **15**, 109–118.
- Trethewey, R.N., Krotzky, A.J., and Willmitzer, L.** (1999a). Metabolic profiling: A rosetta stone for genomics? *Curr. Opin. Plant Biol.* **2**, 83–85.
- Trethewey, R.N., Geigenberger, P., Sonnenwald, U., Hennig, A., Müller-Rober, B., and Willmitzer, L.** (1999b). Induction of the activity of glycolytic enzymes correlates with enhanced hydrolysis of sucrose in the cytosol of transgenic potato tubers. *Plant Cell Environ.* **22**, 71–79.
- Trethewey, R.N., Fernie, A.R., Bachmann, A., Fleischer-Notter, H., Geigenberger, P., and Willmitzer, L.** (2001). Glucose-independent induction of glycolysis in heterotrophic plant tissue. *Plant Cell Environ.*, in press.
- Venables, W.N., and Ripley, B.D.** (1999). *Modern Applied Statistics with S-PLUS*. (New York: Springer-Verlag).
- Veramendi, J., Roessner, U., Renz, A., Willmitzer, L., and Trethewey, R.N.** (1999). Antisense repression of hexokinase 1 leads to an overaccumulation of starch in leaves of transgenic potato plants but not significant changes in tuber metabolism. *Plant Physiol.* **121**, 1–11.
- Zhu, T., Peterson, D.J., Tagliani, L., St. Clair, G., Baszczynski, C.L., and Bowen, B.** (1999). Targeted manipulation of maize genes in vivo using chimeric RNA/DNA oligonucleotides. *Proc. Natl. Acad. Sci. USA* **96**, 8768–8773.

**Metabolic Profiling Allows Comprehensive Phenotyping of Genetically or Environmentally Modified Plant Systems**

Ute Roessner, Alexander Luedemann, Doreen Brust, Oliver Fiehn, Thomas Linke, Lothar Willmitzer and Alisdair R. Fernie  
*Plant Cell* 2001;13;11-29  
DOI 10.1105/tpc.13.1.11

This information is current as of April 24, 2013

<b>References</b>	This article cites 39 articles, 11 of which can be accessed free at: <a href="http://www.plantcell.org/content/13/1/11.full.html#ref-list-1">http://www.plantcell.org/content/13/1/11.full.html#ref-list-1</a>
<b>Permissions</b>	<a href="https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X">https://www.copyright.com/ccc/openurl.do?sid=pd_hw1532298X&amp;issn=1532298X&amp;WT.mc_id=pd_hw1532298X</a>
<b>eTOCs</b>	Sign up for eTOCs at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>CiteTrack Alerts</b>	Sign up for CiteTrack Alerts at: <a href="http://www.plantcell.org/cgi/alerts/ctmain">http://www.plantcell.org/cgi/alerts/ctmain</a>
<b>Subscription Information</b>	Subscription Information for <i>The Plant Cell</i> and <i>Plant Physiology</i> is available at: <a href="http://www.aspb.org/publications/subscriptions.cfm">http://www.aspb.org/publications/subscriptions.cfm</a>