Metabolome Analysis of Biosynthetic Mutants Reveals a Diversity of Metabolic Changes and Allows Identification of a Large Number of New Compounds in Arabidopsis[1][W][OA]

Christoph Böttcher, Edda von Roepenack-Lahaye, Jürgen Schmidt, Constanze Schmotz, Steffen Neumann, Dierk Scheel, and Stephan Clemens*

Leibniz Institute of Plant Biochemistry, Department of Stress and Developmental Biology, 06120 Halle/Saale, Germany (C.B., E.v.R.-L., C.S., S.N., D.S., S.C.); Leibniz Institute of Plant Biochemistry, Department of Bioorganic Chemistry, 06120 Halle/Saale, Germany (J.S.); and University of Bayreuth, Department of Plant Physiology, 95440 Bayreuth, Germany (S.C.)

Metabolomics is facing a major challenge: the lack of knowledge about metabolites present in a given biological system. Thus, large-scale discovery of metabolites is considered an essential step toward a better understanding of plant metabolism. We show here that the application of a metabolomics approach generating structural information for the analysis of Arabidopsis (*Arabidopsis thaliana*) mutants allows the efficient cataloging of metabolites. Fifty-six percent of the features that showed significant differences in abundance between seeds of wild-type, *transparent testa4*, and *transparent testa5* plants could be annotated. Seventy-five compounds were structurally characterized, 21 of which could be identified. About 40 compounds had not been known from Arabidopsis before. Also, the high-resolution analysis revealed an unanticipated expansion of metabolic conversions upstream of biosynthetic blocks. Deficiency in chalcone synthase results in the increased seed-specific biosynthesis of a range of phenolic choline esters. Similarly, a lack of chalcone isomerase activity leads to the accumulation of various naringenin chalcone derivatives. Furthermore, our data provide insight into the connection between *p*-coumaroyl-coenzyme A-dependent pathways. Lack of flavonoid biosynthesis results in elevated synthesis not only of *p*-coumarate-derived choline esters but also of sinapate-derived metabolites. However, sinapoylcholine is not the only accumulating end product. Instead, we observed specific and sophisticated changes in the complex pattern of sinapate derivatives.

The emergence of metabolomics has for good reasons attracted enormous attention in recent years (Fiehn, 2002; Sumner et al., 2003; Bino et al., 2004). Comprehensive detection of small molecules in a biological system has huge potential as a tool for functional genomics and systems biology. Also, metabolome analysis is expected to contribute significantly to economically important endeavors such as the discovery of bioactive compounds or improving food quality (Dixon et al., 2006; Ryan and Robards, 2006). Currently, however, there is no single analytical approach in sight that would cover the chemical diversity of metabolomes (Dunn and Ellis, 2005), and, perhaps even more problematic, most of the metabolites in any higher eukaryote are as yet unknown (Fernie et al., 2004). Plants in particular synthesize myriad compounds. Plasticity and diversity of metabolism are hallmarks of plant biology, yet we are far from understanding the complex networks of plant metabolism (Last et al., 2007). An essential first step toward a better comprehension of regulation and dynamics is the large-scale discovery of metabolites in Arabidopsis (*Arabidopsis thaliana*) and other model plants or crops and the cataloging of all of the metabolites that are synthesized in a system under investigation (Last et al., 2007). A large fraction of the thousands of unknowns in any given plant species are secondary metabolites. Fifteen percent to 20% of the genes in Arabidopsis, for instance, are predicted to encode enzymes of secondary metabolism. These are far more than necessary to produce the metabolites identified so far in Arabidopsis (Facchini et al., 2004; D'Auria and Gershenzon, 2005).

The strategy we are adopting to contribute to cataloging the Arabidopsis metabolome is the application of a powerful metabolite profiling approach that has the potential to generate structural information (von Roepenack-Lahaye et al., 2004; De Vos et al., 2007) for the analysis of mutants with defined blocks in metabolic pathways. The coupling of liquid chromatography to electrospray ionization quadrupole time-of-flight mass spectrometry (LC/ESI-QTOF-MS) allows the

[1] This work was supported by the German Plant Genome Initiative.

* Corresponding author; e-mail stephan.clemens@uni-bayreuth.de.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Stephan Clemens (stephan.clemens@uni-bayreuth.de).

[W] The online version of this article contains Web-only data.

[OA] Open Access articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.108.117754

detection of >1,000 mass signals in roots and leaves with S/N > 5 in a single analysis. In-source fragmentation, accurate mass, and targeted collision-induced dissociation (CID)-MS/MS analyses provide information for structure predictions (Clemens et al., 2006). Coupling to reversed-phase chromatography covers compounds in the medium polarity range, among them representatives of five of the six biosynthetic classes distinguished to date in Arabidopsis secondary metabolism: glucosinolates and their breakdown products (isothiocyanates, nitriles), *N*-containing compounds (indole derivatives), phenylpropanoids, flavonoids/polyketides, and fatty acid derivatives (D'Auria and Gershenzon, 2005). Terpenes cannot be detected here.

For a first systematic study, we chose the phenylpropanoid and flavonoid pathways because of their importance for plant biology. Phenylpropanoid synthesis is a ubiquitous pathway in terrestrial plants. Its evolution was most likely essential for the adaptation to land. Phenolic compounds, most of which are derived from the phenylpropanoid pathway, are major components of cell walls (lignin and suberin), accounting for about 40% of organic carbon in the biosphere (Buchanan et al., 2000). Understanding the regulation of phenylpropanoid synthesis and partitioning will be crucial for the efficient use of lignocellulosic material in biofuel and biomaterial production (Ragauskas et al., 2006). Flavonoids are derived from phenylpropanoids and are ubiquitous in higher plants as well. A large number of different flavonoids (at least 6,000) are known to date (Harborne and Williams, 2000) as well as a multitude of important biological functions (Winkel-Shirley, 2001). Anthocyanin flower pigments attract pollinators. Color is often influenced by complex formation with flavone copigments and metal ions (Shiono et al., 2005). Several flavonoids and isoflavonoids are known to function as signal molecules in plant-*Rhizobium* interactions. Plant protection is provided by UV-absorbing flavonoids, by phytoalexins with antifungal activity, or by antifeedants. The pharmacological activity of flavonoids in animals has elicited interest in their potential health benefits (Ross and Kasum, 2002). Prominent examples are certain isoflavonoids in legumes. More recently, flavonoids have been implicated in developmental processes. Flavonoids negatively regulate polar auxin transport (Brown et al., 2001; Buer and Muday, 2004; Besseau et al., 2007).

The second objective of this study besides the identification of metabolites was to unravel metabolic connections through the sensitive nontargeted detection of alterations in metabolic profiles between wild-type and mutant plants. We focused on Arabidopsis seeds and on the two well-characterized mutants *transparent testa4* (*tt4*) and *tt5* (Shirley et al., 1995). Until recently, there was very little known about seed metabolites in Arabidopsis despite the fact that there is pronounced commercial interest in the phenylpropanoid pathway in seeds of the Brassicaceae, most notably in rapeseed (*Brassica napus*), in which sinapate esters are antinutritive metabolites (Milkowski et al., 2004). *tt4* plants are deficient in chalcone synthase, the entry point into the flavonoid pathway. In Arabidopsis, there is only one chalcone synthase gene (Burbulis et al., 1996), just as for most of the other flavonoid biosynthesis enzymes (Winkel-Shirley, 2001). The following step, synthesis of the flavanone naringenin via cyclization of chalcones, is catalyzed by TT5 (=chalcone isomerase). The *tt4* mutant is completely devoid of flavonoids. Thus, it can be used to address the question of the spectrum of flavonoids found in Arabidopsis. *tt5-1* is also considered a null allele with no detectable chalcone isomerase activity or protein (Shirley et al., 1992; Cain et al., 1997).

Through profiling and structural elucidation via CID-MS/MS and ESI-Fourier transform ion cyclotron resonance (FTICR)-MS analyses, we found a surprising diversification of compounds upstream of metabolic blocks. Also, unanticipated connections between pathways can be inferred from our metabolome data. Finally, about 40 compounds were tentatively identified that had not been known from Arabidopsis. Thus, our study demonstrates the feasibility of the approach and the potential of LC/ESI-QTOF-MS to contribute significantly to cataloging the metabolomes of Arabidopsis and other species of interest.

RESULTS

Establishing LC/ESI-QTOF-MS Analysis of Seed Extracts

The coupling of LC to ESI-QTOF-MS was applied, to our knowledge, for the first time to Arabidopsis seed metabolome analysis in this study. Therefore, we initially validated the method and report the data in Supplemental Data S1, as recommended by the Chemical Analysis Working Group, which is part of the Metabolomics Standards Initiative (Sumner et al., 2007). We assessed the reliability and robustness of mass signal detection and quantification as well as the reproducibility of the extraction procedure by analyzing injection replicates (repetitive injections of a single seed extract; $n = 16$) and extraction replicates (repetitive extractions of a single seed pool; $n = 16$). Eleven features, a feature being defined as a unique *m/z* at a unique time point (Nordström et al., 2006), with known elemental compositions were chosen in order to be able to assess mass and isotope abundance accuracy. These 11 features plus three unknowns covered the entire chromatogram and several orders of magnitude in intensity. They were subjected to a first manual analysis, which is essential to establish parameters for the global data processing (Supplemental Data S1, table I). The average maximum retention time shift observed across both replicate series was 32 ± 13 s. Using an external two-point calibration, an average mass accuracy of 4.8 ± 1.9 mD was reached, which could be significantly improved to 1.1 ± 1.0 mD by applying a single-point recalibration to each chromato-

Table I. *Metabolites accumulating to higher levels in wild-type seeds (Ler) than in tt4 or tt5 seeds*

Alignment and comparison of L*er* and *tt4* (*tt5*) metabolite profiles revealed 79 (87) differential features (fold change ≥ 2, $P \leq 0.05$), of which 57 (60) could be putatively annotated. Quant., Quantifier. EC, Epicatechin; FC, feruloylcholine; G, guaiacyl; Hex, hexose; K, kaempferol; n.d., not determined; Q, quercetin; QMe, 3′-*O*-methylquercetin; SC, sinapoylcholine; SyC, syringoylcholine.

No.	Compound	Annotation Level[a]	Retention Time	Quant. Ion	ESI-FTICR-MS[b] (Fraction)	No. of Detected Differential Features		Experiment 1				Experiment 2			
								Fold Change		$-\log_{10}(P)$		Fold Change		$-\log_{10}(P)$	
						L*er*-*tt4*	L*er*-*tt5*	L*er*/*tt4*	L*er*/*tt5*	L*er*-*tt4*	L*er*-*tt5*	L*er*/*tt4*	L*er*/*tt5*	L*er*-*tt4*	L*er*-*tt5*
			min	*m/z*											
T1	Q-DeoxyHex	2[c,d]	31.2	303.05	III	8	7	n.d. in *tt4*	48.1	–	8.4	n.d. in *tt4*	60.3	–	9.0
T2	Q-(DeoxyHex)$_2$	2[c,d]	26.3	449.11	III	5	5	n.d. in *tt4*	6.4	–	6.0	n.d. in *tt4*	10.8	–	7.4
T3	Q-DeoxyHex-Hex	2[c,d]	24.5	611.16	III	4	4	n.d. in *tt4*	9.0	–	6.1	n.d. in *tt4*	14.4	–	4.5
T4	Q-Hex	2[d]	28.8	465.10	–	1	3	n.d. in *tt4*	n.d. in *tt5*	–	–	n.d. in *tt4*	n.d. in *tt5*	–	–
T5	K-DeoxyHex	2[c,d]	33.9	287.06	–	2	3	n.d. in *tt4*	n.d. in *tt5*	–	–	n.d. in *tt4*	n.d. in *tt5*	–	–
T6	K-(DeoxyHex)$_2$	2[c,d]	28.2	579.17	III	6	5	n.d. in *tt4*	19.7	–	7.5	n.d. in *tt4*	22.6	–	8.6
T7	K-DeoxyHex-Hex	2[d]	34.0	317.07	III	3	3	n.d. in *tt4*	7.0	–	4.9	n.d. in *tt4*	11.1	–	9.4
T9	QMe-(DeoxyHex)$_2$	2[d]	28.5	609.18	III	5	5	n.d. in *tt4*	3.9	–	7.1	n.d. in *tt4*	5.3	–	11.0
T10	QMe-DeoxyHex-Hex	2[d]	26.5	625.18	III	3	1	n.d. in *tt4*	2.4	–	4.1	n.d. in *tt4*	3.5	–	6.7
T11	EC-Hex	2[d]	18.2	291.09	–	3	3	n.d. in *tt4*	n.d. in *tt5*	–	–	n.d. in *tt4*	n.d. in *tt5*	–	–
12	EC	1[e]	23.2	291.09	–	2	3	n.d. in *tt4*	n.d. in *tt5*	–	–	n.d. in *tt4*	n.d. in *tt5*	–	–
T13	Bis(sinapoyl)spermidine	2[c]	27.1	558.28	II	3	2	n.d. in *tt4*	n.d. in *tt5*	–	–	n.d. in *tt4*	n.d. in *tt5*	–	–
T14	**T13**-Hex	2[f]	23.9	720.33	II	3	3	n.d. in *tt4*	n.d. in *tt5*	–	–	n.d. in *tt4*	n.d. in *tt5*	–	–
15	1-*O*-Sinapoyl Glc	1[g]	21.9	225.08	–	–	5	0.8	n.d. in *tt5*	0.6	–	1.6	n.d. in *tt5*	1.7	–
T16	SyC(4-*O*-β)G	2[f]	19.9	480.22	I	2	–	5.4	2.1	3.7	2.0	3.3	2.1	3.7	0.7
17	FC(4-*O*-β)G	1[h]	22.9	476.23	I	2	3	2.6	4.7	2.2	5.0	2.5	5.3	2.2	3.3
T18	SC(4-*O*-β)G	2[f]	25.7	506.24	I	2	2	2.6	3.2	3.2	5.6	2.8	5.1	3.1	4.6
19	FC(5-β)G	1[h]	29.9	458.22	I	1	1	2.4	2.1	3.6	4.2	2.2	2.5	2.6	6.8

[a]Annotation level according to Sumner et al. (2007). [b]Elemental composition confirmed by direct infusion ESI-FTICR-MS after fractionation. For numbering of fractions see "Materials and Methods." [c]Compound described in the literature. [d]Aglycone putatively annotated with a commercially available standard. [e]Compound identified with a commercially available standard. [f]Compound putatively annotated by interpretation and comparison of CID mass spectra of similar known compounds. [g]Compound identified with a standard isolated from plants. [h]Compound identified with a chemically synthesized standard.

gram (Supplemental Data S1, table III and figure I). Isotope patterns displayed an average relative error of 22% ± 13% (Supplemental Data S1, table IV). Mean relative SD (RSD) values of percentile-normalized feature intensities (Fundel et al., 2005) were 8.0% ± 3.5% for the injection replicates and 9.1% ± 4.0% for the extraction replicates (Supplemental Data S1, table V and figure III).

Subsequently, a global analysis using the XCMS package (Smith et al., 2006) was performed (Supplemental Data S1). Sixteen injection replicates yielded 434 features that were detectable with S/N > 3 in at least 75% of the replicates. After percentile normalization, feature intensities showed a mean RSD of 10.5% ± 5.0% (Supplemental Data S1, figure IV). From 16 extraction replicates, 467 features were extracted. Mean RSD of feature intensities was determined to be 11.4% ± 8.0%. Maximum retention time shifts were calculated for the complete feature lists. Seventy-nine percent of the features extracted from both replicate series were below 30 s, with a mean maximum shift of 21 ± 17 s (Supplemental Data S1, figure V).

In order to estimate the number of features exceeding the upper dynamic range limit, 2-fold dilution series were analyzed. Using the XCMS algorithm, response curves of prominent features were constructed. Out of 208 features consistently detected down to an 8-fold dilution in a Landsberg *erecta* (L*er*) seed extract, 147 features (71%) showed linear behavior (i.e. a decrease in intensity across all dilution steps, with a mean correlation coefficient of 0.948). Thirty-five features (17%) did not respond exclusively in the first dilution step. Among the remaining features, the dominating seed metabolite sinapoylcholine was identified as being outside its dynamic range. For quantification of this compound, extracts had to be 10-fold diluted.

Next, we determined the range of compounds covered by our analysis. Due to the reconstitution step of crude seed extracts in methanol:water (1:9, v/v) and the design of the chromatographic gradient, the analysis is restricted to polar and semipolar compounds. These include primary metabolites like amino acids, certain saccharides and nucleosides, as well as secondary metabolites such as aliphatic glucosinolates, phenolic esters, and flavonol glycosides (for identification or putative annotation of these compounds, see Supplemental Data S4). Recovery rates determined for eight model compounds ranging in polarity from phenylGly (t_r = 6.9 min, XlogP = −1.8) to biochanin A (t_r = 45.2 min, XlogP = 2.6) indicated the applicability of the extraction protocol for compounds in this polarity range (Supplemental Data S1, table VI). The method did not allow the detection of fatty acids, glycerolipids, sterols, or compounds of comparable low polarity.

The total number of detectable compounds cannot be accurately determined. It is certainly smaller than the number of features robustly measured in Arabidopsis seed extracts, because adduct formation, in-source fragmentation, and isotope peaks result in some compounds giving rise to more than one feature. Chromatogram correlation-based deconvolution of

434 features extracted in the injection replicates using a recently released extension of the XCMS package (Tautenhahn et al., 2007) revealed about 180 components each represented by an average of 2.4 features (Supplemental Data S1, figure VI).

ESI-MS can be prone to matrix effects (i.e. suppression or enhancement of analyte signal intensity caused by coeluting compounds; Taylor, 2005). Even though the relative quantification that is typical for metabolome studies is much less compromised by matrix effects than previously thought, it is advisable to assess possible matrix effects for every new sample type (Böttcher et al., 2007). In order to study absolute and relative matrix effects for wild-type and *tt4*/*tt5* mutant seed extracts, postextraction spiking and postcolumn infusion experiments were performed. A test set of seven reference compounds (Supplemental Data S1, table VII) was added at different concentrations to pure solvent as well as wild-type and mutant seed extracts. After the establishment of response curves, absolute matrix effects were determined from alterations in sensitivity caused by the matrix in comparison with standard solutions (Supplemental Data S1, table VIII). For five of the seven compounds, we observed modest ion suppression/enhancement below 30%. Two compounds, phenyl-Gly and rutin (the former elutes near the void time), were subject to stronger matrix effects (Supplemental Data S1, figure VII). Nevertheless, comparison of absolute matrix effects gave no indications for significant relative matrix effects between wild-type and mutant seed extracts for five of six chromatographically separated test compounds. Only one of the retained compounds (rutin) showed an approximately 1.6-fold stronger suppression in the wild-type seed extract relative to mutant seed extracts. To dynamically assess matrix effects over the whole chromatographic gradient, two test compounds differing in polarity, kinetin and biochanin A, were continuously delivered into the column eluate and monitored over the gradient after solvent and sample injection (Supplemental Data S1, figure VIII). Suppression or enhancement of signal strength was largely confined to retention time windows of the major components (e.g. sinapoylcholine and 8-methylthiooctylglucosinolate) and the solvent front. Again, comparison of postcolumn infusion chromatograms of different matrices did not reveal significant differences that would seriously compromise the relative quantification between different types of seed extracts.

Detection of Differences between Seeds of L*er* and *tt* Mutants

Two independently grown batches of L*er*, *tt4*, and *tt5* seeds were subjected to LC/ESI(+)-QTOF-MS analysis in four technical replications. Following data processing with the XCMS software, which has been demonstrated to reliably align LC/MS chromatograms and detect differences (Nordström et al., 2006), pairwise comparisons of wild-type and mutant data sets were performed. Features were identified as differential that were detected in at least three of four replicates and displayed a greater than 2-fold ($P < 0.05$) difference in both experiments. Of 618 features detected in the L*er* versus *tt4* comparison, 79 features were consistently more intense in L*er* and 145 features indicated higher abundance in *tt4*. The respective numbers for the L*er* versus *tt5* experiments were 527 features total, of which 87 were significantly stronger in L*er* and 78 in *tt5*.

In addition to the pairwise comparisons, the complete data set was analyzed by principal component analysis after appropriate scaling (Fig. 1). The first three principal components explained 70.0% of the total variance. Both PC1 (39.5%) and PC2 (24.5%) were associated with genotype-specific variation. Whereas PC1 clearly discriminated all of the three genotypes, PC2 was responsible for an additional separation of the wild type from both *tt* mutants. PC3 and higher principal components corresponded mainly to technical variation and led to separation between the respective genotypes. An experiment-specific separation was not observed, indicating that variability between seed pools originating from independent experiments was in the same range as technical variability. As expected by the large fraction of differential features identified in pairwise comparisons between mutants and the wild type, examination of the loading plots of PC1 and PC2 revealed a high number of features contributing to genotype-specific differences.

Annotation of Differential Features and Identification of Metabolites

The assignment of differential features to individual compound mass spectra, followed by identification of (pseudo)molecular ion(s), cluster ions, and in-source fragment ions as well as the determination of their charge states, are the first steps in structure elucidation. Therefore, differential features detected in narrow retention time windows and exhibiting high chromatogram correlation were grouped. In combination with the measured mass spectra, putative compound mass spectra were reconstructed and annotated. In the second step, targeted CID-MS/MS experiments of (pseudo)molecular ions or in-source fragment ions (pseudo-MS^3) were performed. Based on accurate masses, putative elemental compositions were calculated for the (pseudo)molecular ion as well as for fragment ions and neutral losses observed in the CID mass spectra, applying reasonable restrictions on elemental compositions, the number of double bond equivalents, and electron parity. Although isotope abundance information can be used as an orthogonal filter to efficiently reduce the number of potential elemental compositions (Kind and Fiehn, 2006), isotope abundances obtained with the hardware and hardware settings used in this study were of low accuracy and therefore hardly useable for that purpose (Supplemental Data S1, table IV). However, after reconstruction of the fragmentation scheme using infor-

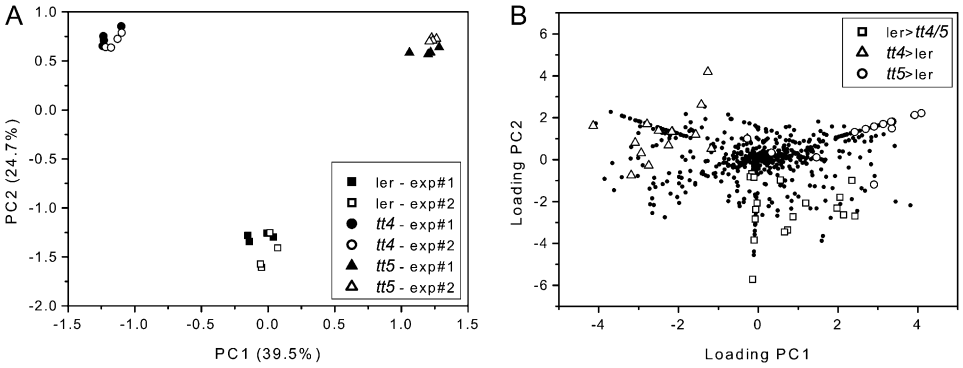

Figure 1. Multivariate analysis of LC/MS profiles of methanolic extracts prepared from Arabidopsis wild-type, *tt4*, and *tt5* mutant seeds. Seed pools of 20 plants each were created in two independent experiments and analyzed in four technical replications. Peak picking and alignment resulted in 660 features, which were log fold-change transformed and subjected to principal component analysis. Projection of the sample vectors onto the first two principal components (A) explaining 64% of total variance reveals genotype-specific sample clustering. The corresponding loading plot (B) indicates a high number of features with genotype-associated alteration. Selected features identified in pairwise comparisons of the wild type versus *tt4/tt5* (quantifier ions listed in Tables I–III) are highlighted.

mation from pseudo-MS^3 experiments, efficient calculation of the precursor molecular formula could be achieved by sequentially adding unique elemental compositions of small neutral losses starting from unambiguously determined product ions (Konishi et al., 2007). In order to validate elemental compositions from QTOF-MS analyses, seed extracts of each genotype were consecutively fractionated on weak cation and anion exchangers and analyzed by direct infusion ESI-FTICR-MS. Identification, putative annotation, or classification of metabolites differing significantly in abundance between L*er* and *tt4* or *tt5* were then pursued through a combination of database searches (Chemical Abstracts Service, PubChem, KNApSAcK [Shinbo et al., 2006]) with putative elemental compositions, interpretation of fragmentation patterns (Fig. 2), and the analysis of commercially available standards. Additionally, basic chemical transformations were used to obtain standard compounds for authentication purposes [e.g. a set of phenolic choline esters was synthesized by alkylation of commercially available phenolic acids with (2-bromoethyl)triethylammonium bromide; Clausen et al., 1982].

Following annotation, the relative quantification of annotated differential metabolites was further validated for linearity through analyses of dilution series of seed extracts and the establishment of response curves for corresponding quantifier ions (Supplemental Data S2).

In agreement with the nomenclature proposed by the Chemical Analysis Working Group (Sumner et al., 2007), we distinguish level 1 of annotation (identified compounds) from level 2 of annotation (putatively annotated compounds) by labeling the latter with a "T." For the L*er* > *tt4* class, we were able to characterize 57 features representing 18 different putative metabolites (Table I; Supplemental Data S1, table III). For 22 features, no identity could be assigned. Twelve of the putatively annotated metabolites were assigned to the flavonoid biosynthetic pathway: the flavan-3-ol epicatechin (**12**) and a derived hexoside (**T11**) as well as various glycoconjugates of the flavonols kaempferol (**T5**–**T7**), quercetin (**T1**–**T4**), and 3′-*O*-methyl quercetin (**T8**–**T10**; Fig. 3). The respective aglycones were not detected. The glycoside composition of compounds **T1** to **T10** was derived from characteristic neutral losses of 146.06 and 162.05 D from the protonated pseudomolecular ions, which correspond to deoxyhexose and hexose moieties, respectively. Those losses were already inducible by in-source CID. Putative annotation of the aglycones was achieved by comparison of pseudo-MS^3 spectra of the deglycosylated product ions with MS^2 spectra obtained from $[M+H]^+$ of authentic standards (Kachlicki et al., 2008). Assuming equal response factors, quercetin conjugates are the dominant flavonols in Arabidopsis seeds. All of the annotated flavonoids were below the detection limit in *tt4* seed extracts. The same applies to two novel hydroxycinnamoyl spermidine conjugates that were annotated as *N*,*N*′-bis(sinapoyl)spermidine (**T13**) and a corresponding *O*-hexoside *N*-sinapoyl-*N*′-(4-hexosyloxy-3,5-dimethoxycinnamoyl)spermidine (**T14**). The reported CID mass spectrum of *N*,*N*′-bis(sinapoyl)spermidine identified in pollen of *Hippeastrum* × *hortorum* was in good agreement with that obtained from compound **T13** (Youhnovski et al., 1998). Unfortunately, the acylation pattern of the spermidine backbone cannot be deduced from CID mass spectra due to intramolecular transamidation reactions occurring under mass spectral conditions (Bigler et al., 1996). Quantitative differences were found for three hydroxycinnamoylcholines oxidatively coupled to coniferyl alcohol via 4-*O*-*β* or 5-*β* linkages [feruloylcholine(4-*O*-*β*)guaiacyl (**17**), sinapoylcholine(4-*O*-*β*)guaiacyl (**T18**), and feruloylcholine(5-*β*)guaiacyl (**19**)] and syringoylcholine(4-*O*-*β*)guaiacyl (**T16**). They were significantly reduced in *tt4* seeds by factors of 2 to 3 yet were still detectable.

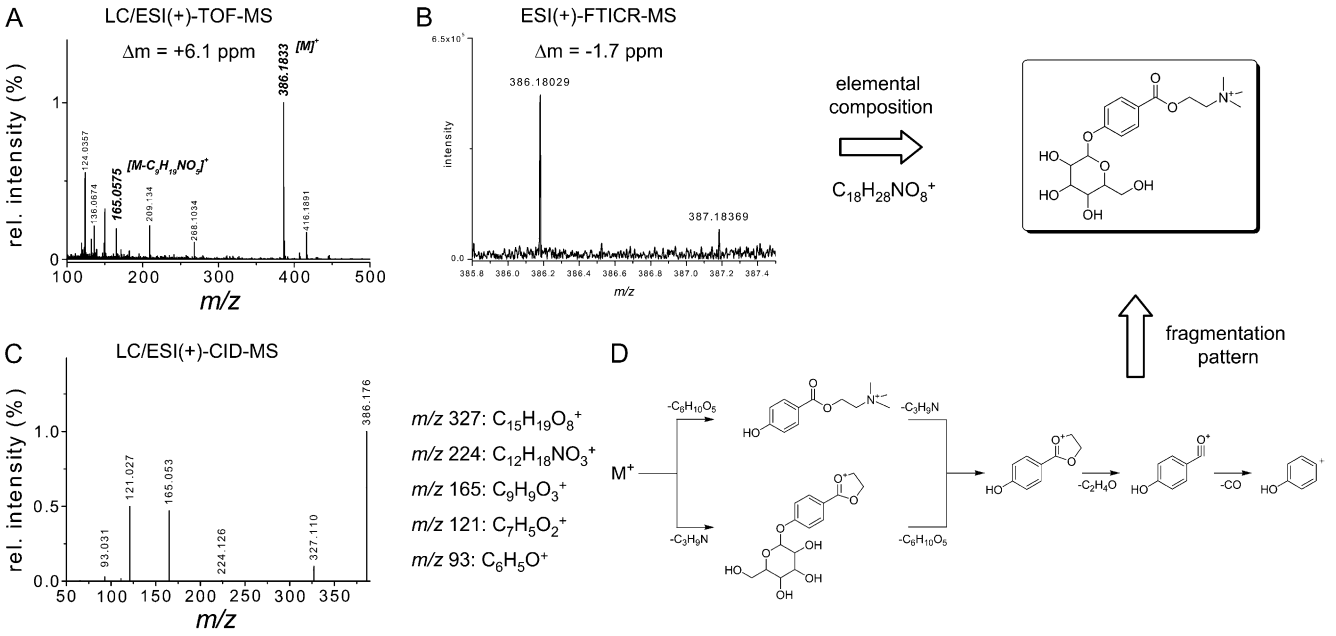

Figure 2. Structure elucidation of a compound overaccumulating in *tt4* seeds. A, Differential analysis of wild-type and *tt4* LC/ESI(+)-QTOF-MS profiles revealed altered features at *m/z* 386.183, 387.188, 388.189, and 165.058 coeluting at 7.0 min. These features belong to a single compound mass spectrum due to their chromatogram correlation and are annotated as molecular ion (*m/z* 386), isotope (*m/z* 387 and 388), and in-source fragment (*m/z* 165). Calculation of elemental composition (constraints: ^{1}H, ^{12}C, ^{16}O without limitation, $^{14}N \leq 5$, $^{32}S \leq 2$, even electron state, DBE ≥ -2) yields 10 potential sum formulas for the molecular ion within 15 ppm error. B, The molecular ion is detected by direct infusion ESI(+)-FTICR-MS in fraction I of a *tt4* seed extract at *m/z* 386.18029. Determination of elemental composition applying the same constraints results in a single sum formula within 2 ppm error. C, The molecular ion is collisionally fragmented at different collision energies (15 and 25 eV) by QTOF-MS. Elemental compositions were calculated for each fragment ion and each neutral loss. D, A hypothetical fragmentation scheme is constructed, and substructure information from characteristic fragment ions (*m/z* 121.017, 4-hydroxybenzoyl) and neutral losses (*m/z* 59.07 D, C_3H_9N; *m/z* 162.05 D, $C_6H_{10}O_5$) was combined to a biologically reasonable molecular structure. To further validate the structural assignment, the aglycone of the compound was chemically synthesized. The CID mass spectrum of synthetic 4-hydroxybenzoylcholine shows fragment ions at *m/z* 165, 121, and 93, as observed in the CID mass spectra of the new compound.

Identification of compounds **17** and **19** was achieved by comparison of synthetic standards prepared from ferulic acid(4-*O*-β)guaiacyl and ferulic acid(5-β)guaiacyl, respectively. Compounds **T16** and **T18** were putatively annotated based on their fragmentation patterns, which are similar to that observed for compound **17**.

The list of compounds being more abundant in L*er* relative to *tt5* confirmed exactly the flavonol glycoside spectrum of Arabidopsis seeds as determined through the L*er* versus *tt4* comparison (Table I; Supplemental Data S1, table III). In this case, however, flavonol glycosides were not completely absent in mutant seeds but were significantly reduced even though *tt5-1* is assumed to be a null mutant. The two hydroxycinnamoyl spermidine conjugates **T13** and **T14** identified in L*er* were undetectable in *tt5* as well. Similarly, the lower abundance of three of four phenolic choline esters cross-coupled to coniferyl alcohol (**17**, **T18**, and **19**) was confirmed. In this case, a reduction by factors of 2 to 5 was found. A *tt5*-specific difference was the reduced amount of 1-*O*-sinapoyl-Glc (**15**). Again, about 70% of the features could be characterized. Sixty features represented 18 different putative compounds, and 27 features remained unidentified.

A total of 145 features were significantly stronger in *tt4* seed extracts than in L*er* seed extracts (Table II; Supplemental Data S1, table III). We were able to characterize 59 of them, originating from 19 putative compounds. The most pronounced metabolic change apparent from the list of structurally elucidated compounds is a strong increase in the abundance of phenolic choline esters in *tt4*. *p*-Coumarate-derived choline esters, namely 4-hydroxybenzoylcholine (**20**), 4-hexosyloxybenzoylcholine (**T21**), a putative pair of *E*/*Z* isomers of 4-hexosyloxycinnamoylcholine (**T22/T23**), and 3-(4-hexosyloxyphenyl)propanoylcholine (**T24**), as well as five structurally uncharacterized phenolic choline esters (**T25**–**T29**) were either not detectable in L*er* seeds or showed increases in abundance in *tt4* by factors between 4 and 29. Compound **20** was unequivocally identified by comparison with a synthesized standard. Syntheses and mass spectrometric analyses of the aglycones of compounds **T21** to **T24** allowed putative annotation of these glycosylated phenolic choline esters. Also apparent from the metabolite spectrum in *tt4* seeds are putative dimerization products of sinapoylcholine, the major phenolic choline ester in Arabidopsis seeds. Two isomers of

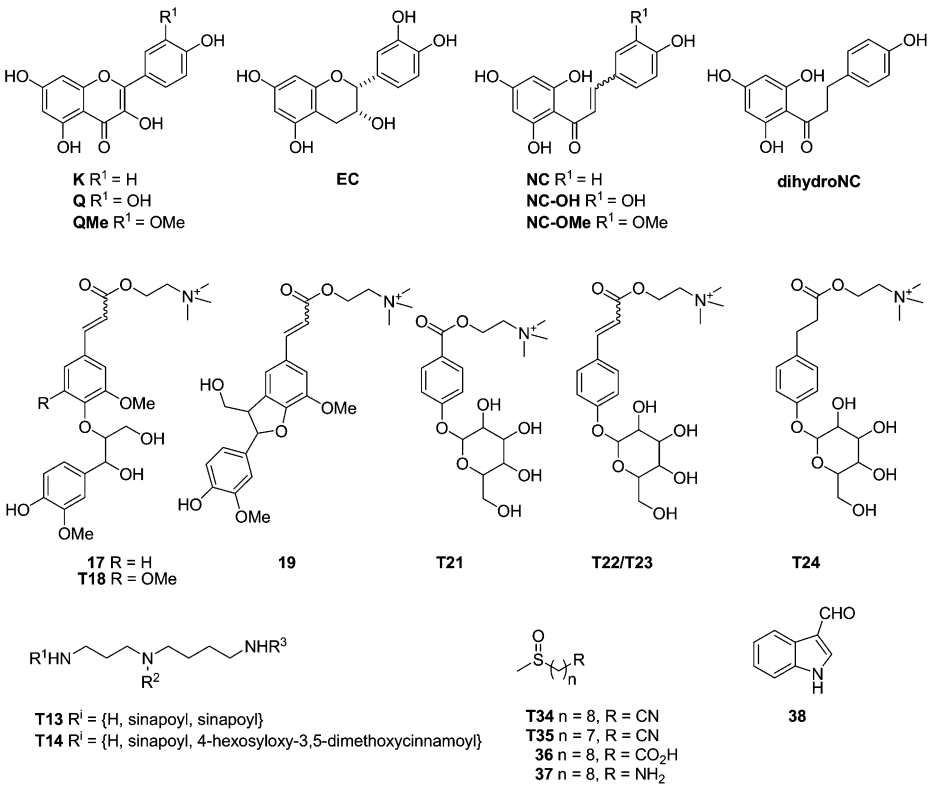

Figure 3. Putative structures of differentially accumulating compounds in wild-type and *tt4*/*tt5* mutant seeds. For the multitude of flavonoid glycosides only, the structure of the aglycone is depicted. Note that in numerous cases, neither the substitution pattern nor the stereochemistry could be exactly determined by mass spectrometry. For the complete mass spectrometric characterization and additional remarks on identification, see Supplemental Data S3.

sinapoylcholine dimers (**T30/T31**) and two isomers of sinapoylcholine dehydro dimers (**T32**/**T33**) were 10- to 25-fold more abundant in *tt4*, while the amount of sinapoylcholine was not significantly different between *tt4* and L*er* seeds. Bischoline esters **T30**/**T31** and **T32**/**T33** were detected as doubly charged ions at *m/z* 310.17 and *m/z* 309.16, respectively, and CID-MS/MS analyses revealed complex product ion spectra. However, characteristic fragment ions and neutral losses also observed in CID mass spectra of sinapoylcholine (**52**) were identified. Other identified changes included the accumulation of aliphatic glucosinolate breakdown products [9-(methylsulfinyl)nonanenitrile (**T34**), 8-(methylsulfinyl)octanenitrile (**T35**), 9-(methylsulfinyl)nonanoic acid (**36**), and 8-(methylsulfinyl)octane-1-amine (**37**)] and indole-3-carbaldehyde (**38**). Compound **36** was authenticated by comparison with the product obtained by acidic hydrolysis of 8-methylsulfinyloctylglucosinolate isolated from wild-type seeds (Olsen and Sorensen, 1980). Analogously, compound **37** was prepared by acidic hydrolysis of commercially available 8-methylsulfinyloctylisothiocyanate.

In the absence of chalcone isomerase activity, one would expect the accumulation of naringenin chalcone. Tentative identification of 14 metabolites being more abundant in *tt5* seeds, however, revealed that naringenin chalcone is further metabolized (Table III; Supplemental Data S3). Glycosylation results in two isomers of a naringenin chalcone hexoside (**T46**/**T47**) and a naringenin chalcone dihexoside (**T45**). Dihydronaringenin chalcone (phloretin) was identified as aglycone in two monohexosides (**T46**/**T47**) and one dihexoside (**T42**). Furthermore, B ring modifications analogous to flavonoid biosynthesis occur: 3′-hydroxylation and 3′-*O*-methylation and subsequent glycosylation result in the respective modified naringenin chalcone hexosides (**T48** and **T49**/**T50**). For putative annotation of compounds **T42** to **T50**, the same strategy as described for flavonol glycosides was applied. As observed for the flavonols in wild-type seed, aglycones were not detected. Another observation for *tt5* seeds was the 3-fold increase in choline (**39**) levels, which might correspond with the decrease in 1-*O*-sinapoyl-Glc content (**15**). The dramatic increase in phenolic choline esters found for *tt4* was not detected in *tt5*. Only 4-hydroxybenzoylcholine (**20**), feruloylcholine (**41**), and a sinapoylcholine dimer (**T40**) were found to be elevated.

In addition to annotating differential signals, we used mass spectral characteristics to identify structurally similar metabolites that were not found to show changes in abundance between the wild type and the *tt* mutants. Characteristic neutral losses and fragment ions observed upon CID as well as syntheses of additional compounds allowed the identification of three phenolic choline esters (**52**, **60**, and **61**) and putative annotation of eight additional representatives of this compound class (**T51** and **T53**–**T59**). Spectroscopic data and structures are shown in Supplemental Data S4. Furthermore, compounds known from primary and secondary metabolism of Arabidopsis could be identified, among them several Met-derived ali-

Table II. *Metabolites accumulating to higher levels in tt4 seeds than in wild-type (Ler) seeds*

Alignment and comparison of L*er* and *tt4* metabolite profiles revealed 145 differential features (fold change ≥ 2, $P \leq 0.05$), of which 59 could be putatively annotated. Quant., Quantifier. PCE, Phenolic choline ester; SC, sinapoylcholine.

No.	Compound	Annotation Level[a]	Retention Time	Quant. Ion	ESI-FTICR-MS[b] (Fraction)	No. of Detected Differential Features	Experiment 1		Experiment 2	
							Fold Change	$-\log_{10}(P)$	Fold Change	$-\log_{10}(P)$
			min	*m/z*						
20	4-Hydroxybenzoylcholine	1[c]	10.1	165.05	–	4	3.9	2.2	6.5	3.4
T21	4-Hexosyloxybenzoylcholine	2[d]	7.0	386.18	I	6	16.4	2.6	29.0	2.9
T22	4-Hexosyloxycinnamoylcholine isomer 1	2[d]	8.1	412.20	I	2	n.d. in L*er*	–	n.d. in L*er*	–
T23	4-Hexosyloxycinnamoylcholine isomer 2	2[d]	9.5	412.20	I	0	n.d. in L*er*	–	n.d. in L*er*	–
T24	3-(4-Hexosyloxyphenyl)propanoylcholine	2[d]	9.0	414.21	I	1	n.d. in L*er*	–	n.d. in L*er*	–
T25	Unknown PCE 1	3	18.1	504.18	–	2	13.7	1.9	28.2	2.2
T26	Unknown PCE 2	3	19.7	534.23	I	1	n.d. in L*er*	–	n.d. in L*er*	–
T27	Unknown PCE 3	3	24.3	370.19	I	2	44.6	4.6	43.0	3.1
T28	Unknown PCE 4	3	26.8	548.25	I	1	n.d. in L*er*	–	n.d. in L*er*	–
T29	Unknown PCE 5	3	37.8	516.22	I	2	5.6	4.2	7.1	5.4
T30	SC dimer isomer 1	3	10.5	310.17	–	3	22.8	3.0	51.3	3.4
T31	SC dimer isomer 2	3	27.5	310.17	–	9	21.1	5.7	34.9	4.8
T32	SC dehydro dimer isomer 1	3	21.0	279.62	–	4	9.8	4.1	24.5	3.3
T33	SC dehydro dimer isomer 2	3	25.9	309.16	–	7	15.5	5.1	33.4	4.5
T34	9-(Methylsulfinyl)nonanenitrile	2[e]	32.5	202.13	III	5	7.5	8.9	7.7	4.7
T35	8-(Methylsulfinyl)octanenitrile	2[f]	26.2	188.11	–	1	6.3	4.6	6.1	3.4
T36	9-(Methylsulfinyl)nonanoic acid	1[c]	29.5	203.11	–	5	11.8	5.9	9.0	5.5
T37	8-(Methylsulfinyl)octane-1-amine	1[c]	8.9	192.14	–	1	2.3	1.5	10.1	3.3
38	Indole-3-carbaldehyde	1[g]	31.1	146.06	–	2	7.6	5.5	4.7	4.5

[a]Annotation level according to Sumner et al. (2007). [b]Elemental composition confirmed by direct infusion ESI-FTICR-MS after fractionation. For numbering of fractions, see "Materials and Methods." [c]Compound identified with a chemically synthesized standard. [d]Aglycone putatively annotated with a chemically synthesized standard. [e]Compound described in the literature. [f]Compound putatively annotated by interpretation and comparison of CID mass spectra of similar known compounds. [g]Compound identified with a commercially available standard.

phatic glucosinolates (**T62–T66**), 2-*O*-sinapoylmalate (**67**), and several amino acids (**70–75**; Supplemental Data S4).

DISCUSSION

LC/ESI-QTOF-MS-based metabolite profiling is on the verge of becoming an established functional genomics approach, applicable for the detection and identification of a large fraction of a plant's metabolome (i.e. the semipolar secondary metabolites; von Roepenack-Lahaye et al., 2004; De Vos et al., 2007). Still, it is important to assess at this stage of LC/ESI-QTOF-MS method development the quality and scope of analysis for any new biological matrix and to report data describing method validation (Sumner et al., 2007; Fiehn et al., 2008). Besides the assessment of technical reproducibility and polarity range covered, method validation should consider technology-inherent limitations that might hamper the identification and reliable relative quantification of differential features in metabolite profiling studies. These include the restricted dynamic range of ESI (Tang and Kebarle, 1993; Zook and Bruins, 1997; Tang et al., 2004) and time-to-digital converter-based MS detection systems (Chernushevich et al., 2001) as well as the influence of absolute and relative matrix effects (Böttcher et al., 2007).

Through analysis of dilution series, we determined for aqueous methanolic extracts of Arabidopsis seeds that the majority of features were detected within their linear range. Manual inspection of 50 response curves (Supplemental Data S2) of annotated differential features revealed linear behavior for 46 features upon dilution. Sinapoylcholine (**52**), the major phenolic choline ester in Arabidopsis seeds, was identified to be detected outside its dynamic range and therefore quantified in 10-fold dilution. Postextraction spiking and postcolumn addition experiments indicated significant absolute matrix effects, mostly in areas of coelution with major components. However, both experimental strategies revealed only marginal relative matrix effects between different types of seed extracts.

The reproducibility that we determined for our analysis was very similar to that reported recently for LC/ESI-QTOF-MS profiling of human serum: the large majority of detected features showed RSD between 5% and 25% (Nordström et al., 2006). Observed retention time shifts could efficiently be corrected using the iterative XCMS algorithm. Altogether, about 660 features in wild-type and mutant lines were robustly detectable. For the characterized compounds, we were able to determine that, on average, about three features were detected per metabolite under our experimental conditions, confirming the estimate derived from the chromatogram correlation-based deconvolution (see above). With all due caution (adduct formation, in-source fragmentation, etc., are metabolite dependent and can vary greatly), this could be extrap-

Table III. *Metabolites accumulating to higher levels in tt5 seeds than in wild-type seeds*

Alignment and comparison of L*er* and *tt5* metabolite profiles revealed 78 differential features (fold change ≥ 2, $P \leq 0.05$), of which 42 could be putatively annotated. Quant., Quantifier. Hex, Hexose; NC, naringenin chalcone; SC, sinapoylcholine.

No.	Compound	Annotation Level[a]	Retention Time	Quant. Ion	ESI-FTICR-MS[b] (Fraction)	No. of Detected Differential Features	Experiment 1		Experiment 2	
							Fold Change	$-\log_{10}$ (P)	Fold Change	$-\log_{10}$ (P)
			min	*m/z*						
39	Choline	1[c]	5.8	104.11	–	2	2.4	4.4	5.3	4.8
20	4-Hydroxybenzoylcholine	1[d]	10.5	165.06	–	2	2.3	1.9	7.8	5.2
T40	SC dimer isomer 3	3	19.3	310.17	–	1	n.d. in L*er*	–	n.d. in L*er*	–
41	Feruloylcholine	1[d]	20.8	221.08	–	4	3.9	1.9	3.5	2.8
T42	Dihydro NC-$(Hex)_2$	2[e]	28.6	275.10	III	6	n.d. in L*er*	–	n.d. in L*er*	–
T43	Dihydro NC-Hex isomer 1	2[e]	33.8	275.10	III	5	n.d. in L*er*	–	n.d. in L*er*	–
T44	Dihydro NC-Hex isomer 2	2[e]	35.7	275.10	III	7	n.d. in L*er*	–	n.d. in L*er*	–
T45	NC-$(Hex)_2$	2[e]	28.7	273.08	III	4	n.d. in L*er*	–	n.d. in L*er*	–
T46	NC-Hex isomer 1	2[e]	32.1	273.08	III	3	n.d. in L*er*	–	n.d. in L*er*	–
T47	NC-Hex isomer 2	2[e]	35.2	273.08	III	3	n.d. in L*er*	–	n.d. in L*er*	–
T48	(NC-OH)-Hex	2[f]	32.0	289.08	–	2	n.d. in L*er*	–	n.d. in L*er*	–
T49	(NC-OMe)-Hex isomer 1	2[f]	32.0	465.14	III	0	n.d. in L*er*	–	n.d. in L*er*	–
T50	(NC-OMe)-Hex isomer 2	2[f]	32.9	465.14	III	2	n.d. in L*er*	–	n.d. in L*er*	–
T34	9-(Methylsulfinyl)nonanenitrile	2[g]	33.0	202.12	III	1	4.6	3.7	5.8	5.0

[a]Annotation level according to Sumner et al. (2007). [b]Elemental composition confirmed by direct infusion ESI-FTICR-MS after fractionation. For numbering of fractions, see "Materials and Methods." [c]Compound identified with a commercially available standard. [d]Compound identified with a chemically synthesized standard. [e]Aglycone putatively annotated with a commercially available standard. [f]Compound putatively annotated by interpretation and comparison of CID mass spectra of similar known compounds. [g]Compound described in the literature.

olated to approximately 220 compounds detected. The range of detected metabolites encompasses many physiologically and economically important semipolar compound classes, including phenolic acids and esters, phenylpropanoids, flavonoids, and glucosinolates (De Vos et al., 2007).

Identification and Classification of Compounds Distinguishing the Wild Type and Mutants

Having established efficient and reproducible seed metabolome analysis via LC/ESI-QTOF-MS, we initiated nontargeted profiling of known flavonoid biosynthesis mutants. The objective was to enhance our knowledge of the Arabidopsis seed metabolome with respect to both metabolites and metabolic pathways. A general strategy for nontargeted metabolome analysis is to compare sample sets, to generate lists of features being different, and finally to elucidate the structure of these differential features (Nordström et al., 2006). The latter is particularly important. As emphasized recently, meaningful metabolome analysis requires the identification of metabolites that are important to distinguish two plant lines, individuals, et cetera (Nielsen and Oliver, 2005). The mass accuracy of TOF-MS and CID represent two means of obtaining structural information. We applied these, and in some cases a combination of prefractionation and direct infusion ESI-FTICR-MS, to features differing significantly between L*er* and *tt4* or *tt5*. The overall success rate was 56%: 218 of 389 features could be characterized to represent 50 compounds. In addition, a number of nondifferential features were annotated based on spectral characteristics of identified compounds and literature data corresponding to another 25 compounds. In all, about 40 metabolites were identified that had not been found in Arabidopsis before. It should be noted, however, that in several cases the exact positions of functional groups and stereochemical issues cannot be determined by MS but require subsequent enrichment and NMR analysis. Thus, following the nomenclature proposed by the Chemical Analysis Working Group (Sumner et al., 2007), 43 of 75 metabolite identifications are of level 2 (putatively annotated compounds), while for 21 compounds level 1 (identified compounds) was reached owing to the availability of authentic standards (Supplemental Data S4). For 11 compounds, only the compound class could be putatively annotated (annotation level 3). For the remaining uncharacterized features, retention time and mass data are reported as suggested (Supplemental Data S3).

The use of mutants with defined metabolic blocks helped in two ways. Lack of chalcone synthase or chalcone isomerase activity resulted in a loss of metabolites downstream of these enzymes. Thus, a comparison with the wild type potentially reveals the whole spectrum of flavonoids normally present in seeds. Indeed, the spectra of flavonol glycosides and flavan-3-ols that we determined confirm recently reported data (Routaboul et al., 2006). Ten different flavonol glycosides with quercetin, kaempferol, and 3′-*O*-methyl quercetin as aglycones and three different conjugation patterns (DeoxyHex, DeoxyHex-DeoxyHex,

and DeoxyHex-Hex) were identified, as well as a quercetin hexoside, epicatechin, and an epicatechin hexoside. Quercetin-3-*O*-rhamnoside appears to be the most prominent flavonol glycoside. Flavonol aglycones were not detected. A comparison of this pattern with that reported before clearly validates our approach and underscores the potential to catalog metabolomes efficiently. The known seed flavonoid spectrum was apparently detected, except for procyanidin oligomers, which are outside our scan range of *m/z* 75 to 1,000 D.

Conversely, defined metabolic blocks can lead to higher signal strength for precursors and their derivatives. For example, most of the phenolic choline esters accumulating in *tt4* seeds were also detectable in the wild type, yet the signal strength in many cases did not allow obtaining structural data from wild-type extracts.

Finally, spectral and biological information on unidentified features that differ significantly between L*er* and *tt* mutants can be stored in databases such as MassBank (http://www.massbank.jp/) and compound/species databases such as KNApSAcK (Shinbo et al., 2006) and help in identifying the respective metabolites in future studies. The availability of such databases will greatly enhance identification success (Bino et al., 2004; Moco et al., 2006).

Phenolic Choline Ester in Seeds of Arabidopsis

In addition to sinapoylcholine (**52**), the well-known major phenolic choline ester in Arabidopsis seeds, structure elucidation of differential and nondifferential metabolites revealed the presence of about 20 minor representatives of this compound class in wild-type seeds. These include substituted hydroxycinnamoylcholines derived from the intermediary acids of the hydroxycinnamate pathway as well as corresponding substituted hydroxybenzoylcholines. Further modification by either 4-*O*-hexosylation or oxidative coupling to monolignols via 4-*O*-β or 5-β linkages gives rise to a combinatorial plethora of compounds. Sinapoylcholine itself accounts for approximately 80% to 90% of total phenolic choline ester content.

Enzymes possibly catalyzing the formation of phenolic choline esters include UDP-Glc:hydroxycinnamate glucosyltransferases and hydroxycinnamoyl-Glc:choline hydroxycinnamoyltransferases. 1-*O*-Sinapoyl-β-Glc is the immediate biosynthetic precursor of sinapoylcholine. In Arabidopsis, at least four enzymes (UGT84A1 to -A4, encoded by *At4g15480*, *At3g21560*, *At4g15490*, and *At4g15500*) are known to catalyze the formation of such β-acetal esters from UDP-Glc and hydroxycinnamic acids (Lim et al., 2001). Only UGT84A2 displays pronounced glucosyl acceptor specificity (Milkowski et al., 2000). Partial redundancy of those glucosyltransferases was demonstrated also through the characterization of the *brt1* mutant defective in UGT84A2 (Sinlapadech et al., 2007). According to the Genevestigator database (Zimmermann et al., 2004), at least three UGT84As (UGT84A2, -3, and -4) are expressed in mature siliques, with UGT84A2/BRT1 showing the highest expression level, followed by UGT84A3. SCT/SNG2 (*At5g09640*) catalyzes the acyl transfer reaction from Glc to choline (Shirley et al., 2001). Characterization of heterologously expressed SCT from Arabidopsis revealed a clear preference for choline as acyl acceptor but low substrate specificity toward a variety of hydroxycinnamoyl donors (Shirley and Chapple, 2003). Furthermore, a large number of other Ser carboxypeptidase-like proteins encoded in the Arabidopsis genome represent acyl transferases with presumably varying and overlapping acyl donor specificities (Fraser et al., 2007). Thus, enzyme diversity and substrate ranges are apparently broad enough to provide an explanation for the observed chemical diversity of phenolic choline esters in seeds. Future studies will aim at dissecting the contributions of these and other candidate enzymes.

A further interesting feature is the occurrence of several substituted hydroxybenzoylcholines. Since 4-hydroxybenzoylcholine (**20**) and its 4-*O*-hexoside (**T21**) were found to be strongly accumulating in *tt4* seeds, it can be assumed that *p*-coumarate or *p*-coumaroyl-CoA are precursors of these compounds. Thus, we hypothesize that biosynthesis of substituted hydroxybenzoylcholines occurs via the cleavage of two carbon atoms from the C3 side chain of precursor cinnamic acids, either via CoA-dependent β-oxidative or CoA-dependent/independent nonoxidative routes, and subsequent conjugation to choline (Wildermuth, 2006).

Detecting Metabolic Connections

Metabolism is far more diverse and interconnected than indicated by typical maps or charts (Nielsen and Oliver, 2005). Nontargeted metabolite profiling studies can help unravel complexity as well as unknown or unexpected connections. This is shown by our data in various ways. First, the accumulation of precursors yields an increase in the abundance of several derivatives (Fig. 4). This is evident for both *p*-coumaroyl-CoA and naringenin chalcone. LC/ESI-QTOF-MS allows a ''zooming in'' on the metabolism and documentation of this diversity (e.g. of phenolic choline esters). Interestingly, a number of phenolic choline esters showed no alterations in the mutants, indicating that the changes are still specific. Moreover, it should be noted that most of the accumulating choline esters are detectable also in the wild type and are thus not just the result of an unnatural block in the flavonoid pathway. Rather, our data reveal an additional branch point in seed phenylpropanoid metabolism. Overall, synthesis of *p*-coumarate-derived choline esters was less pronounced in *tt5*. Apparently, there is less accumulation of this key intermediate. Residual flavonol accumulation in *tt5* is probably due to spontaneous cyclization of naringenin chalcone, as suggested earlier (Peer et al., 2001).

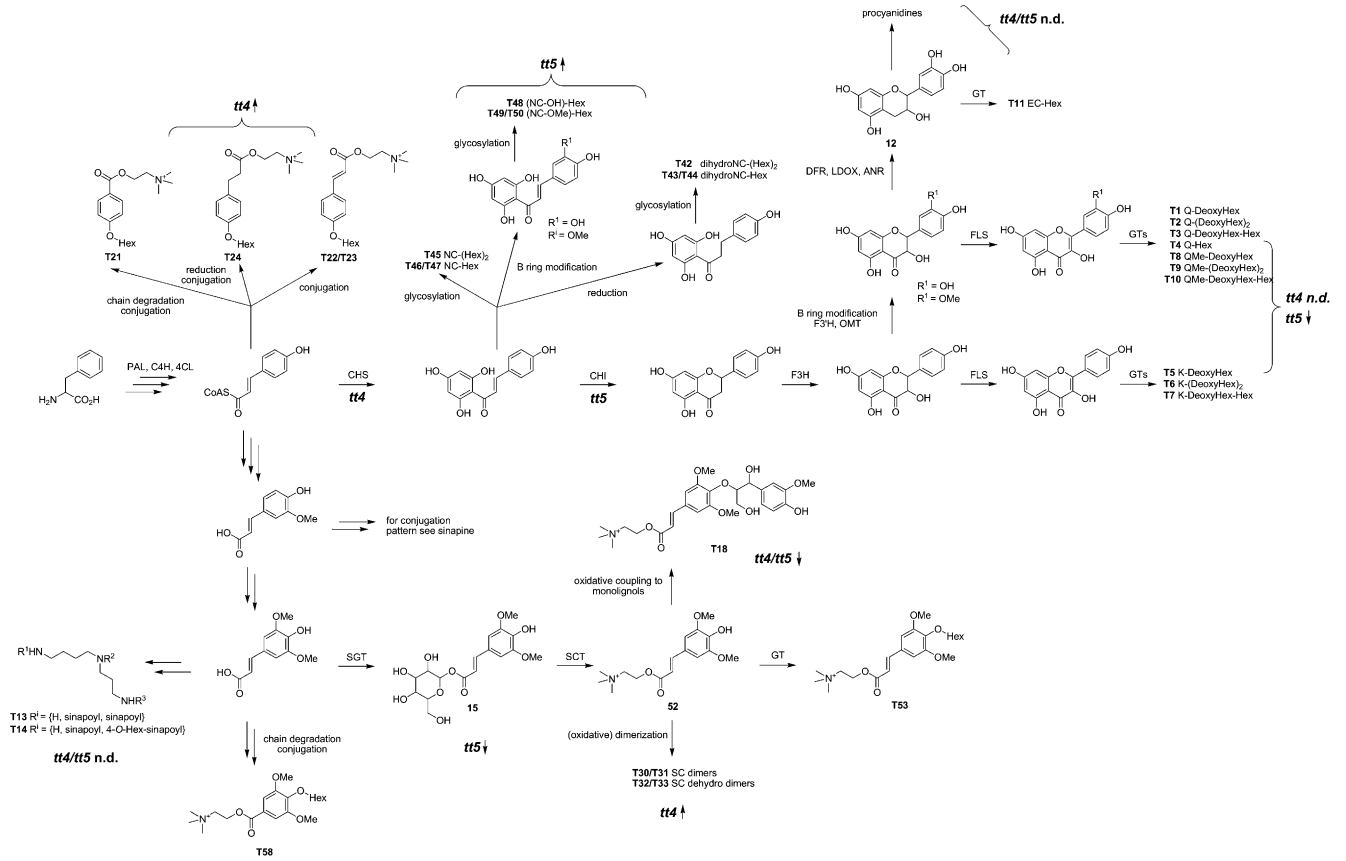

Figure 4. Simplified scheme of the flavonoid and phenylpropanoid biosynthetic pathways and tentative assignment of observed alterations in seeds of plants deficient in either chalcone synthase (CHS; *tt4*) or chalcone isomerase (CHI; *tt5*). These metabolic blocks lead to the accumulation of specific arrays of phenolic choline esters in the case of *tt4* and of naringenin chalcone derivatives in the case of *tt5*. In addition, a complex pattern of changes in hydroxycinnamate-derived metabolites is observed. Hydroxycinnamoylcholines oxidatively coupled to coniferyl alcohol showed a decrease and hydroxycinnamoyl spermidine conjugates were completely absent in both *tt* mutants. Sinapine levels remained constant, while dimerization products of sinapine accumulated strongly. ANR, Anthocyanidin reductase; C4H, cinnamate-4-hydroxylase; 4CL, 4-coumarate:CoA ligase; DFR, dihydroflavonol-4-reductase; F3H, flavonol-3-hydroxylase; F3′H, flavonol-3′-hydroxylase; FLS, flavonol synthase; GT, glycosyl transferase; LDOX, leucoanthocyanindin dioxygenase; OMT, *O*-methyl transferase; PAL, Phe ammonia lyase; SCT, sinapoyl-Glc:choline sinapoyltransferase; SGT, UDP-Glc:sinapic acid glucosyltransferase.

Second, besides elevated synthesis of *p*-coumarate-derived choline esters, there is redirection of *p*-coumaroyl-CoA into the sinapate pathway. Earlier work on *tt4* leaves suggested shunting of *p*-coumaroyl-CoA into sinapate biosynthesis (Li et al., 1993; Peer et al., 2001). We detected a more sophisticated array of specific changes in *tt4* and *tt5* seeds (Fig. 4). Some of the sinapate-derived metabolites showed a decrease in abundance (hydroxycinnamoylcholines oxidatively coupled to coniferyl alcohol) or were completely absent (hydroxycinnamoyl spermidine conjugates) in *tt4* and *tt5*. Others accumulated more strongly (e.g. dimerization products of sinapine, which increased by 10- to 25-fold). Sinapine levels remained stable, indicating that the dimerization might represent an overflow mechanism. Again, the resolution of our analysis revealed the nontrivial complexity, specificity, and diversity of changes.

Comparison with metabolome changes in leaves and roots of *tt4* mutant plants (von Roepenack-Lahaye et al., 2004) suggests that the accumulation of choline esters is a seed-specific phenomenon. The reverse effect with respect to partitioning between phenylpropanoids and flavonoids was described recently for hydroxycinnamoyl-CoA shikimate/quinate hydroxycinnamoyl transferase silencing (Besseau et al., 2007). A reduction in lignin precursor biosynthesis leads to stronger accumulation of flavonoids. Similarly, a block in one *p*-coumaroyl-CoA-dependent pathway, the 3′-hydroxylation of *p*-coumaroyl-CoA, leads to increased feeding into flavonol glycosides (Abdulrazzak et al., 2006).

An unexpected metabolic connection is indicated by the accumulation of glucosinolate breakdown products exclusively in *tt4* seeds. The glucosinolate pathway is assumed to be independent of phenylpropanoid and flavonoid pathways. Our data do not

immediately reveal the underlying mechanism. However, cross talk between these pathways has been observed previously. The *ref2* mutant of Arabidopsis was isolated based on its reduced sinapoylmalate content but was found to be defective in alkylglucosinolate biosynthesis due to a mutation in CYP83A1 (Hemm et al., 2003). Observations such as these emphasize the need for nontargeted metabolite profiling capable of annotating mass signals and its combination with genetic, physiological, and genomics data.

MATERIALS AND METHODS

Plant Material

Arabidopsis (*Arabidopsis thaliana*) *tt4* and *tt5* mutants in the L*er* background were isolated by Maarten Koornneef (Koornneef, 1990; Shirley et al., 1995). Seeds of *tt4-1* were kindly provided by Bernd Weisshaar (University of Bielefeld), and *tt5-1* seeds (N86) were obtained from the Nottingham Arabidopsis Stock Centre. Wild-type and mutant plants were grown in parallel on a soil:vermiculite mixture (3:2) in a fully climatized greenhouse at 22°C and 60% relative humidity under a photoperiod of 16 h of light/8 h of dark until final seed set (growth stage 9.70 according to Boyes et al. [2001]). Seed pools of 20 plants per genotype were created and stored at room temperature until analysis. Two biologically (and temporally) independent experiments were performed resulting in 3 × 2 (genotype × experiment) seed pools.

Chemicals

All solvents used were of LC/MS grade quality (Riedel-de Haën). 3-[2-(4-Hydroxy-3-methoxyphenyl)-3-hydroxymethyl-7-methoxy-2,3-dihydrobenzofuran-5-yl]acrylic acid (ferulic acid 5-β cross-coupled to coniferyl alcohol) was purchased from Herbstandard. 3-{4-[2-Hydroxy-2-(4-hydroxy-3-methoxyphenyl)-1-hydroxymethylethoxy]-3-methoxyphenyl}acrylic acid (ferulic acid 4-*O*-β cross-coupled to coniferyl alcohol) was kindly provided by John Ralph (U.S. Dairy Forage Research Center). 2-*O*-Sinapoyl-L-malate and 1-*O*-sinapoyl-β-D-Glc were obtained from Alfred Baumert (Leibniz Institute of Plant Biochemistry). Further sources of authentic compounds are detailed in Supplemental Data S4.

Sample Preparation

Seeds (20 ± 1 mg) were homogenized in 1,000 μL of methanol:water (4:1, v/v) using 0.4 g of zirconium beads in a MiniBeadBeater (BioSpec Products) for 1 min at 3,200 rpm. After centrifugation at 11,000*g* for 10 min, 700 μL of the supernatant was evaporated to dryness in a vacuum centrifuge at ambient temperature. The remaining residue was redissolved in 70 μL of methanol by vigorous vortexing and diluted with 630 μL of water. Prior to LC/MS analysis, the extract was filtered through a 0.45-μm PTFE syringe filter (Whatman). Each of the six seed pools was extracted in quadruplicate. For quantification of sinapoylcholine, extracts were diluted 10-fold.

Fractionation of Seed Extracts

Two hundred fifty microliters of a seed extract was consecutively fractionated on Strata-X-CW and Strata-X-AW solid-phase extraction cartridges (60 mg, 33 μm; Phenomenex), both being solvated with methanol and equilibrated with methanol:water (1:9, v/v). After sample loading, the weak cation exchanger was washed with 1 mL of 25 mM ammonium acetate and 2 mL of methanol. Fraction I was eluted with 1 mL of methanol:acetonitrile:formic acid (19:79:2, v/v). Afterward, fraction II was obtained by elution with 1 mL of methanol:acetonitrile:formic acid (17.5:77.5:5, v/v). The combined wash solution was evaporated to dryness, reconstituted in 250 μL of methanol:water (1:9, v/v), and applied to the weak anion exchanger. Fractions III and IV were obtained by consecutive elution with 2 mL of methanol:water (4:1, v/v) and 2 mL of methanol:ammonia (98:2, v/v), respectively. All fractions were evaporated to dryness and reconstituted in 250 μL of methanol:water (4:1, v/v).

Capillary LC/ESI(+)-QTOF-MS

One microliter of the extract was separated using a capillary LC system (Ultimate; Dionex) equipped with an autosampler (Famos; Dionex) on a modified C_{18} column (GROMSIL ODS 4 HE; 150 × 0.3 mm; particle size, 3 μm; pore size, 120 Å; guard column, 10 × 0.3 mm [Alltech Grom]) applying the following binary gradient: 0 to 5 min, isocratic 95% A (water:formic acid, 99.9:0.1 [v/v]) and 5% B (acetonitrile:formic acid, 99.9:0.1 [v/v]); 5 to 55 min, linear from 5% to 55% B; 55 to 65 min, isocratic 95% B; 65 to 75 min, isocratic 5% B. Flow rate was 5 μL min^{-1}. Eluted compounds were detected by an API QSTAR Pulsar Hybrid QTOF mass spectrometer (Applied Biosystems/MDS Sciex) equipped with an ion spray source in positive ion mode. Typical instrument settings were as follows: ion spray voltage, 5.5 kV; DP1, 50 V; DP2, 15 V; FP, 220 V; nebulizer gas (nitrogen), 25 arbitrary units; curtain gas (nitrogen), 20 arbitrary units; collision gas (nitrogen), 4 arbitrary units; pulser frequency, 9.986 kHz; accumulation time, 2 s. Ions were detected in enhance-all mode from *m/z* 75 to 1,000 within four Q_1 transmission windows: *m/z* 55 to 113, 10% scan time; *m/z* 113 to 225, 20% scan time; *m/z* 225 to 450, 35% scan time; *m/z* 450 to 1,000, 35% scan time. The mass spectrometer was operated under Analyst QS 1.0. Mass calibration was performed using protonated ALILTLVS (*m/z* 829.5393) and a fragment ion (*m/z* 149.0233) originating from phthalate-type plasticizers. Mass resolution for $[M+H]^+$ of the calibration peptide was R_{FWHM} (resolution at full width at half maximum) = 8,500. For improvement of mass accuracy, internal single-point recalibration was applied using already identified mass signals with *m/z*, retention time, and intensity comparable to the unknown mass signal.

Data Analysis

Raw data files were converted to netCDF format using the Analyst QS file translator utility and processed using the XCMS package (http://metlin.scripps.edu/download/). For the pairwise comparison of two genotypes, chromatograms were grouped in four classes (genotype × experiment), each containing four technical replicates. Peak detection was performed using the parameter settings snr = 3, fwhm = 30 s, step = 0.1 D, mzdiff = 0.1 D, profmethod = binlinbase. Retention time correction was achieved in two iterations considering approximately 150 to 200 peak groups applying the parameter settings minfrac = 1, bw = 60 s, mzwid = 0.1 D, span = 1, missing = extra = 0 for the first iteration and minfrac = 1, bw = 30 s, mzwid = 0.1 D, span = 0.6, missing = extra = 0 for the second iteration. After final peak grouping (minfrac = 0.75, bw = 20 s) and filling in of missing features using the fillPeaks routine of the XCMS package, a data matrix (feature × sample) was obtained that was imported into Microsoft Excel. For further analysis, only consistent mass signals were considered, which were at least present in a single genotype in three of four technical replicates in both experiments. For this subset of signals, fold changes and *P* values (*t* test, two-sided, unequal variance) were calculated for each experiment. Signals were called differential when fold change was higher than 2 at a significance level of 0.05 in both experiments. After tentative identification, fold changes and *P* values were recalculated for corresponding quantifier ions using manually integrated peak areas from Analyst QS QuantWizard. For multivariate analysis of the whole data set, chromatograms were grouped in six classes (genotype × experiment) each containing four technical replicates. Peak detection and retention time correction were performed using a similar set of parameters as described above. After filling in missing features and elimination of inconsistent features, a data matrix consisting of 660 features (rows) × 24 samples (columns) was obtained. Signal intensities were transformed to fold changes by dividing each row of the matrix by the median of the wild-type intensities. After log transformation, principal component analysis was performed using SPSS 11.0.0 (SPSS).

MS/MS

Product ion spectra were acquired with Q_1 operating at unit resolution applying collision energies in the range of 15 to 55 eV. Nitrogen was used as collision gas at a pressure of 4 arbitrary units. Product ions were detected in enhance-all mode using q_2 transmission windows and pulser frequencies as suggested by the software. For fragmentation of precursor ions that are prone to in-source fragmentation, DP1 was decreased to 20 V. For acquisition of product ion spectra of in-source fragments (pseudo-MS^3), DP1 was increased to 80 V.

ESI-FTICR-MS

The positive ion high-resolution ESI mass spectra were obtained on a Bruker Apex III FTICR mass spectrometer (Bruker Daltonics) equipped with an Infinity cell, a 7.0-Tesla superconducting magnet (Bruker), a radiofrequency-only hexapole ion guide, and an APOLLO electrospray ion source (Agilent, off axis spray; voltages: end plate, −3,700 V; capillary, −4,200 V; capillary exit, 100 V; skimmer 1, 15.0 V; skimmer 2, 6.0 V). Nitrogen was used as drying gas at 150°C. The sample solutions were introduced continuously via a syringe pump with a flow rate of 120 μL h^{-1}. All data were acquired with 512 k data points and zero filled to 2,048 k by averaging 32 scans. The resolution at *m/z* 310.1649 (M^+ of sinapoylcholine) was approximately 50,000. The used mass range (*m/z* 100–2,000) was externally calibrated by the ES tuning mix (Agilent).

Synthesis of Phenolic Choline Ester

One hundred micromoles of the corresponding benzoic/cinnamic acid was dissolved in 100 μL of 1 M NaOH and 200 μL of dimethylformamide. After addition of 200 μL of 1 M (2-bromoethyl)trimethylammonium bromide (Merck) in water, the reaction mixture was heated for 24 h at 90°C. Fifty microliters of the crude reaction mixture was diluted with 1 mL of water and purified on Strata-X-CW (60 mg, 33 μm; Phenomenex) as described above. The eluate was evaporated to dryness in a vacuum centrifuge, reconstituted in methanol:water (1:9, v/v), and analyzed by LC/ESI-QTOF-MS. Syntheses of feruloylcholine(4-*O*-β)guaiacyl (**17**) and feruloylcholine(5-β)guaiacyl (**19**) were performed at 5-μmol scale.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Data S1. Method validation for seed metabolite profiling by LC/ESI-QTOF-MS (figures I–VIII; tables I–VIII).

Supplemental Data S2. Response curves of quantifier ions of tentatively identified metabolites.

Supplemental Data S3. Lists of all differential features.

Supplemental Data S4. Spectroscopic data of all identified and putatively annotated compounds.

ACKNOWLEDGMENTS

We thank Bernd Weisshaar (Bielefeld) for providing *tt4* seeds and the Nottingham Arabidopsis Stock Centre for *tt5* seeds. John Ralph (U.S. Dairy Forage Research Center) and Alfred Baumert (Leibniz Institute of Plant Biochemistry) kindly provided standards. Expert technical assistance by Michaela Winkler is gratefully acknowledged.

Received February 13, 2008; accepted June 11, 2008; published June 13, 2008.

LITERATURE CITED

Abdulrazzak N, Pollet B, Ehlting J, Larsen K, Asnaghi C, Ronseau S, Proux C, Erhardt M, Seltzer V, Renou JP, et al (2006) A coumaroyl-ester-3-hydroxylase insertion mutant reveals the existence of nonredundant meta-hydroxylation pathways and essential roles for phenolic precursors in cell expansion and plant growth. Plant Physiol **140:** 30–48

Besseau S, Hoffmann L, Geoffroy P, Lapierre C, Pollet B, Legrand M (2007) Flavonoid accumulation in *Arabidopsis* repressed in lignin synthesis affects auxin transport and plant growth. Plant Cell **19:** 148–162

Bigler L, Schnider CF, Hu W, Hesse M (1996) Electrospray-ionization mass spectrometry. 3. Acid catalyzed isomerization of *N*,*N*′-bis[(*E*)-3-(4-hydroxyphenyl)prop-2-enoyl]spermidines by the *ZIP* reaction. Helv Chim Acta **79:** 2152–2163

Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, Draper J, Nikolau BJ, Mendes P, Roessner-Tunali U, Beale MH, et al (2004) Potential of metabolomics as a functional genomics tool. Trends Plant Sci **9:** 418–425

Böttcher C, Roepenack-Lahaye EV, Willscher E, Scheel D, Clemens S (2007) Evaluation of matrix effects in metabolite profiling based on capillary liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry. Anal Chem **79:** 1507–1513

Boyes DC, Zayed AM, Ascenzi R, McCaskill AJ, Hoffman NE, Davis KR, Gorlach J (2001) Growth stage-based phenotypic analysis of *Arabidopsis*: A model for high throughput functional genomics in plants. Plant Cell **13:** 1499–1510

Brown DE, Rashotte AM, Murphy AS, Normanly J, Tague BW, Peer WA, Taiz L, Muday GK (2001) Flavonoids act as negative regulators of auxin transport in vivo in Arabidopsis. Plant Physiol **126:** 524–535

Buchanan B, Gruissem W, Jones R (2000) Biochemistry and Molecular Biology of Plants. American Society of Plant Biologists, Rockville, MD

Buer CS, Muday GK (2004) The transparent testa4 mutation prevents flavonoid synthesis and alters auxin transport and the response of *Arabidopsis* roots to gravity and light. Plant Cell **16:** 1191–1205

Burbulis IE, Iacobucci M, Shirley BW (1996) A null mutation in the first enzyme of flavonoid biosynthesis does not affect male fertility in *Arabidopsis*. Plant Cell **8:** 1013–1025

Cain CC, Saslowsky DE, Walker RA, Shirley BW (1997) Expression of chalcone synthase and chalcone isomerase proteins in Arabidopsis seedlings. Plant Mol Biol **35:** 377–381

Chernushevich IV, Loboda AV, Thomson BA (2001) An introduction to quadrupole-time-of-flight mass spectrometry. J Mass Spectrom **36:** 849–865

Clausen S, Olsen O, Sorensen H (1982) 4-Hydroxybenzoylcholine: a natural product in Sinapis alba. Phytochemistry **21:** 917–922

Clemens S, Böttcher C, Franz M, Willscher E, von Roepenack-Lahaye E, Scheel D (2006) Capillary HPLC coupled to electrospray ionization quadrupole time-of-flight mass spectrometry. *In* K Saito, RA Dixon, L Willmitzer, eds, Plant Metabolomics. Springer, Heidelberg, pp 65–79

D'Auria JC, Gershenzon J (2005) The secondary metabolism of Arabidopsis thaliana: growing like a weed. Curr Opin Plant Biol **8:** 308–316

De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat Protocols **2:** 778–791

Dixon RA, Gang DR, Charlton AJ, Fiehn O, Kuiper HA, Reynolds TL, Tjeerdema RS, Jeffery EH, German JB, Ridley WP, et al (2006) Applications of metabolomics in agriculture. J Agric Food Chem **54:** 8984–8994

Dunn WB, Ellis DI (2005) Metabolomics: current analytical platforms and methodologies. TrAC Trends Anal Chem **24:** 285–294

Facchini PJ, Bird DA, St-Pierre B (2004) Can Arabidopsis make complex alkaloids? Trends Plant Sci **9:** 116–122

Fernie AR, Trethewey RN, Krotzky AJ, Willmitzer L (2004) Metabolite profiling: from diagnostics to systems biology. Nat Rev Mol Cell Biol **5:** 763–769

Fiehn O (2002) Metabolomics: the link between genotypes and phenotypes. Plant Mol Biol **48:** 155–171

Fiehn O, Wohlgemuth G, Scholz M, Kind T, Lee DY, Moon S, Nikolau B (2008) Quality control for plant metabolomics: reporting MSI-compliant studies. Plant J **53:** 691–704

Fraser CM, Thompson MG, Shirley AM, Ralph J, Schoenherr JA, Sinlapadech T, Hall MC, Chapple C (2007) Related Arabidopsis serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities. Plant Physiol **144:** 1986–1999

Fundel K, Küffner R, Aigner T, Zimmer R (2005) Data processing effects on the interpretation of microarray gene expression experiments. *In* A Torda, S Kurtz, M Rarey, eds, German Conference on Bioinformatics (GCB) 2005. GI Lecture Notes in Informatics, P-71. Gesellschaft für Informatik, Bonn, pp 77–91

Harborne JB, Williams CA (2000) Advances in flavonoid research since 1992. Phytochemistry **55:** 481–504

Hemm MR, Ruegger MO, Chapple C (2003) The *Arabidopsis* ref2 mutant is defective in the gene encoding CYP83A1 and shows both phenylpropanoid and glucosinolate phenotypes. Plant Cell **15:** 179–194

Kachlicki P, Einhorn J, Muth D, Kerhoas L, Stobiecki M (2008) Evaluation of glycosylation and malonylation patterns in flavonoid glycosides during LC/MS/MS metabolite profiling. J Mass Spectrom **43:** 572–586

Kind T, Fiehn O (2006) Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics **7:** 234

Konishi Y, Kiyota T, Draghici C, Gao J, Yeboah F, Acoca S, Jarussophon S, Purisima E (2007) Molecular formula analysis by an MS/MS/MS technique to expedite dereplication of natural products. Anal Chem **79:** 1187–1197

Koornneef M (1990) Mutations affecting the testa color in Arabidopsis. Arabidopsis Inf Serv **28:** 1–4

Last RL, Jones AD, Shachar-Hill Y (2007) Towards the plant metabolome and beyond. Nat Rev Mol Cell Biol **8:** 167–174

Li J, Ou-Lee TM, Raba R, Amundson RG, Last RL (1993) *Arabidopsis* flavonoid mutants are hypersensitive to UV-B irradiation. Plant Cell **5:** 171–179

Lim EK, Li Y, Parr A, Jackson R, Ashford DA, Bowles D (2001) Identification of glucosyltransferase genes involved in sinapate metabolism and lignin synthesis in Arabidopsis. J Biol Chem **276:** 4344–4349

Milkowski C, Baumert A, Schmidt D, Nehlin L, Strack D (2004) Molecular regulation of sinapate ester metabolism in Brassica napus: expression of genes, properties of the encoded proteins and correlation of enzyme activities with metabolite accumulation. Plant J **38:** 80–92

Milkowski C, Baumert A, Strack D (2000) Identification of four Arabidopsis genes encoding hydroxycinnamate glucosyltransferases. FEBS Lett **486:** 183–184

Moco S, Bino RJ, Vorst O, Verhoeven HA, De Groot J, Van Beek TA, Vervoort J, De Vos CHR (2006) A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol **141:** 1205–1218

Nielsen J, Oliver S (2005) The next wave in metabolome analysis. Trends Biotechnol **23:** 544–546

Nordström A, O'Maille G, Qin C, Siuzdak G (2006) Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: quantitative analysis of endogenous and exogenous metabolites in human serum. Anal Chem **78:** 3289–3295

Olsen O, Sorensen H (1980) Sinalbin and other glucosinolates in seeds of double low rape species and Brassica napus cv. Bronowski. J Agric Food Chem **28:** 43–48

Peer WA, Brown DE, Tague BW, Muday GK, Taiz L, Murphy AS (2001) Flavonoid accumulation patterns of transparent testa mutants of Arabidopsis. Plant Physiol **126:** 536–548

Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ Jr, Hallett JP, Leak DJ, Liotta CL, et al (2006) The path forward for biofuels and biomaterials. Science **311:** 484–489

Ross JA, Kasum CM (2002) Dietary flavonoids: bioavailability, metabolic effects, and safety. Annu Rev Nutr **22:** 19–34

Routaboul JM, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J, Lepiniec L (2006) Flavonoid diversity and biosynthesis in seed of Arabidopsis thaliana. Planta **224:** 96–107

Ryan D, Robards K (2006) Metabolomics: the greatest omics of them all? Anal Chem **78:** 7954–7958

Shinbo Y, Nakamura Y, Altaf-Ul-Amin M, Asahi H, Kurokawa K, Arita M, Saito K, Ohta D, Shibata D, Kanaya S (2006) KNApSAcK: a comprehensive species-metabolite relationship database. *In* K Saito, RA Dixon, L Willmitzer, eds, Plant Metabolomics. Springer, Heidelberg, pp 165–184

Shiono M, Matsugaki N, Takeda K (2005) Phytochemistry: structure of the blue cornflower pigment. Nature **436:** 791

Shirley AM, Chapple C (2003) Biochemical characterization of sinapoylglucose:choline sinapoyltransferase, a serine carboxypeptidase-like protein that functions as an acyltransferase in plant secondary metabolism. J Biol Chem **278:** 19870–19877

Shirley AM, McMichael CM, Chapple C (2001) The *sng2* mutant of *Arabidopsis* is defective in the gene encoding serine carboxypeptidase-like protein sinapoylglucose:choline sinapoyltransferase. Plant J **28:** 83–94

Shirley BW, Hanley S, Goodman HM (1992) Effects of ionizing radiation on a plant genome: analysis of two *Arabidopsis* transparent testa mutations. Plant Cell **4:** 333–347

Shirley BW, Kubasek WL, Storz G, Bruggemann E, Koornneef M, Ausubel FM, Goodman HM (1995) Analysis of Arabidopsis mutants deficient in flavonoid biosynthesis. Plant J **8:** 659–671

Sinlapadech T, Stout J, Ruegger MO, Deak M, Chapple C (2007) The hyper-fluorescent trichome phenotype of the *brt1* mutant of Arabidopsis is the result of a defect in a sinapic acid:UDPG glucosyltransferase. Plant J **49:** 655–668

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem **78:** 779–787

Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, et al (2007) Proposed minimum reporting standards for chemical analysis. Metabolomics **3:** 211–221

Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics era. Phytochemistry **62:** 817–836

Tang L, Kebarle P (1993) Dependence of ion intensity in electrospray mass spectrometry on the concentration of the analytes in the electrosprayed solution. Anal Chem **65:** 3654–3668

Tang KQ, Page JS, Smith RD (2004) Charge competition and the linear dynamic range of detection in electrospray ionization mass spectrometry. J Am Soc Mass Spectrom **15:** 1416–1423

Tautenhahn R, Böttcher C, Neumann S (2007) Annotation of LC/ESI-MS mass signals. *In* Lecture Notes in Computer Science. Bioinformatics Research and Development. Springer, Heidelberg, pp 371–380

Taylor PJ (2005) Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography-electrospray-tandem mass spectrometry. Clin Biochem **38:** 328–334

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

Wildermuth MC (2006) Variations on a theme: synthesis and modification of plant benzoic acids. Curr Opin Plant Biol **9:** 288–296

Winkel-Shirley B (2001) Flavonoid biosynthesis: a colorful model for genetics, biochemistry, cell biology, and biotechnology. Plant Physiol **126:** 485–493

Youhnovski N, Bigler L, Werner C, Hesse M (1998) On-line coupling of high-performance liquid chromatography to atmospheric pressure chemical ionization mass spectrometry (HPLC/APCI-MS and MS/MS): the pollen analysis of Hippeastrum x hortorum (Amaryllidaceae). Helv Chim Acta **81:** 1654–1671

Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR: Arabidopsis microarray database and analysis toolbox. Plant Physiol **136:** 2621–32

Zook DR, Bruins AP (1997) On cluster ions, ion transmission, and linear dynamic range limitations in electrospray (ionspray) mass spectrometry. Int J Mass Spectrom Ion Process **162:** 129–147