Metabolomic Characterization of Knockout Mutants in Arabidopsis: Development of a Metabolite Profiling Database for Knockout Mutants in Arabidopsis[1][W][OPEN]

Atsushi Fukushima*, Miyako Kusano, Ramon Francisco Mejia, Mami Iwasa, Makoto Kobayashi, Naomi Hayashi, Akiko Watanabe-Takahashi, Tomoko Narisawa, Takayuki Tohge, Manhoi Hur, Eve Syrkin Wurtele, Basil J. Nikolau, and Kazuki Saito*

RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230–0045, Japan (A.F., Mi.K., R.F.M., M.I., Ma.K., N.H., A.W.-T., T.N., T.T., K.S.); Japan Science and Technology Agency, National Bioscience Database Center, Chiyoda-ku, Tokyo 102–0081, Japan (A.F.); Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki 305–8572, Japan (Mi.K.); Nissan Chemical Industries, Funabashi, Chiba 274–8507, Japan (M.I.); Max-Planck-Institute of Molecular Plant Physiology, 14476 Potsdam-Golm, Germany (T.T.); Department of Genetics Development and Cell Biology (M.H., E.S.W.), Center for Metabolic Biology (E.S.W., B.J.N.), Center for Biorenewable Chemicals (E.S.W., B.J.N.), and Biochemistry, Biophysics, and Molecular Biology (B.J.N.), Iowa State University, Ames, Iowa 50011; and Graduate School of Pharmaceutical Sciences, Chiba University, Chiba-shi, Chiba 263–8522, Japan (K.S.)

Despite recent intensive research efforts in functional genomics, the functions of only a limited number of Arabidopsis (*Arabidopsis thaliana*) genes have been determined experimentally, and improving gene annotation remains a major challenge in plant science. As metabolite profiling can characterize the metabolomic phenotype of a genetic perturbation in the plant metabolism, it provides clues to the function(s) of genes of interest. We chose 50 Arabidopsis mutants, including a set of characterized and uncharacterized mutants, that resemble wild-type plants. We performed metabolite profiling of the plants using gas chromatography-mass spectrometry. To make the data set available as an efficient public functional genomics tool for hypothesis generation, we developed the Metabolite Profiling Database for Knock-Out Mutants in Arabidopsis (MeKO). It allows the evaluation of whether a mutation affects metabolism during normal plant growth and contains images of mutants, data on differences in metabolite accumulation, and interactive analysis tools. Nonprocessed data, including chromatograms, mass spectra, and experimental metadata, follow the guidelines set by the Metabolomics Standards Initiative and are freely downloadable. Proof-of-concept analysis suggests that MeKO is highly useful for the generation of hypotheses for genes of interest and for improving gene annotation. MeKO is publicly available at http://prime.psc.riken.jp/meko/.

[1] This work was supported by the Japan Science and Technology Agency Strategic International Collaborative Research Program, by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, by the Japan Advanced Plant Science Network, by the U.S. National Science Foundation (grant nos. MCB–0820823, IOS–1139489, and MCB–0951170 to B.J.N. and E.S.W.), and by the National Institutes of Health Institute of General Medical Sciences (grant no. NIGM–99511 to B.J.N. and E.S.W.).

* Address correspondence to a-fukush@psc.riken.jp and kazuki.saito@riken.jp.

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: Atsushi Fukushima (a-fukush@psc.riken.jp).

A.F., Mi.K., and T.T. designed the research; A.F., Mi.K., M.I., and T.T. performed the research; R.F.M. provided computational assistance to A.F.; Ma.K., N.H., A.W.-T., and T.N. performed the experiments; A.F., Mi.K., M.I., Ma.K., N.H., and T.T. analyzed the data; M.H., E.S.W., and B.J.N. contributed the analysis/visualization tools; A.F. and Mi.K. wrote the article; K.S. supervised and coordinated the project consortium.

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

[OPEN] Articles can be viewed online without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.114.240986

The availability of the genome sequence of Arabidopsis (*Arabidopsis thaliana*) and of postgenomic tools such as transcriptomics and metabolomics greatly facilitated the large-scale identification of both gene functions and physiological roles. Based on direct experimental evidence, approximately 20% of all Arabidopsis genes (approximately 28,000 genes in total) have been assigned at least one Gene Ontology (GO) term for a biological process, approximately 27% for a cellular component, and approximately 13% for a molecular function (Lamesch et al., 2012). Computational evidence, including homology-based searches, has also been used to assign at least one GO term to approximately 77% of all Arabidopsis genes (Lamesch et al., 2012). The determination and improvement of gene annotation remain important challenges in the field of plant systems biology.

Metabolomics has become a powerful tool for providing a global snapshot of metabolic behavior in a cell (Fernie et al., 2004; Fukushima et al., 2009b; Saito and Matsuda, 2010; Allwood et al., 2011; Lei et al., 2011). Gas chromatography (GC)-mass spectrometry (MS), liquid chromatography (LC)-MS, and NMR are widely used in this field. GC-MS and LC-MS are sensitive, robust, and

reproducible techniques for the comprehensive metabolomic profiling of a wide range of metabolites (Tohge and Fernie, 2009). They facilitated study of the coordination of carbon and nitrogen metabolism in plants (Stitt and Fernie, 2003; Rubin et al., 2009; Pracharoenwattana et al., 2010; Kusano et al., 2011a; Amiour et al., 2012) and studies to gain a better understanding of regulatory networks involved in genetic perturbation (Roessner et al., 2001; Weckwerth et al., 2004; Tohge et al., 2005), to characterize diurnal/circadian behaviors (Urbanczyk-Wochniak et al., 2005; Gibon et al., 2006; Fukushima et al., 2009a; Espinoza et al., 2010; Hoffman et al., 2010), to assess altered regulatory responses to various abiotic stresses (Kaplan et al., 2004; Urano et al., 2009; Caldana et al., 2011; Kusano et al., 2011b; Maruyama et al., 2014; Nakabayashi et al., 2014), and to identify comprehensive metabolite quantitative trait loci (Morreel et al., 2006; Schauer et al., 2006; Carreno-Quintero et al., 2012; Matsuda et al., 2012).

Database resources, including information on compounds and mass spectra, are essential for MS-based metabolomics because metabolite identification based on such information is central for gaining insights into physiological and cellular responses (Fukushima and Kusano, 2013). The Metabolomics Standard Initiative (MSI; Sumner et al., 2007) published guidelines regarding experimental and analytical metadata, including metabolite identification in metabolomics. Fernie et al. (2011) have also presented additional practical recommendations for reporting metabolome data. In this context, Bais et al. (2010, 2012) developed the first metabolite profiling database with statistical data analysis tools for Arabidopsis mutants using multiple analytical instruments. The data include MSI-compliant experimental metadata and freely available metabolite profiles. The data and metadata for these mutants are archived at the Plant and Microbial Metabolomics Resource (PMR; http://www.metnetdb.org/pmr/). The PMR is a flexible, multispecies, Not Only Structured Query Language-empowered database that is designed for data sharing and has computational tools that enable the analysis of metabolomics and transcriptomics data (Hur et al., 2013). Chloroplast 2010 (Lu et al., 2011) includes information on the large-scale phenotypic screening of Arabidopsis chloroplast mutants using assays of fatty acids and amino acids of more than 10,000 transfer DNA insertion mutants in leaves and seeds using GC-MS and LC-MS (Gu et al., 2007; Bell et al., 2012). Although such data resources can provide an opportunity to generate testable hypotheses from investigations and evaluations of changes in metabolite levels for mutants with unknown gene functions (Quanbeck et al., 2012), only a few databases facilitate the sharing of metabolomics data.

Because metabolite profiling can characterize the metabolomic phenotype (metabotype) of a genetic perturbation of the plant metabolism, it provides clues to the functions of genes of interest. We chose 50 Arabidopsis plants, including a set of characterized and uncharacterized mutants, that share the same genetic background. We performed metabolite profiling of these plants using GC-MS. For efficient use of the data set, we developed the Metabolite Profiling Database for Knock-Out Mutants in Arabidopsis (MeKO; http://prime.psc.riken.jp/meko/), which can be used to browse and visualize metabotype data. It contains images of mutants, data on differences in the accumulation of metabolites, and the results of statistical data analyses. Nonprocessed data, including chromatograms and mass spectra, and the corresponding experimental metadata are MSI compatible and are freely available from the database. MeKO is a public functional genomics tool that can be used to gain insights into individual genes by evaluating whether their mutation affects metabolism during normal plant growth.

RESULTS

Experimental Setup and Design Overview

We focused on 180 Arabidopsis mutants with altered metabolite levels (e.g. ascorbate-deficient mutants); they were obtained from the Arabidopsis Biological Resource Center (https://abrc.osu.edu/). We tested their seed germination on Murashige and Skoog agar medium containing 1% Suc under strictly controlled conditions (see "Materials and Methods") and chose 50 mutants whose seedlings developed healthily under our laboratory conditions (Supplemental Table S1). Visual inspection of all mutant plants (390 individuals in total) showed a resemblance to wild-type plants in 91.5% for the leaf color, in 92.3% for the leaf shape, and in 98.9% for the root shape (Fig. 1A). This indicates that most of neither plant growth nor development was affected in the mutants. Several mutants manifested mutations in genes of unknown function. Figure 1B shows the results of GO enrichment analysis for the functional annotation of the 50 mutants. The significantly overrepresented functional biological process categories were involved in the response to abiotic stress, cell growth, carbohydrate metabolic process, and secondary metabolic process. For example, in the case of response to abiotic stress, this is consistent with our choosing some abscisic acid mutants (Supplemental Table S1). All mutants used in this study are of the Columbia background. To facilitate large-scale comparative metabolomics studies, we planted the Columbia-0 (Col-0) wild type on each plate as a control. Using previously established GC-MS-based metabolite profiling (Kusano et al., 2007), we investigated changes in the metabolite levels of the aerial parts of the mutants and wild-type plants harvested 18 d after germination. Based on these data, we constructed MeKO for use as a functional genomics tool to browse and visualize metabotypes (Fig. 1C).

Reducing the Systematic Batch Effect in Large-Scale Data Sets Renders Them Accessible to a Global Comparison of Arabidopsis Metabotypes

We analyzed more than 650 samples, including 629 plants and 27 quality control samples, by GC-MS (see

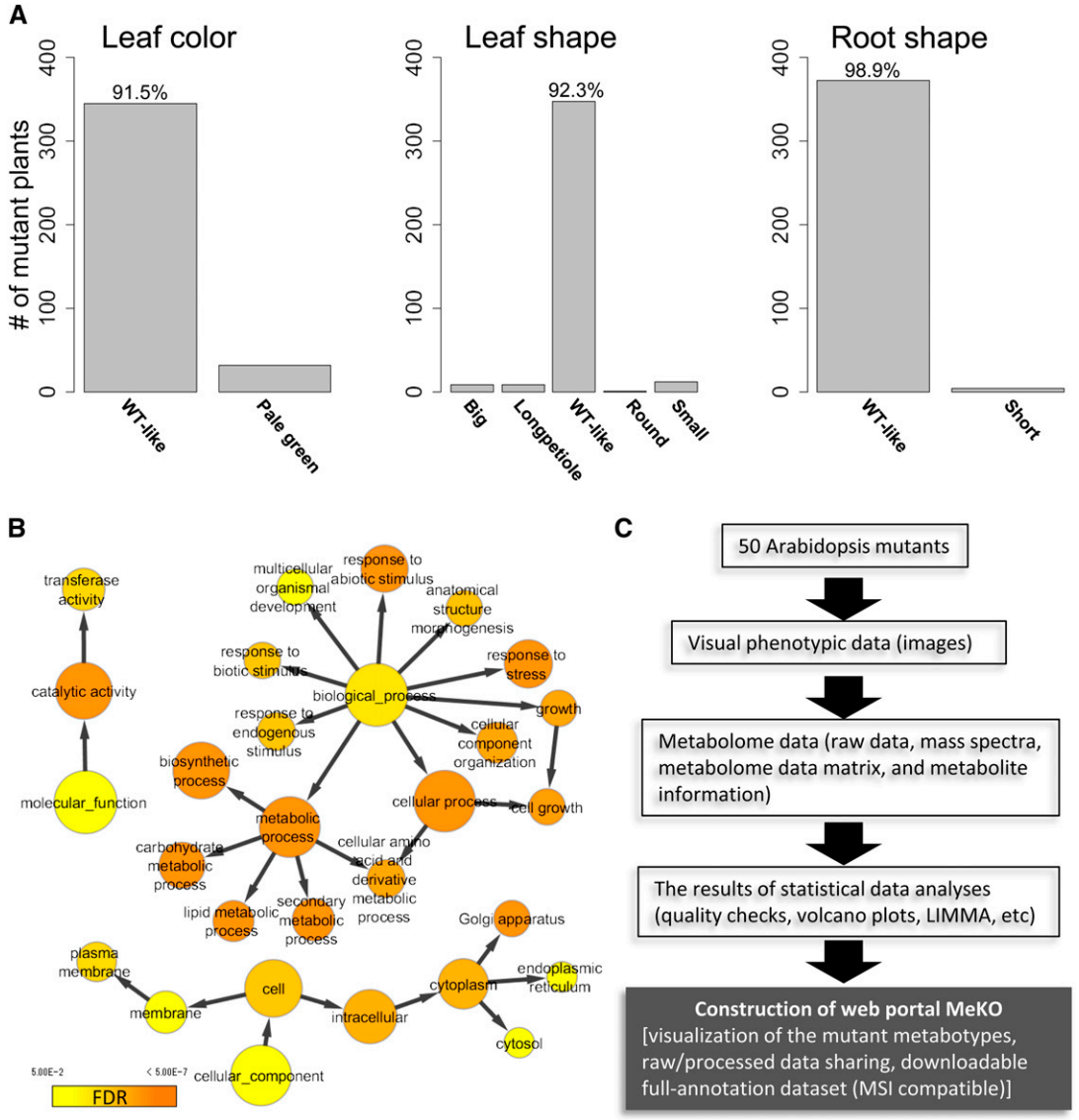

Figure 1. The morphological phenotypes of the 50 mutants, the flow used to assess Arabidopsis mutants by GC-MS-based metabolomics, and the construction of MeKO. A, Statistical distribution of the visualized characteristics of the mutant phenotypes on medium plates. Based on visual inspection, 91.5% of all mutant plants (390 individuals in total) resembled wild-type plants (WT) with respect to the leaf color, 92.3% with respect to the leaf shape, and 98.9% with respect to the root shape. Plants were grown on petri plates for 18 d after germination before assessment. B, Results of GO term enrichment analysis for functional annotation of the 50 mutant genes using BiNGO (Maere et al., 2005). Significantly overrepresented categories include biological processes involved in response to abiotic stress, cell growth, carbohydrate metabolism, and secondary metabolism. C, Schematic representation of the information workflow. Users can access and use various contents, such as visualization of the mutant metabotypes, nonprocessed/processed data, and the results of statistical data analyses.

"Materials and Methods"). Briefly, our quality control samples were treated exactly like the experimental samples and were prepared by mixing 100-μL extracts of each Col-0 sample. Because these data sets contained a large number of samples, we analyzed them on 5 different days and preprocessed them separately. In Table I and Supplemental Table S2, we show all metabolites detected within each batch data set. On average, 95 metabolites were identified/annotated, including 16 mass spectral tags (MSTs; Schauer et al., 2005). To compare the metabolic phenotypes of the 50 mutants simultaneously, we constructed a single data matrix consisting of the five data sets (designated the All data set) with appropriate preprocessing and normalizing methods. Using this design, we carefully normalized the data to reduce the so-called batch effect (i.e. unwanted nonbiological variations; Leek et al., 2010; Lazar et al., 2013). Before batch normalization, we observed

Table I. *Numbers of metabolite peaks detected in this study*

Batch Name	No. of Detected Metabolites	No. of Identified/ Annotated Metabolites	No. of MSTs	No. of Unknown Peaks	No. of Samples	Mutant Name	No. of Mutants
1	240	96	15	129	144	*aba1-5*, *aba2-3*, *constitutive immunity6* (*cim6*), *cim7*, Col-0, *eto1-1*, *eto3*, *glutamine sensing regulator1-1*, *isoxaben-resistant1-2* (*ixr1-2*), and *rsw2-1*	9
2	214	102	16	96	175	*aba2-1*, *aba3-1*, *cim11*, *cim13*, *cim9*, Col-0, *fatty acid desaturase5-1* (*fad5-1*), *fad6-1*, *mur9-1*, *pad2-1*, *pad4-1*, *pap1-D*, and *Procuste1-1*	12
3	229	93	17	119	147	*cim1-4*, *cloroplastos alterados 1S*, *COBRA-2*, Col-0, *FUSCA6 1S*, *mur2-1*, *mur4-2*, *mur5-1*, *mur6-1*, and *pac-1S*	9
4	245	93	15	137	161	*aba1-6*, *ammonium transporter1;1*, *cim10*, Col-0, *ferulic acid 5-hydroxylase1-2*, *mur11-1*, *mur7-1*, *mur8-1*, *pad3-1*, *sinapoylglucose accumulator1-1*, *trichome birefringence1*, and *vtc1-1*	11
5	251	93	19	139	127	*aat2-2*, Col-0, *fluorouridine insensitive1-1*, *ixr1-1*, *mur1-1*, *mur1-2*, *mur3-2*, *phytochrome B9*, *rsw1-1*, and *rsw3-1*	9
Average	235.8	95.4	16.4	124	–	–	50

nonbiological variations in the score scatterplot of our principal component analysis (PCA; Fig. 2, top). Normalization removed batch effects in the data set, as evidenced by the score scatterplot (Fig. 2, bottom). Because adjusting the batch effects renders samples in different batches directly comparable, we were able to inspect changes in the metabolite levels across the 50 mutants. We used this single matrix, All, for further considerations.

Systematic Data Comparison Demonstrates That MeKO Complements the Information in PlantMetabolomics.org

When we compared the mutants on our list (Supplemental Table S1) with those generated by the Arabidopsis 2010 Plant Metabolomics Consortium (Bais et al., 2010, 2012; Quanbeck et al., 2012; Hur et al., 2013; http://www.metnetdb.org/pmr/), we found that on both, *pad4-1* (Arabidopsis Biological Resource Center identifier CS3806; AT3G52430, encoding the ARABIDOPSIS PHYTOALEXIN DEFICIENT4 [PAD4] α/β-hydrolase superfamily protein; Fig. 3A) was analyzed by GC-MS. Although there are some differences in the growth conditions and developmental ages between MeKO and the Arabidopsis 2010 Plant Metabolomics Consortium data, the metabolite profiles can be compared with GC-time-of-flight (TOF)-MS using data from this mutant. Therefore, we downloaded the profile data for *pad4-1* (experiment 13 E4) from PlantMetabolomics.org. Using MetMask (Redestig et al., 2010), the chemical identifier conversion tool, we identified 43 commonly detected metabolites in the two data sets. Figure 3B is a pseudocolored heat map reflecting the Pearson correlation matrices for each data set. The matrix values show the correlation coefficients for sample replicates. Overall, reproducibility within each data set was very high (red color in Fig. 3B); however, the correlation between our samples and the PlantMetabolomics.org samples was not as high as was the reproducibility (r = approximately 0.3; blue color in Fig. 3B). This lack of correlation was probably due to the different growth conditions used.

Cluster Analysis Suggests the Most Pronounced Changes in the Metabolite Levels among the Examined Mutants

After establishing that our metabolomics data are amenable to mutant characterization, we attempted to produce quantitative descriptions of the metabotypes of the 50 mutants. Figure 4 is an overview of the metabolite levels of the 50 Arabidopsis mutants by multidimensional scaling (MDS), a statistical technique to represent similarity among a mutant's metabotypes, based on tentatively annotated metabolites as well as MSTs that were consistently observed but incompletely identified (87 metabolite peaks in total). We observed that the mutant lines *ethylene-overproducer1-1* (*eto1-1*), *vitamin c1-1* (*vtc1-1*), *murus9-1* (*mur9-1*) and *mur11-1*, *aspartate aminotransferase2-2* (*aat2-2*), and *root swelling2-1* (*rsw2-1*) strongly exhibited metabotypes from the other mutants on coordinate 1 (magenta dotted circles in Fig. 4). For a wide range of metabolism, the metabolite levels in these mutants were higher than in Col-0. In contrast, they were lower in the *mur1-1*, *rsw1-1*, and *abscisic acid* (*ABA*)*-deficient mutant1-5* (*aba1-5*) mutants (cyan solid circles in Fig. 4). Coordinate 2 may reflect specific local changes in a metabolic pathway. For example, while *mur9-1* and *eto1-1* exhibited significant increases in the polyamine levels (false discovery rate [FDR] < 0.05), these levels were not changed in *eto3* and *mur4-3*. Compared with Col-0 (e.g. *production of anthocyanin pigmentation1-dominant* [*pap1-D*]; indicated by the arrow in Fig. 4), the impact on the primary metabolite levels of mutants at around the origin of the coordinate axes may be small. We focused on two mutants, *mur9-1* and *eto1-1*, that exhibited similar changes in the levels of a wide range of primary metabolites to

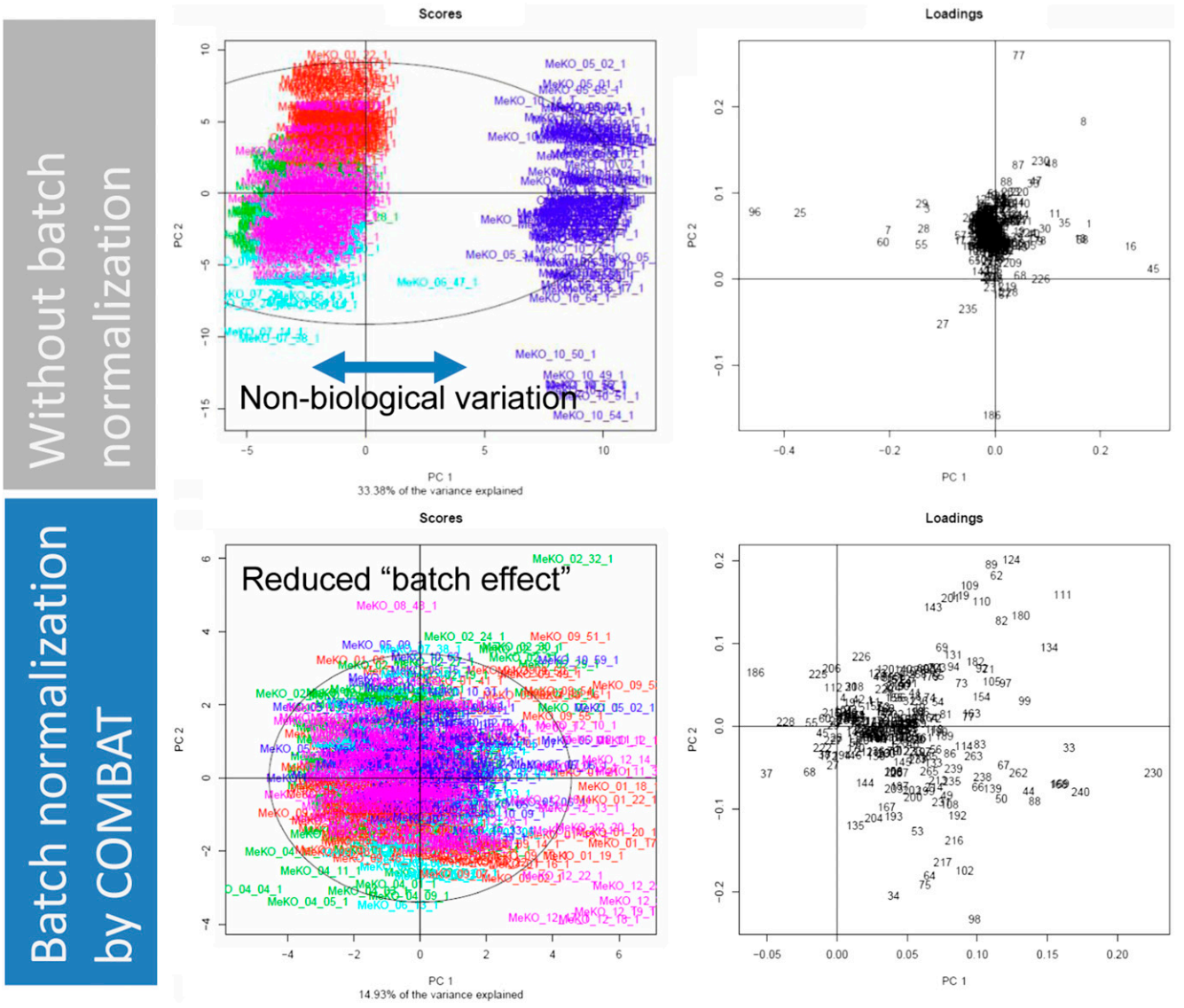

Figure 2. An example of unwanted batch effects in metabolomics data. PCA of metabolite data from GC-MS-based metabolomics without COMBAT (Johnson et al., 2007; top) and with COMBAT normalization to reduce batch effects (bottom) are shown. Normalization removed batch effects from the data set, as evidenced by the score scatterplot.

investigate the effects of unknown mutations on metabolism (Fig. 5). Both mutants exhibited a significant increase in the succinate level in the tricarboxylic acid cycle (FDR < 0.05). In addition to an increase in the level of γ-aminobutyric acid, they manifested a dramatic increase in the level of polyamines. Taken together, such large-scale comparative metabolomics studies may provide clues to the mechanisms responsible for functional loss attributable to mutations.

The Main Contents of MeKO

MeKO features a mutant page, a metabolite page, and an analysis page. Our database stores information on the mutants analyzed in this study (Fig. 6A); it contains a list of entries for the mutant selected from a panel, including its name, the Arabidopsis Genome Initiative (AGI) code, its gene name and description (Fig. 6B), mutant images (Fig. 6C), the results of replicate quality checks, and the volcano plot (Fig. 6D). In the plot, the top 30 metabolites with significant changes in the metabolite levels are highlighted in blue ($|\log_2$ fold change$| \geq 1$, FDR < 0.05). The reader can inspect changes in the metabolite levels in the selected mutant (Fig. 6E). Detailed information is provided below.

Mutant Annotation

The main menu provides information on each mutant we analyzed; this facilitates browsing the mutant/gene of interest (Fig. 6A). Each mutant page includes

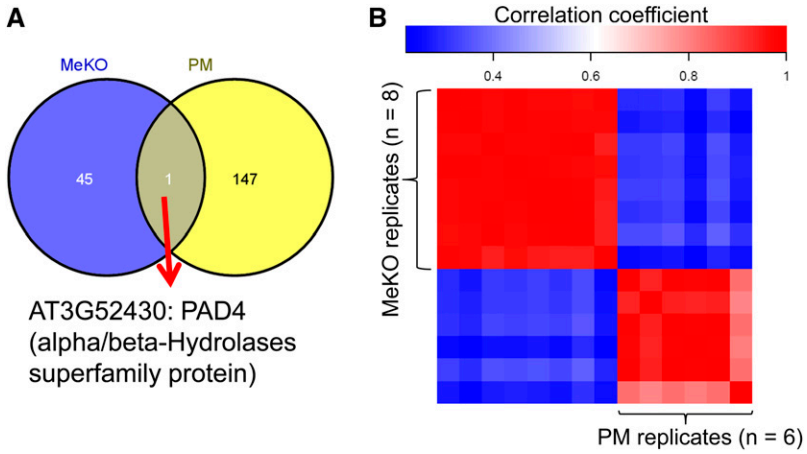

Figure 3. Comparison of MeKO and PlantMetabolomics.org (PM; Bais et al., 2010, 2012). A, Venn diagram of the mutants analyzed by the two projects. Only a single mutant was analyzed by both efforts. B, Reproducibility of metabolite profiling in MeKO and PlantMetabolomics.org tested using the Arabidopsis mutant *pad4* (AT3G52430; encoding an α/β-hydrolase superfamily protein) based on GC-TOF-MS-derived metabolomics data. Values in the matrix are the correlation coefficients between sample replicates. Reproducibility is quite high within each data set, but the correlation between samples from this study and PlantMetabolomics.org is not as high (Pearson correlation [*r*] = approximately 0.3). The data matrix includes 43 metabolites detected in the two projects.

the allele symbol, the AGI code, the gene name and description, and a cross-link to The Arabidopsis Information Resource (Lamesch et al., 2012; Fig. 6B). In addition, images of mutants grown under our growth conditions are available (Fig. 6C). For example, Figure 6C shows *aba1-5* lines that are small in size and weight and have a wilted shape. Users can compare the mutant of interest with a Col-0 wild type grown on each plate, thus allowing integration with morphological phenotype data.

Data Quality Check and Evaluation of Differences in the Accumulation of Metabolites

Assessment of our metabolomics measurements revealed adequate reproducibility across biological replicates. This can be easily verified by checking the scatterplots and correlation coefficients among different replicates (Fig. 6D, top). Using volcano plots, which measure differentially accumulated metabolites based on *t* statistics and fold changes simultaneously, we highlighted the top 30 metabolites with a statistically significant difference between a particular mutant and the wild type (LIMMA method [Smyth, 2004]; FDR < 0.05; blue plots in Fig. 6D, bottom). A tab-delimited text file of the results of the statistical data analysis can be downloaded from this page.

Profile Mapping to Metabolic Pathways

A first step in the interpretation of metabolome data is to visualize and map the detected metabolites in the context of metabolic pathways. MeKO features such visualization and provides a quick overview of the pathway level of metabolites that display significant increases/decreases in the mutant (Fig. 6E). This visualization functionality in MeKO contributes to the generation of testable hypotheses of gene function.

Metabolite Pages

Each detected metabolite is linked to relevant compound databases, such as Chemical Abstracts Service (http://www.cas.org/), PubChem (Wang et al., 2009), Chemical Entities of Biological Interest (Degtyarenko et al., 2008), and KNApSAcK (Afendi et al., 2012). Table II lists all linked databases. Investigators can use the links to obtain further information on chemical compounds of interests. This page also contains links to other

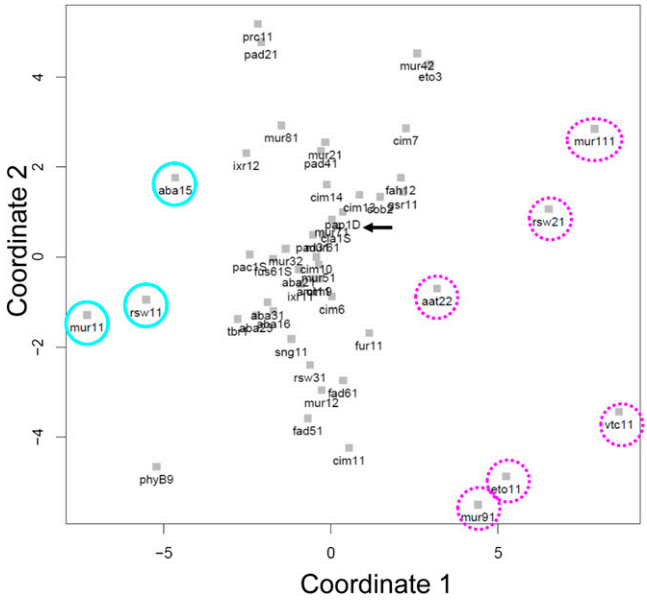

Figure 4. Overview of the metabolite levels of the 50 Arabidopsis mutants by MDS, a statistical technique to represent similarity among the metabotypes of a mutant. We used identified and tentatively annotated metabolites as well as MSTs (Schauer et al., 2005). The two-dimensional MDS plot represents similarity in the metabolite profiles among the 50 mutants. This MDS plot was based on a $\log_2$ fold change calculated by dividing the metabolite level in the mutant by the level in the Col-0 wild type. Magenta dotted circles indicate strong metabotypes from the other mutants in coordinate 1; in contrast, cyan solid circles show mutants that exhibited a decrease. In mutants at around the origin of the coordinate axes, there was little impact on the primary metabolite levels compared with Col-0 (e.g. *pap1-D*; indicated by the arrow).

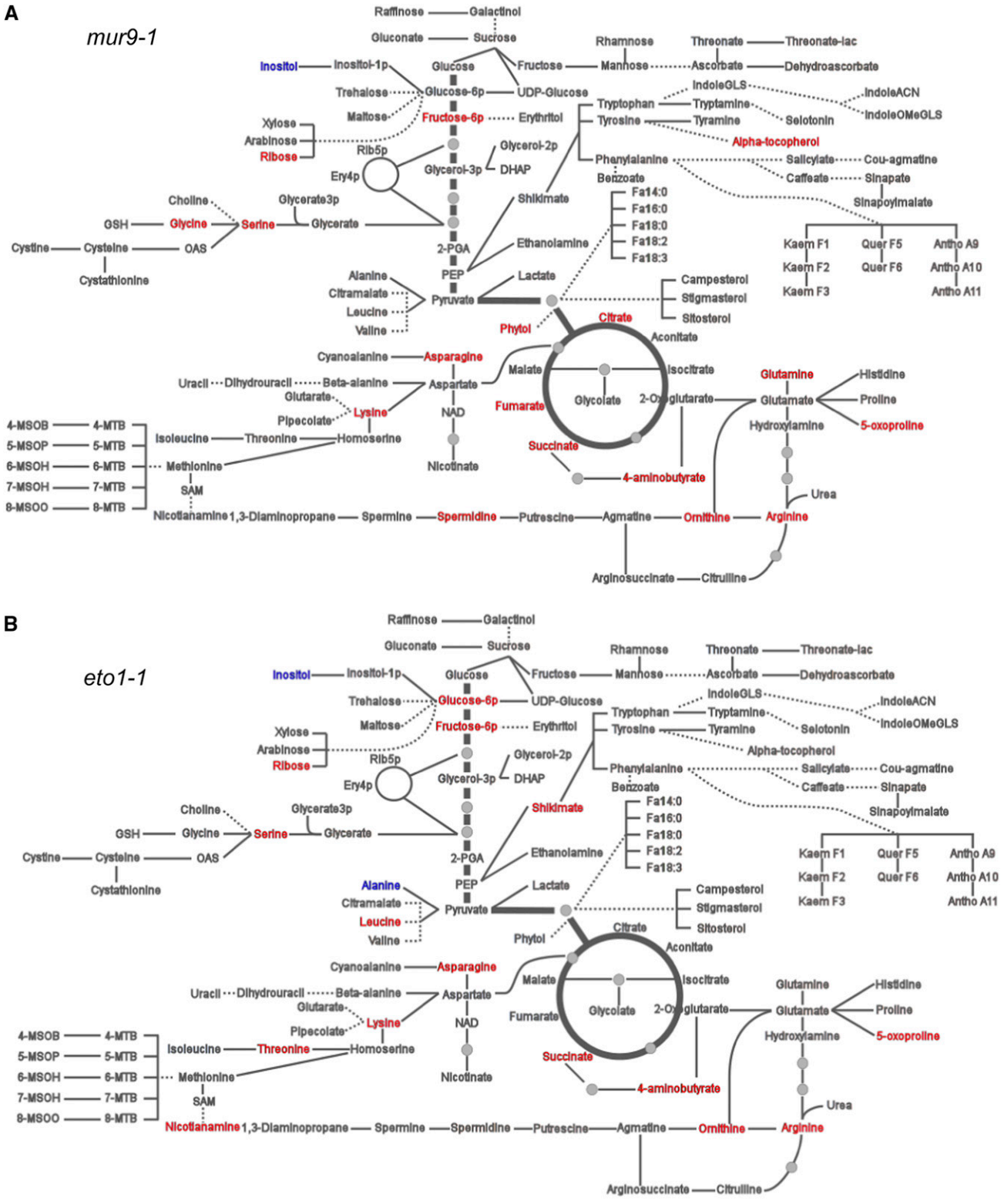

Figure 5. Changes in the metabolite levels in *mur9-1* (A) and *eto1-1* (B) plants compared with the wild type. Differences in metabolite abundance were calculated by dividing the metabolite level in the mutant by the level in the Col-0 wild type and applying a LIMMA statistical test (Smyth, 2004). The level of significance was set at FDR < 0.05 using the procedure proposed by Benjamini and Hochberg (1995). Blue and red fonts represent decreases and increases, respectively, as compared with the Col-0 wild type ($|\log_2$ fold change$| \geq 1$ and FDR < 0.05).

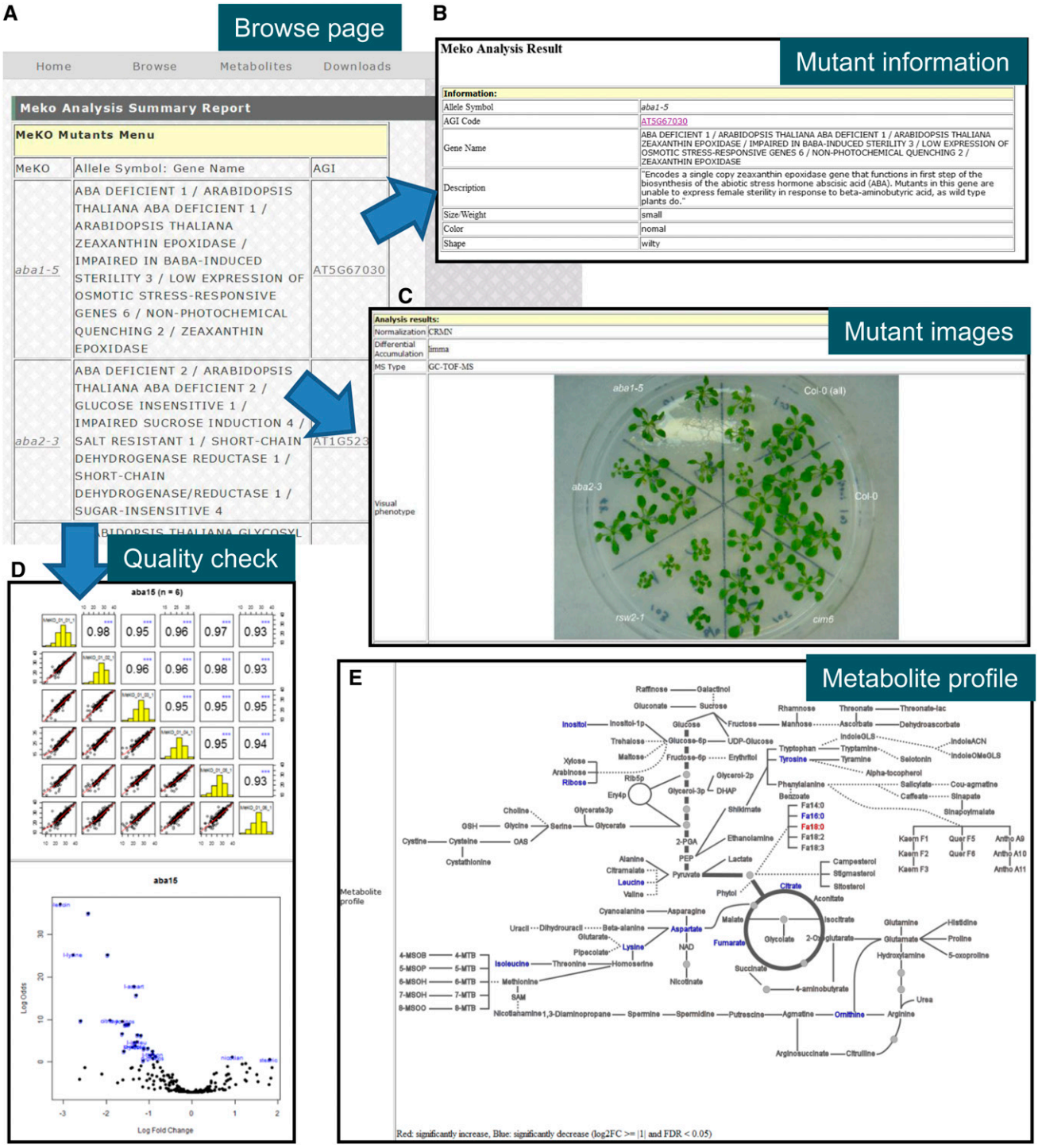

Figure 6. Screen shots of MeKO pages illustrating information on selected mutants. A, Main page of the mutants analyzed in this study. B, List of entries for the mutants selected from A, including the mutant name, AGI code, gene name, and description. C, Image of mutants and their respective wild-type controls before harvest. D, Results of replicate quality checks and volcano plot. The 30 metabolites with the most significant changes in level are highlighted in blue. E, Changes in metabolite levels of a mutant selected from A. Blue and red fonts represent decreases and increases, respectively, as compared with the Col-0 wild type ($|\log_2$ fold change$| \geq 1$ and FDR < 0.05).

Table II. *Links to pages containing resources for compounds, mass spectra, metabolic pathways, and gene description*

Database Name	URL
PubChem	http://pubchem.ncbi.nlm.nih.gov/
Chemical Entities of Biological Interest	http://www.ebi.ac.uk/chebi/
Kyoto Encyclopedia of Genes and Genomes	http://www.genome.jp/kegg/
AraCyc	http://www.arabidopsis.org/biocyc/
LipidMaps	http://www.lipidmaps.org/
LipidBank	http://lipidbank.jp/
Golm Metabolome Database	http://gmd.mpimp-golm.mpg.de/
KNApSAcK	http://kanaya.aist-nara.ac.jp/KNApSAcK/
The Arabidopsis Information Resource	http://www.arabidopsis.org/

metabolome databases, including the Golm Metabolome Database (Kopka et al., 2005), LipidBank (Watanabe et al., 2000), and LipidMaps (Fahy et al., 2007). Metabolites associated with metabolic pathways are also linked to other pathway databases, such as the Kyoto Encyclopedia of Genes and Genomes (Kanehisa et al., 2012) and AraCyc (Zhang et al., 2005).

Analysis Page

Figure 7 shows how our interactive analysis tools (e.g. hierarchical cluster analysis and a pseudocolor heat map) facilitate the visualization of MeKO data. Users can (i) choose our MeKO data set, (ii) download the results, (iii) edit parameters (e.g. normalization methods), (iv) import their own data, (v) select parameters for multivariate analysis (e.g. distance metrics in hierarchical cluster analysis), (vi) visualize data, (vii) select a data source, and (viii) perform metaanalysis. This page shows novel aspects of our study, such as user interactive data analyses, including multivariate statistics and concatenation/summarization of multiple data sets from different analytical platforms (Redestig et al., 2010).

Download Page

Nonprocessed data, including chromatograms (in netCDF format), extracted mass spectra (in NIST format), and experimental metadata, are MSI compatible and freely downloadable (Table III). All raw data in netCDF are also available from DropMet (http://prime.psc.riken.jp/?action=drop_index), MetaboLights (Salek et al., 2013; accession no. MTBLS47), and The MetabolomeExpress (Carroll et al., 2010). Users can also see our data in the PMR (Hur et al., 2013). The URLs for the All data and the five batches are as follows: for All, http://www.metnetdb.org/PMR/experiments/?expid=208; for batch 1, http://www.metnetdb.org/PMR/experiments/?expid=209; for batch 2, http://www.metnetdb.org/PMR/experiments/?expid=210; for batch 3, http://www.metnetdb.org/PMR/experiments/?expid=211; for batch 4, http://www.metnetdb.org/PMR/experiments/?expid=212; and for batch 5, http://www.metnetdb.org/PMR/experiments/?expid=213.

DISCUSSION

Our aims were to (1) characterize the metabotype of a genetic perturbation of plant metabolism and (2) develop data resources for generating testable hypotheses on the functions of genes of interest, such as uncharacterized Arabidopsis mutants. The MeKO presented here consists of integrated genotype, metabotype, and phenotype data and yields consistent information from raw data to pathway mapping of detected metabolites using interactive tools for data analysis (Fig. 1). MeKO allows viewing and browsing of profile data, of morphological image data of the mutants and the control Col-0, of the results of statistical data analyses, and of annotated metabolites of 50 Arabidopsis mutants with links to other relevant databases. These data are accessible to all researchers interested in metabolomics.

High-throughput data typically contain nonbiological variations. As the sample quality is affected by the day of processing and differences in the staff and laboratories, so-called batch effects are produced (Leek et al., 2010; De Livera et al., 2012; Lazar et al., 2013). Because we included more than 650 samples in our study, we could demonstrate that removing such batch effects from our data set by appropriate experimental designs and by quality control samples allows comparison between mutants (Fig. 2). Our experimental design and data processing highlight the importance of reducing nonbiological variations for large-scale comparative metabolomics. The COMBAT algorithm (Johnson et al., 2007) may not be the best choice for adjusting batch effects in metabolomics data sets (Chen et al., 2011). A number of batch adjustment methods are available (De Livera et al., 2012; Lazar et al., 2013), and we are in the process of comparing these methods for their enhancement of the interpretation of metabolomics data sets.

Based on a comparison of our data set and the PlantMetabolomics.org data set (Bais et al., 2010, 2012), we report at least partial mutual data complementarity (comparing only GC-MS metabolomics; Fig. 3A). For the interlaboratory comparison of metabolite profiles of *pad4-1*, the mutant that was analyzed by GC-MS in common, both data sets exhibited very high reproducibility

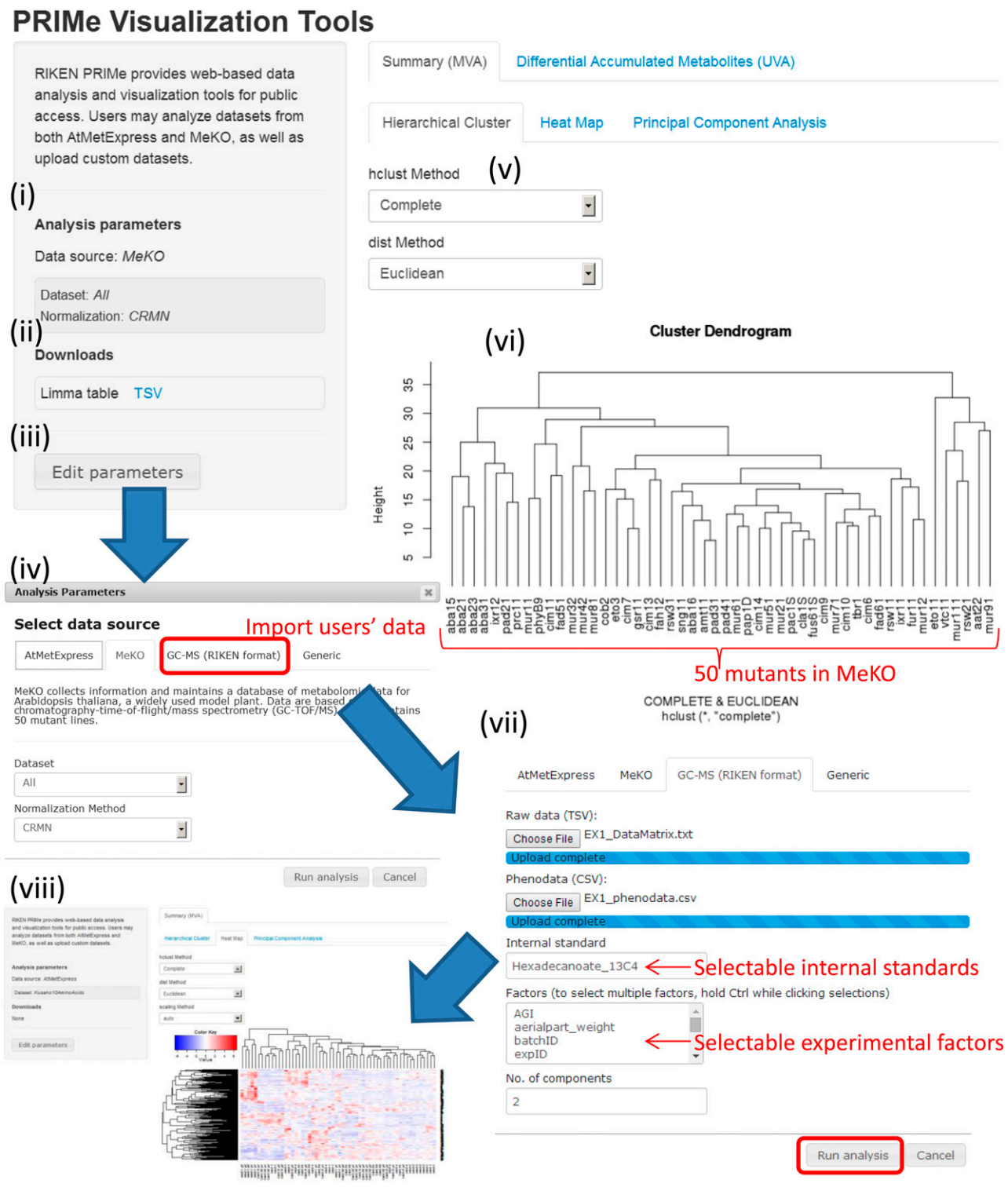

Figure 7. Visualization of MeKO and user data using our interactive analysis tools (e.g. hierarchical cluster analysis and a pseudocolor heat map). Elements are as follows: (i) data set name; (ii) downloadable results; (iii) editing parameters; (iv) importing user data; (v) selectable parameters for multivariate analysis; (vi) visualization; (vii) selecting the data source; and (viii) metaanalysis.

between biological replicates in intralaboratory comparisons (Fig. 3B). Given that analytical variations in the metabolite levels obtained from repeated measurements of individual Arabidopsis plants amounted to approximately 20% (Morgenthal et al., 2006) and that the variations can be reduced with batch normalization,

Table III. *List of downloadable data in MeKO*

All files are downloadable from DropMet (http://prime.psc.riken.jp/?action=drop_index), MetaboLights (Salek et al., 2013; http://www.ebi.ac.uk/metabolights/), and MetabolomeExpress.org (Carroll et al., 2010; https://www.metabolome-express.org/).

Data	File Format	Notes
Chromatogram data	netCDF (*.netcdf)	From LECO GC-TOF-MS
Mass spectra data	NIST (*.MSP)	Extracted in this study
Processed data matrix	CSV (*.csv)	Cross-contribution compensating multiple standard normalization (CCMN)-normalized data with COMBAT
Nonprocessed data matrix	CSV (*.csv)	Without any normalizations
Phenodata file	CSV (*.csv)	Computer-readable description file of our experimental setup
Experimental metadata	PDF (*.pdf)	MSI-compliant description of our methods
Mutant information	XLSX (*.xlsx)	Mutant lines analyzed in this study

the remaining variations may represent intrinsic metabolic fluctuations within the individual plants and/or altered metabolic behaviors that are influenced by differences in the environment. The correlation (r = approximately 0.3) between samples in MeKO and PlantMetabolomics.org probably reflects the effects of different developmental stages, ages, light conditions, and sampling days (Massonnet et al., 2010). Our results imply that for interlaboratory comparisons of normalized responses in the metabolite profiles from complex plant samples, care must be taken in the experimental design, especially with respect to the use of appropriate quality control samples, experimental setup, and the execution of the experiments.

The identification of metabolomic behaviors under abiotic stresses such as UV irradiation, cold, and drought may help us understand differences in the responses to such stresses. Our findings emphasize that in conducting metaanalyses of publicly available metabolomics data from different laboratories, careful processing of experimental metadata and the use of NIST Standard Reference Material (SRM 1950, Metabolites in Human Plasma; Simón-Manso et al., 2013) are necessary. This may facilitate the sharing of similar metabolomics data sets among laboratories.

We found that two mutants, *mur9-1* and *eto1-1*, exhibited similar metabotypes (Figs. 4 and 5). Both plants showed dramatic increases in succinate, γ-aminobutyric acid, and polyamine metabolites. More than 15 years ago, Reiter et al. (1997) used biochemical screening to isolate *mur* mutants that manifested an altered monosaccharide composition. For example, in *mur9-1*, Fuc and Xyl were decreased (Reiter et al., 1997; Reiter, 2008). However, the genes corresponding to this locus remain to be identified. Next-generation mapping developed by Austin et al. (2011) strongly suggested that *MUR11* and *SUPPRESSOR OF ACTIN9* (At3g59770) were the same gene. Studies on the metabolic coordination between cell wall synthesis and primary metabolism in plants have focused on different nucleotide sugars (Reiter, 2008). A close link between reduced tricarboxylic acid cycle activity and the inhibition of secondary cell wall synthesis in transgenic tomato (*Solanum lycopersicum*) roots has also been suggested (van der Merwe et al., 2010). The evident relationship in the aerial parts of Arabidopsis remains largely uncharacterized, and experimental validation of this hypothesis is beyond the scope of our study. Nonetheless, an unsupervised approach using mutants with cellulose defects in our data groups may be useful for the further characterization of *mur* mutants. Our report documents that MeKO is highly informative for the generation of testable hypotheses regarding plant functional genomics.

MeKO will continue to grow with the (1) analysis of other mutants, (2) consideration of genetic and environmental perturbations, and (3) full use of various ecotypes. Increasing the extent of metabolomic detection is important for a complete understanding of gene functions. The combined use of multiple analytical instruments, such as LC-MS, will facilitate the detection of more diverse metabolites. Like the PMR (Hur et al., 2013), MeKO with full metabolomic annotation is a promising functional genomics tool, although the generation of such data sets is laborious. We expect our data sets to significantly accelerate the development and improvement of informatics tools, making them more useful for the study of metabolomics.

CONCLUSION

We have presented the comprehensive metabolite profiles of 50 Arabidopsis mutants, including a set of characterized and uncharacterized mutants. We detected previously hidden alterations in the production of metabolites, particularly in primary metabolism. We expect our data to help improve our understanding of gene function as well as plant growth and development. We developed MeKO to use the data set as a functional genomics tool. Metabolites whose accumulation differs from that of wild-type plants are conveniently displayed in a metabolic map that includes relational information. Our analyses not only revealed the metabotype of each mutant, but MeKO also can be used for the generation of testable hypotheses concerning the functions of genes of interest and is highly useful for improving gene annotation.

MATERIALS AND METHODS

Plant Materials

Details on Arabidopsis (*Arabidopsis thaliana*) growth and harvest are found in Supplemental Text S1.

Metabolite Profiling

Detailed information on extraction, MS conditions, and GC-MS data processing is provided in the Metabolomics Metadata section in Supplemental Text S1. The All raw data are also available in MetaboLights (Salek et al., 2013; accession no. MTBLS47) and The MetabolomeExpress (Carroll et al., 2010).

Normalization and Data Preprocessing

Data preprocessing, such as baseline correction, peak alignment, and peak deconvolution, was as reported by Kusano et al. (2007). In brief, most processes were based on hierarchical multivariate curve resolution (Jonsson et al., 2004, 2006) and an in-house mass spectral search program. Detailed information is included in Supplemental Text S1. We subjected more than 650 samples, including 629 plants and 27 quality control samples, to GC-MS analysis. Because these data sets were composed of a large number of samples, we analyzed them on 5 different days. After careful preprocessing and peak annotation, we constructed a single data matrix consisting of five data sets (the All data set) and separately analyzed each data set (Table I). We prepared mixed extracts of each sample to reduce unavoidable batch effects (i.e. the nonbiological effects of different MS conditions, such as detector sensitivity, detector responses, and column performance) among the data sets. We corrected the variations using the COMBAT algorithm (Johnson et al., 2007) based on the empirical Bayes approach. We then applied CCMN normalization (Redestig et al., 2009) to batch-corrected data to eliminate cross-contribution problems from the data set. Metabolite identifiers were integrated using MetMask (Redestig et al., 2010), a conversion tool for metabolite identifiers.

Statistical Data Analysis

We utilized the BiNGO program (Maere et al., 2005), a software to evaluate overrepresentation or underrepresentation in a set of genes, for the analysis of significantly overrepresented GO categories. To evaluate batch effects, PCA score scatterplots and ordinary scatterplots were constructed using the pcaMethods package (Stacklies et al., 2007). A correlation heat map, a scatterplot for replicate quality checks, and a volcano plot were constructed using statistical R software (http://cran.r-project.org/). Differences in the metabolite levels were detected using the LIMMA package (Smyth, 2004) in R with the FDR procedure (Benjamini and Hochberg, 1995). We also calculated MDS with Euclidean distance as implemented in R. This calculation was based on the $\log_2$ ratio of the samples to the Col-0 wild-type samples in each data set. For mapping the metabolite levels onto the diagram of the metabolic pathways, the map file in SVG format was applied manually. The final mapped files were generated with an in-house Ruby script. Changes in the metabolite levels were calculated by dividing the metabolite abundance in the mutant by the abundance in the Col-0 wild-type plants; they are represented on a pseudocolor scale where red and blue indicate increases and decreases, respectively. For constructing our interactive analysis tools, we used the Shiny package in R (http://www.rstudio.com/shiny/).

Supplemental Data

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

Supplemental Table S1. List of the mutant lines analyzed in this study.

Supplemental Table S2. List of all metabolites detected in this study.

Supplemental Text S1. Metabolomics metadata.

ACKNOWLEDGMENTS

We thank Drs. Thomas Moritz, Pär Jonsson, and Hans Stenlund (Umeå Plant Science Centre and Umeå University) for assistance with hyphenated data analysis and raw data analysis, Koji Takano and Miho Misawa (RIKEN Center for Sustainable Resource Science) for preparing the standard compounds, Ursula Petralia for editorial assistance, and Dr. Tetsuya Sakurai and Yutaka Yamada (RIKEN Center for Sustainable Resource Science) for computational assistance.

Received April 8, 2014; accepted May 5, 2014; published May 14, 2014.

LITERATURE CITED

Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Altaf-Ul-Amin M, Darusman LK, et al (2012) KNApSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol **53:** e1

Allwood JW, De Vos RC, Moing A, Deborde C, Erban A, Kopka J, Goodacre R, Hall RD (2011) Plant metabolomics and its potential for systems biology research background concepts, technology, and methodology. Methods Enzymol **500:** 299–336

Amiour N, Imbaud S, Clément G, Agier N, Zivy M, Valot B, Balliau T, Armengaud P, Quilleré I, Cañas R, et al (2012) The use of metabolomics integrated with transcriptomic and proteomic studies for identifying key steps involved in the control of nitrogen metabolism in crops such as maize. J Exp Bot **63:** 5017–5033

Austin RS, Vidaurre D, Stamatiou G, Breit R, Provart NJ, Bonetta D, Zhang J, Fung P, Gong Y, Wang PW, et al (2011) Next-generation mapping of Arabidopsis genes. Plant J **67:** 715–725

Bais P, Moon SM, He K, Leitao R, Dreher K, Walk T, Sucaet Y, Barkan L, Wohlgemuth G, Roth MR, et al (2010) PlantMetabolomics.org: a Web portal for plant metabolomics experiments. Plant Physiol **152:** 1807–1816

Bais P, Moon-Quanbeck SM, Nikolau BJ, Dickerson JA (2012) Plantmetabolomics.org: mass spectrometry-based Arabidopsis metabolomics—database and tools update. Nucleic Acids Res **40:** D1216–D1220

Bell SM, Burgoon LD, Last RL (2012) MIPHENO: data normalization for high throughput metabolite analysis. BMC Bioinformatics **13:** 10

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B **57:** 289–300

Caldana C, Degenkolbe T, Cuadros-Inostroza A, Klie S, Sulpice R, Leisse A, Steinhauser D, Fernie AR, Willmitzer L, Hannah MA (2011) High-density kinetic analysis of the metabolomic and transcriptomic response of Arabidopsis to eight environmental conditions. Plant J **67:** 869–884

Carreno-Quintero N, Acharjee A, Maliepaard C, Bachem CW, Mumm R, Bouwmeester H, Visser RG, Keurentjes JJ (2012) Untargeted metabolic quantitative trait loci analyses reveal a relationship between primary metabolism and potato tuber quality. Plant Physiol **158:** 1306–1318

Carroll AJ, Badger MR, Harvey Millar A (2010) The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets. BMC Bioinformatics **11:** 376

Chen C, Grennan K, Badner J, Zhang D, Gershon E, Jin L, Liu C (2011) Removing batch effects in analysis of expression microarray data: an evaluation of six batch adjustment methods. PLoS ONE **6:** e17238

Degtyarenko K, de Matos P, Ennis M, Hastings J, Zbinden M, McNaught A, Alcántara R, Darsow M, Guedj M, Ashburner M (2008) ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res **36:** D344–D350

De Livera AM, Dias DA, De Souza D, Rupasinghe T, Pyke J, Tull D, Roessner U, McConville M, Speed TP (2012) Normalizing and integrating metabolomics data. Anal Chem **84:** 10768–10776

Espinoza C, Degenkolbe T, Caldana C, Zuther E, Leisse A, Willmitzer L, Hincha DK, Hannah MA (2010) Interaction with diurnal and circadian regulation results in dynamic metabolic and transcriptional changes during cold acclimation in Arabidopsis. PLoS ONE **5:** e14101

Fahy E, Sud M, Cotter D, Subramaniam S (2007) LIPID MAPS online tools for lipid research. Nucleic Acids Res **35:** W606–W612

Fernie AR, Aharoni A, Willmitzer L, Stitt M, Tohge T, Kopka J, Carroll AJ, Saito K, Fraser PD, DeLuca V (2011) Recommendations for reporting metabolite data. Plant Cell **23:** 2477–2482

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

Fukushima A, Kusano M (2013) Recent progress in the development of metabolome databases for plant systems biology. Front Plant Sci **4:** 73

Fukushima A, Kusano M, Nakamichi N, Kobayashi M, Hayashi N, Sakakibara H, Mizuno T, Saito K (2009a) Impact of clock-associated Arabidopsis pseudo-response regulators in metabolic coordination. Proc Natl Acad Sci USA **106:** 7251–7256

Fukushima A, Kusano M, Redestig H, Arita M, Saito K (2009b) Integrated omics approaches in plant systems biology. Curr Opin Chem Biol **13:** 532–538

Gibon Y, Usadel B, Blaesing OE, Kamlage B, Hoehne M, Trethewey R, Stitt M (2006) Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. Genome Biol **7:** R76

Gu L, Jones AD, Last RL (2007) LC-MS/MS assay for protein amino acids and metabolically related compounds for large-scale screening of metabolic phenotypes. Anal Chem **79:** 8067–8075

Hoffman DE, Jonsson P, Bylesjö M, Trygg J, Antti H, Eriksson ME, Moritz T (2010) Changes in diurnal patterns within the Populus transcriptome and metabolome in response to photoperiod variation. Plant Cell Environ **33:** 1298–1313

Hur M, Campbell AA, Almeida-de-Macedo M, Li L, Ransom N, Jose A, Crispin M, Nikolau BJ, Wurtele ES (2013) A global approach to analysis and interpretation of metabolic data for plant natural product discovery. Nat Prod Rep **30:** 565–583

Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics **8:** 118–127

Jonsson P, Gullberg J, Nordström A, Kusano M, Kowalczyk M, Sjöström M, Moritz T (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. Anal Chem **76:** 1738–1745

Jonsson P, Johansson ES, Wuolikainen A, Lindberg J, Schuppe-Koistinen I, Kusano M, Sjöström M, Trygg J, Moritz T, Antti H (2006) Predictive metabolite profiling applying hierarchical multivariate curve resolution to GC-MS data: a potential tool for multi-parametric diagnosis. J Proteome Res **5:** 1407–1414

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res **40:** D109–D114

Kaplan F, Kopka J, Haskell DW, Zhao W, Schiller KC, Gatzke N, Sung DY, Guy CL (2004) Exploring the temperature-stress metabolome of Arabidopsis. Plant Physiol **136:** 4159–4168

Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmüller E, Dörmann P, Weckwerth W, Gibon Y, Stitt M, et al (2005) GMD@CSB. DB: the Golm Metabolome Database. Bioinformatics **21:** 1635–1638

Kusano M, Fukushima A, Arita M, Jonsson P, Moritz T, Kobayashi M, Hayashi N, Tohge T, Saito K (2007) Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in Arabidopsis thaliana. BMC Syst Biol **1:** 53

Kusano M, Tabuchi M, Fukushima A, Funayama K, Diaz C, Kobayashi M, Hayashi N, Tsuchiya YN, Takahashi H, Kamata A, et al (2011a) Metabolomics data reveal a crucial role of cytosolic glutamine synthetase 1;1 in coordinating metabolic balance in rice. Plant J **66:** 456–466

Kusano M, Tohge T, Fukushima A, Kobayashi M, Hayashi N, Otsuki H, Kondou Y, Goto H, Kawashima M, Matsuda F, et al (2011b) Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light. Plant J **67:** 354–369

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, et al (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res **40:** D1202–D1210

Lazar C, Meganck S, Taminau J, Steenhoff D, Coletta A, Molter C, Weiss-Solis DY, Duque R, Bersini H, Nowe A (2013) Batch effect removal methods for microarray gene expression data integration: a survey. Brief Bioinform **14:** 469–490

Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, Geman D, Baggerly K, Irizarry RA (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. Nat Rev Genet **11:** 733–739

Lei Z, Huhman DV, Sumner LW (2011) Mass spectrometry strategies in metabolomics. J Biol Chem **286:** 25435–25442

Lu Y, Savage LJ, Larson MD, Wilkerson CG, Last RL (2011) Chloroplast 2010: a database for large-scale phenotypic screening of Arabidopsis mutants. Plant Physiol **155:** 1589–1600

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in biological networks. Bioinformatics **21:** 3448–3449

Maruyama K, Urano K, Yoshiwara K, Morishita Y, Sakurai N, Suzuki H, Kojima M, Sakakibara H, Shibata D, Saito K, et al (2014) Integrated analysis of the effects of cold and dehydration on rice metabolites, phytohormones, and gene transcripts. Plant Physiol **164:** 1759–1771

Massonnet C, Vile D, Fabre J, Hannah MA, Caldana C, Lisec J, Beemster GT, Meyer RC, Messerli G, Gronlund JT, et al (2010) Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three Arabidopsis accessions cultivated in ten laboratories. Plant Physiol **152:** 2142–2157

Matsuda F, Okazaki Y, Oikawa A, Kusano M, Nakabayashi R, Kikuchi J, Yonemaru J, Ebana K, Yano M, Saito K (2012) Dissection of genotype-phenotype associations in rice grains using metabolome quantitative trait loci analysis. Plant J **70:** 624–636

Morgenthal K, Weckwerth W, Steuer R (2006) Metabolomic networks in plants: transitions from pattern recognition to biological interpretation. Biosystems **83:** 108–117

Morreel K, Goeminne G, Storme V, Sterck L, Ralph J, Coppieters W, Breyne P, Steenackers M, Georges M, Messens E, et al (2006) Genetical metabolomics of flavonoid biosynthesis in Populus: a case study. Plant J **47:** 224–237

Nakabayashi R, Yonekura-Sakakibara K, Urano K, Suzuki M, Yamada Y, Nishizawa T, Matsuda F, Kojima M, Sakakibara H, Shinozaki K, et al (2014) Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. Plant J **77:** 367–379

Pracharoenwattana I, Zhou W, Keech O, Francisco PB, Udomchalothorn T, Tschoep H, Stitt M, Gibon Y, Smith SM (2010) Arabidopsis has a cytosolic fumarase required for the massive allocation of photosynthate into fumaric acid and for rapid plant growth on high nitrogen. Plant J **62:** 785–795

Quanbeck SM, Brachova L, Campbell AA, Guan X, Perera A, He K, Rhee SY, Bais P, Dickerson JA, Dixon P, et al (2012) Metabolomics as a hypothesis-generating functional genomics tool for the annotation of Arabidopsis thaliana genes of "unknown function." Front Plant Sci **3:** 15

Redestig H, Fukushima A, Stenlund H, Moritz T, Arita M, Saito K, Kusano M (2009) Compensation for systematic cross-contribution improves normalization of mass spectrometry based metabolomics data. Anal Chem **81:** 7974–7980

Redestig H, Kusano M, Fukushima A, Matsuda F, Saito K, Arita M (2010) Consolidating metabolite identifiers to enable contextual and multi-platform metabolomics data analysis. BMC Bioinformatics **11:** 214

Reiter WD (2008) Biochemical genetics of nucleotide sugar interconversion reactions. Curr Opin Plant Biol **11:** 236–243

Reiter WD, Chapple C, Somerville CR (1997) Mutants of Arabidopsis thaliana with altered cell wall polysaccharide composition. Plant J **12:** 335–345

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

Rubin G, Tohge T, Matsuda F, Saito K, Scheible WR (2009) Members of the LBD family of transcription factors repress anthocyanin synthesis and affect additional nitrogen responses in *Arabidopsis*. Plant Cell **21:** 3567–3584

Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. Annu Rev Plant Biol **61:** 463–489

Salek RM, Haug K, Conesa P, Hastings J, Williams M, Mahendraker T, Maguire E, González-Beltrán AN, Rocca-Serra P, Sansone SA, et al (2013) The MetaboLights repository: curation challenges in metabolomics. Database (Oxford) **2013:** bat029

Schauer N, Semel Y, Roessner U, Gur A, Balbo I, Carrari F, Pleban T, Perez-Melis A, Bruedigam C, Kopka J, et al (2006) Comprehensive metabolic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat Biotechnol **24:** 447–454

Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K, Roessner-Tunali U, Forbes MG, Willmitzer L, et al (2005) GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett **579:** 1332–1337

Simón-Manso Y, Lowenthal MS, Kilpatrick LE, Sampson ML, Telu KH, Rudnick PA, Mallard WG, Bearden DW, Schock TB, Tchekhovskoi DV, et al (2013) Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. Anal Chem **85:** 11725–11731

Smyth GK (2004) Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol **3:** Article 3

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods: a Bioconductor package providing PCA methods for incomplete data. Bioinformatics **23:** 1164–1167

Stitt M, Fernie AR (2003) From measurements of metabolites to metabolomics: an 'on the fly' perspective illustrated by recent studies of carbon-nitrogen interactions. Curr Opin Biotechnol **14:** 136–144

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: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics **3:** 211–221

Tohge T, Fernie AR (2009) Web-based resources for mass-spectrometry-based metabolomics: a user's guide. Phytochemistry **70:** 450–456

Tohge T, Nishiyama Y, Hirai MY, Yano M, Nakajima J, Awazuhara M, Inoue E, Takahashi H, Goodenowe DB, Kitayama M, et al (2005) Functional genomics by integrated analysis of metabolome and transcriptome of Arabidopsis plants over-expressing an MYB transcription factor. Plant J **42:** 218–235

Urano K, Maruyama K, Ogata Y, Morishita Y, Takeda M, Sakurai N, Suzuki H, Saito K, Shibata D, Kobayashi M, et al (2009) Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. Plant J **57:** 1065–1078

Urbanczyk-Wochniak E, Baxter C, Kolbe A, Kopka J, Sweetlove LJ, Fernie AR (2005) Profiling of diurnal patterns of metabolite and transcript abundance in potato (Solanum tuberosum) leaves. Planta **221:** 891–903

van der Merwe MJ, Osorio S, Araújo WL, Balbo I, Nunes-Nesi A, Maximova E, Carrari F, Bunik VI, Persson S, Fernie AR (2010) Tricarboxylic acid cycle activity regulates tomato root growth via effects on secondary cell wall production. Plant Physiol **153:** 611–621

Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH (2009) PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res **37:** W623–W633

Watanabe K, Yasugi E, Oshima M (2000) How to search the glycolipid data in LIPIDBANK for Web: the newly developed lipid database in Japan. Trends Glycosci Glycotechnol **12:** 175–184

Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. Proc Natl Acad Sci USA **101:** 7809–7814

Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, Rhee SY (2005) MetaCyc and AraCyc: metabolic pathway databases for plant research. Plant Physiol **138:** 27–37