Patterns of Metabolite Changes Identified from Large-Scale Gene Perturbations in Arabidopsis Using a Genome-Scale Metabolic Network[1][OPEN]

Taehyong Kim, Kate Dreher, Ricardo Nilo-Poyanco, Insuk Lee, Oliver Fiehn, Bernd Markus Lange, Basil J. Nikolau, Lloyd Sumner, Ruth Welti, Eve S. Wurtele, and Seung Y. Rhee*

Department of Plant Biology, Carnegie Institution for Science, Stanford, California 94305 (T.K., K.D., R.N.-P., S.Y.R.); Department of Biotechnology, Yonsei University, Seoul 120–749, South Korea (I.L.); Genome Center, University of California, Davis, California 95616 (O.F.); M. J. Murdock Metabolomics Laboratory, Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164 (B.M.L.); Center for Metabolic Biology, Department of Biochemistry, Biophysics, and Molecular Biology (B.J.N.), and Department of Genetics, Development, and Cell Biology (E.S.W.), Iowa State University, Ames, Iowa 50011; Plant Biology Division, The Samuel Roberts Noble Foundation, Ardmore, Oklahoma 73401 (L.S.); and Division of Biology, Kansas State University, Manhattan, Kansas 66506 (R.W.)

Metabolomics enables quantitative evaluation of metabolic changes caused by genetic or environmental perturbations. However, little is known about how perturbing a single gene changes the metabolic system as a whole and which network and functional properties are involved in this response. To answer this question, we investigated the metabolite profiles from 136 mutants with single gene perturbations of functionally diverse Arabidopsis (*Arabidopsis thaliana*) genes. Fewer than 10 metabolites were changed significantly relative to the wild type in most of the mutants, indicating that the metabolic network was robust to perturbations of single metabolic genes. These changed metabolites were closer to each other in a genome-scale metabolic network than expected by chance, supporting the notion that the genetic perturbations changed the network more locally than globally. Surprisingly, the changed metabolites were close to the perturbed reactions in only 30% of the mutants of the well-characterized genes. To determine the factors that contributed to the distance between the observed metabolic changes and the perturbation site in the network, we examined nine network and functional properties of the perturbed genes. Only the isozyme number affected the distance between the perturbed reactions and changed metabolites. This study revealed patterns of metabolic changes from large-scale gene perturbations and relationships between characteristics of the perturbed genes and metabolic changes.

Rational and quantitative assessment of metabolic changes in response to genetic modification (GM) is an open question and in need of innovative solutions. Nontargeted metabolite profiling can detect thousands of compounds, but it is not easy to understand the significance of the changed metabolites in the biochemical and biological context of the organism. To better assess the changes in metabolites from nontargeted metabolomics studies, it is important to examine the changed metabolites in the context of the genome-scale metabolic network of the organism.

Metabolomics is a technique that aims to quantify all the metabolites in a biological system (Nikolau and Wurtele, 2007; Nicholson and Lindon, 2008; Roessner and Bowne, 2009). It has been used widely in studies ranging from disease diagnosis (Holmes et al., 2008; DeBerardinis and Thompson, 2012) and drug discovery (Cascante et al., 2002; Kell, 2006) to metabolic reconstruction (Feist et al., 2009; Kim et al., 2012) and metabolic engineering (Keasling, 2010; Lee et al., 2011). Metabolomic studies have demonstrated the possibility of identifying gene functions from changes in the relative concentrations of metabolites (metabotypes or metabolic signatures; Ebbels et al., 2004) in various species including yeast (*Saccharomyces cerevisiae*; Raamsdonk et al., 2001; Allen et al., 2003), Arabidopsis (*Arabidopsis thaliana*; Brotman et al., 2011), tomato (*Solanum lycopersicum*; Schauer et al., 2006), and maize (*Zea mays*; Riedelsheimer et al., 2012). Metabolomics has also been used to better understand how plants interact with their environments (Field and Lake, 2011), including their responses to biotic and abiotic stresses (Dixon et al., 2006; Arbona et al., 2013), and to predict important agronomic traits (Riedelsheimer et al., 2012). Metabolite profiling has been performed on many plant species, including angiosperms such as Arabidopsis, poplar (*Populus trichocarpa*), and *Catharanthus roseus* (Sumner et al., 2003; Rischer et al., 2006), basal land

[1] This work was supported in part by the National Science Foundation (grant nos. MCB–0820823, IOS–1026003, and DBI–0640769), the Department of Energy (grant no. BER–65472), and a postdoctoral fellowship award from Becas Chile-Conicyt (to R.N.-P.).

* Address correspondence to srhee@carnegiescience.edu.

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: Seung Y. Rhee (srhee@carnegiescience.edu).

[OPEN] Articles can be viewed without a subscription.
www.plantphysiol.org/cgi/doi/10.1104/pp.114.252361

plants such as *Selaginella moellendorffii* and *Physcomitrella patens* (Erxleben et al., 2012; Yobi et al., 2012), and *Chlamydomonas reinhardtii* (Fernie et al., 2012; Davis et al., 2013). With the availability of whole genome sequences of various species, metabolomics has the potential to become a useful tool for elucidating the functions of genes using large-scale systematic analyses (Fiehn et al., 2000; Saito and Matsuda, 2010; Hur et al., 2013).

Although metabolomics data have the potential for identifying the roles of genes that are associated with metabolic phenotypes, the biochemical mechanisms that link functions of genes with metabolic phenotypes are still poorly characterized. For example, we do not yet know the principles behind how perturbing the expression of a single gene changes the metabolic system as a whole. Large-scale metabolomics data have provided useful resources for linking phenotypes to genotypes (Fiehn et al., 2000; Roessner et al., 2001; Tikunov et al., 2005; Schauer et al., 2006; Lu et al., 2011; Fukushima et al., 2014). For example, Lu et al. (2011) compared morphological and metabolic phenotypes from more than 5,000 Arabidopsis chloroplast mutants using gas chromatography (GC)- and liquid chromatography (LC)-mass spectrometry (MS). Fukushima et al. (2014) generated metabolite profiles from various characterized and uncharacterized mutant plants and clustered the mutants with similar metabolic phenotypes by conducting multidimensional scaling with quantified metabolic phenotypes. Nonetheless, representation and analysis of such a large amount of data remains a challenge for scientific discovery (Lu et al., 2011). In addition, these studies do not examine the topological and functional characteristics of metabolic changes in the context of a genome-scale metabolic network. To understand the relationship between genotype and metabolic phenotype, we need to investigate the metabolic changes caused by perturbing the expression of a gene in a genome-scale metabolic network perspective, because metabolic pathways are not independent biochemical factories but are components of a complex network (Berg et al., 2002; Merico et al., 2009).

Much progress has been made in the last 2 decades to represent metabolism at a genome scale (Terzer et al., 2009). The advances in genome sequencing and emerging fields such as biocuration and bioinformatics enabled the representation of genome-scale metabolic network reconstructions for model organisms (Bassel et al., 2012). Genome-scale metabolic models have been built and applied broadly from microbes to plants. The first step toward modeling a genome-scale metabolism in a plant species started with developing a genome-scale metabolic pathway database for Arabidopsis (AraCyc; Mueller et al., 2003) from reference pathway databases (Kanehisa and Goto, 2000; Karp et al., 2002; Zhang et al., 2010). Genome-scale metabolic pathway databases have been built for several plant species (Mueller et al., 2005; Zhang et al., 2005, 2010; Urbanczyk-Wochniak and Sumner, 2007; May et al., 2009; Dharmawardhana et al., 2013; Monaco et al., 2013, 2014; Van Moerkercke et al., 2013; Chae et al., 2014; Jung et al., 2014). Efforts have been made to develop predictive genome-scale metabolic models using enzyme kinetics and stoichiometric flux-balance approaches (Sweetlove et al., 2008). de Oliveira Dal'Molin et al. (2010) developed a genome-scale metabolic model for Arabidopsis and successfully validated the model by predicting the classical photorespiratory cycle as well as known key differences between redox metabolism in photosynthetic and nonphotosynthetic plant cells. Other genome-scale models have been developed for Arabidopsis (Poolman et al., 2009; Radrich et al., 2010; Mintz-Oron et al., 2012), *C. reinhardtii* (Chang et al., 2011; Dal'Molin et al., 2011), maize (Dal'Molin et al., 2010; Saha et al., 2011), sorghum (*Sorghum bicolor*; Dal'Molin et al., 2010), and sugarcane (*Saccharum officinarum*; Dal'Molin et al., 2010). These predictive models have the potential to be applied broadly in fields such as metabolic engineering, drug target discovery, identification of gene function, study of evolutionary processes, risk assessment of genetically modified crops, and interpretations of mutant phenotypes (Feist and Palsson, 2008; Ricroch et al., 2011).

Here, we interrogate the metabotypes caused by 136 single gene perturbations of Arabidopsis by analyzing the relative concentration changes of 1,348 chemically identified metabolites using a reconstructed genome-scale metabolic network. We examine the characteristics of the changed metabolites (the metabolites whose relative concentrations were significantly different in mutants relative to the wild type) in the metabolic network to uncover biological and topological consequences of the perturbed genes.

RESULTS

Comprehensive Metabolite Profiles from Single Gene Perturbations

To systematically characterize metabotypes resulting from gene perturbations, we selected 136 Arabidopsis genes across 11 functional categories of metabolism in the Arabidopsis metabolic database AraCyc (Mueller et al., 2003; Fig. 1; Supplemental Fig. S1; Supplemental Table S1). Among these, 29 are genes with experimentally demonstrated enzymatic function, while 33 have significant sequence similarity to known enzymatic genes. In addition, we selected 45 experimentally uncharacterized genes whose involvement in specific metabolic pathways was predicted computationally using the gene cofunction network AraNet ("Materials and Methods;" Lee et al., 2010). AraNet integrates functional genomics data to make probabilistic functional links between genes, which can be used to predict functions of uncharacterized genes using the guilt-by-association principle. Finally, 29 genes that could not be assigned to any functional category by existing methods were selected. We selected the genes based on the absence of a visible phenotype from the gene perturbation, the absence of homologs from recent gene duplications, and easily detected expression in 2-week-old leaves in AtGenExpress data (levels of at least 100; Supplemental Fig. S1). Functions of the well-characterized and predicted genes in this study spanned all metabolism categories (Fig. 1).

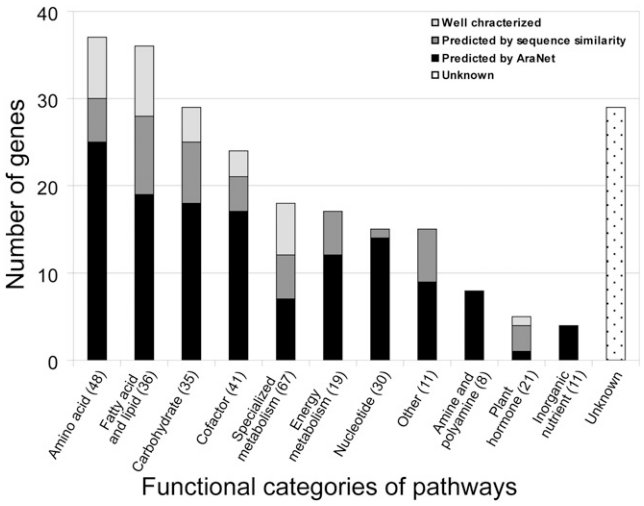

Figure 1. Distribution of the genes selected for this study across 11 functional domains of metabolism in the Arabidopsis metabolic pathway database AraCyc (Mueller et al., 2003). Different colored bars in each domain of metabolism indicate proportions of genes with different degrees of knowledge about their roles in metabolism. Genes associated with more than one functional domain were categorized to multiple domains of metabolism (Supplemental Table S1). The number in parenthesis on the *x* axis indicates the number of pathways belonging to each functional domain of metabolism.

Wild-type ecotype Columbia (Col-0) seedlings and mutant seedlings of the 136 selected genes were maintained under standard growth conditions, and metabolites were analyzed as described previously (Bais et al., 2010, 2012; Quanbeck et al., 2012). We detected 4,483 distinct metabolite peaks from the leaves of 2-week-old seedlings using 11 mass spectrometric platforms, including seven targeted and four nontargeted platforms. We identified 1,348 metabolites based on mass fragmentation patterns and chromatographic behavior (comparison with authentic standards; Fig. 2A) and classified them into nine categories based on systematic names, molecular structures, and functional roles in metabolism. Among the 1,348 chemically defined metabolites, fatty acids and lipids constituted the largest class (51%), followed by specialized metabolites (18%), carbohydrates (10%), amino acids (7%), and other less represented categories (Fig. 2B). The chemically defined metabolites were mapped to 839 metabolic reactions in AraCyc to evaluate the global metabolic changes in the metabolic network of reactions ("Materials and Methods;" Supplemental Fig. S2). The entire metabolite profiling data for all the mutants are available online (http://www.plantmetabolomics.org; Bais et al., 2010) and (http://www.metnetdb.org/PMR; Wurtele et al., 2012; Hur et al., 2013).

Specificity of Changed Metabolites across 136 Single Gene Perturbations

To determine the patterns of the changed metabolites caused by a gene perturbation, we identified metabolites whose relative concentrations were significantly different in mutants relative to the wild type among the 1,348 chemically defined metabolites (Supplemental Table S2; Supplemental Fig. S3). Over 80% (113/136) of the mutants showed fewer than 10 changed metabolites (Fig. 3A). Thirty-one mutants (23%) showed no changed metabolites. Only two mutants, SALK_015522C (AT5G36880) and SALK_069657C (AT1G32200), showed more than 50 changed metabolites. These results indicate that most single gene perturbations of metabolic genes did not rewire the metabolic network globally. The identity of changed metabolites varied among the mutants, suggesting that metabolic changes were specific to the perturbed genes (Fig. 3B). The same pattern of metabolic specificity and diversity was observed when genes with different degrees of characterization (e.g. well characterized, predicted by homology, predicted by AraNet, and unknown) were compared (Supplemental Figs. S4 and S5). Lipids and fatty acids were most commonly affected in the mutants, which may result from the prevalence of this class of metabolites in the analytical platforms in this study. The composition of the changed metabolites was independent from the number of changed metabolites

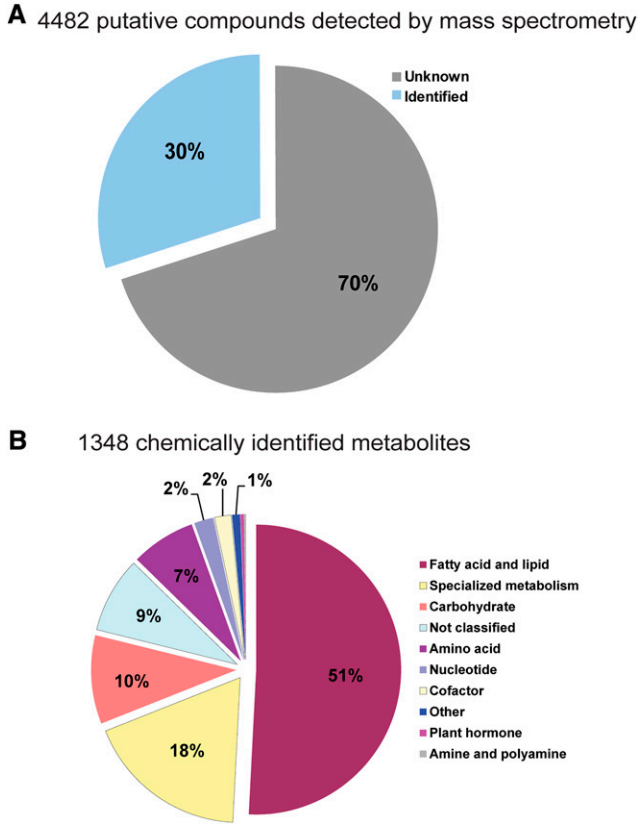

Figure 2. Distribution of metabolites detected from Arabidopsis leaves using 11 MS platforms. A, Among 4,482 putative compounds, 1,348 (30%) compounds were chemically identified and 3,134 (70%) compounds remained unnamed. B, One thousand three hundred forty-eight metabolites were categorized into nine domains of metabolism based on their name, molecular structure, and functional role.

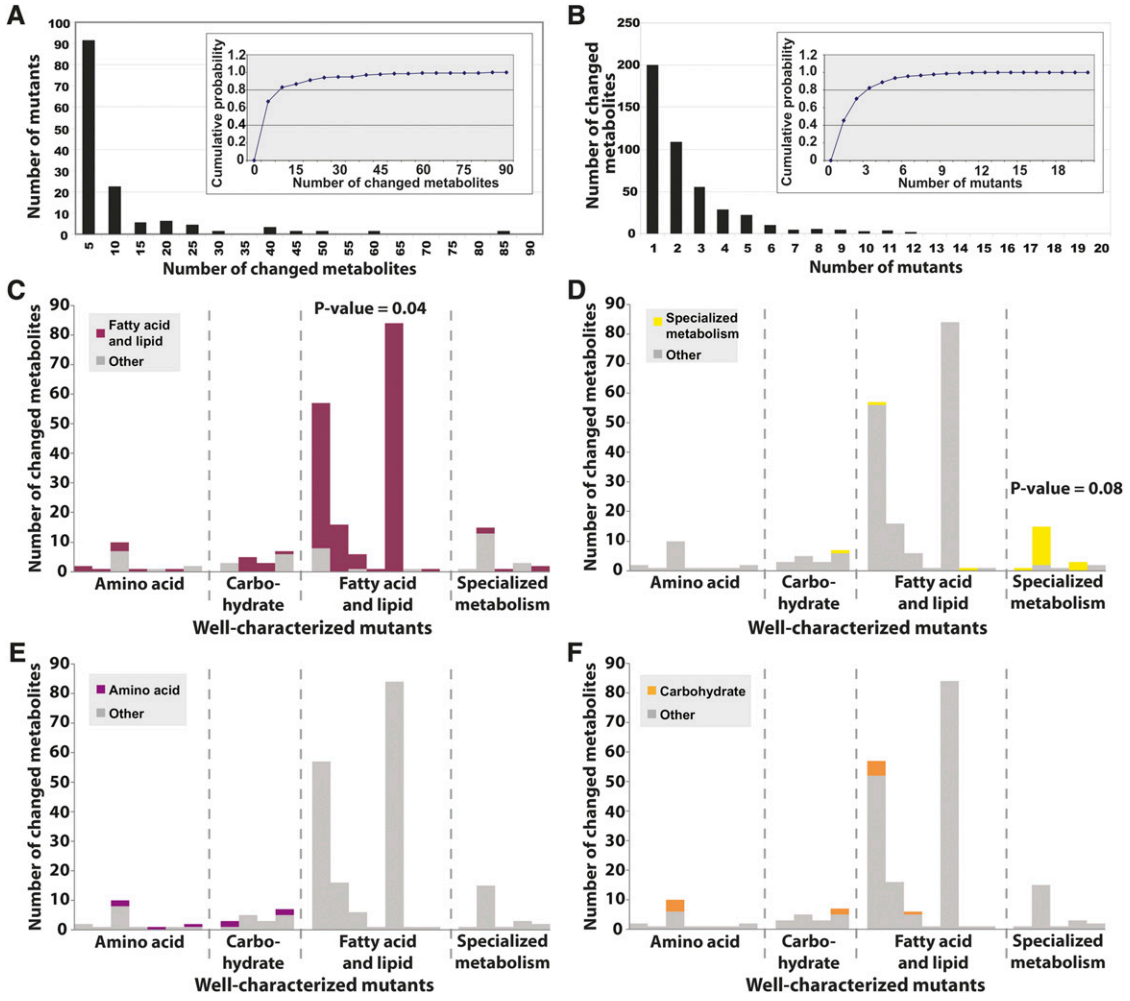

Figure 3. Specificity and diversity of significantly changed metabolites across 136 mutants. A, Distribution of the number of changed metabolites per mutant. B, Distribution of the overlap of changed metabolites across the mutants. Insets show cumulative distributions. C, Proportions of the changed metabolites in fatty acid and lipid metabolism in groups of mutants belonging to four categories. D, Proportion of the changed metabolites in specialized metabolism. E and F, Proportion of the changed metabolites in amino acid and carbohydrate metabolism. *P* values were determined from hypergeometric tests.

across the mutants (Supplemental Figs. S6, A and B, and S7; Supplemental Table S3). These results demonstrate that metabolic responses from single gene perturbations are quantitatively and qualitatively specific and diverse.

To test whether the specificity of metabolic responses from the mutants corroborated the functional role of the genes, we examined the changed metabolites in the mutants of well-characterized genes. Fatty acids and lipids were enriched in the changed metabolites of perturbed genes involved in fatty acid and lipid metabolism ($P = 0.04$, hypergeometric test; Fig. 3C). Specialized metabolites were also enriched in the changed metabolites of perturbed genes involved in specialized metabolism, though it was not statistically significant ($P = 0.08$, hypergeometric test; Fig. 3D). We did not observe a statistical enrichment of amino acids and carbohydrates in the changed metabolites of perturbed genes involved in amino acid and carbohydrate metabolism (Fig. 3, E and F). Composition of the changed metabolites remained similar across the metabolic domains whereas the number of changed metabolites decreased when only three replicates were used to identify the changed metabolites (Supplemental Fig. S8). These results suggest that mutations in different classes of metabolic genes have different impacts on metabolism and that genes involved in fatty acid and lipid metabolism tend to specifically affect fatty acids and lipids. However, we cannot exclude the possibility that the different number of detected metabolites in the metabolic domains may have led to the observed statistical enrichment patterns.

Evidence of Functional Links between Changed Metabolites and Perturbed Genes

To test whether the metabolite profiles we generated were consistent with previously observed changes in

metabolism, we examined the changed metabolites and the perturbed enzymatic functions of two mutants associated to well-characterized genes involved in fatty acid and lipid metabolism and specialized metabolism. One of the well-characterized genes in our study, AT1G32200 (*Acyltransferase1* [*ATS1*]), encodes a chloroplast glycerol-3-P acyltransferase that synthesizes 1-acylglycerol 3-P (lysophosphatidic acid) through the acyl-CoA-dependent acylation of sn-glycerol-3-P (Nishida et al., 1993). Lines carrying a transfer DNA insertion in the gene (*ats1*, SALK_069657C) lack most of the plastidic glycerol-3-P acyltransferase activity and are disrupted in the galactoglycerolipid biosynthesis pathway (Fig. 4A; Kunst et al., 1988; Xu et al., 2006). In this mutant, the amount of galactolipids that contained the plastid-derived diacylglycerol (DAG) moiety decreased, whereas the amount of galactolipids that contained the endoplasmic reticulum (ER)-derived DAG moiety increased, perhaps to maintain plastid membrane function (Fig. 4A; Okazaki et al., 2013). Plastid-derived glycerolipids have a 16-carbon fatty acid at the sn-2 position, and ER-derived glycerolipids preferentially carry an 18-carbon fatty acid at the sn-2 position (Heinz and Roughan, 1983; Browse et al., 1986; Okazaki et al., 2013). The metabotype of *ats1* lines showed a significant decrease in the plastid-derived 16-carbon fatty acids of monogalactosyl DAGs, digalactosyl DAGs, and sulfoquinovosyl DAGs and a significant increase in the 18-carbon fatty acids of monogalactosyl DAG, digalactosyl DAG, and sulfoquinovosyl DAG made in the ER compared with the wild type (Fig. 4B). These results show that the metabotypes of *ats1* were consistent to changes that would be caused by perturbation of the enzyme encoded by *ATS1*.

Another well-characterized gene in our study, AT5G13930 (*Transparent Testa4* [*TT4*]), encodes a naringenin-chalcone synthase that produces a precursor of naringenin, 2′4′6′4-tetrahydroxychalcone (Shirley et al., 1995; Dana et al., 2006). Lines carrying a transfer DNA insertion in *TT4* (*tt4*, SALK_020583) lack the naringenin-chalcone synthase activity and are blocked in the flavonoid biosynthesis pathway (Supplemental Fig. S9; Shirley et al., 1995). The metabotypes of *tt4* lines showed that one metabolite, naringenin, was significantly changed in abundance in the mutant compared with the wild type (4-fold decrease in the mutant, $P = 2.39 \times 10^{-3}$, false discovery rate [FDR]-adjusted Student's *t* test). This indicates that the lack of naringenin-chalcone synthase activity from *TT4* perturbation decreased the concentration of naringenin. Kusano et al. (2011) and Yonekura-Sakakibara et al. (2008) found that the relative concentrations of kaempferol 3-*O*-rhamnoside 7-*O*-rhamnoside, kaempferol 3-*O*-glucoside 7-*O*-rhamnoside, and quercetin 3-*O*-glucoside 7-*O*-rhamnoside from *tt4* were significantly changed when compared with the wild type. These results also indicate that metabolic changes are associated with the loss of the enzymatic function of *TT4*, because all these metabolites are derived from the glycosylation of kaempferol, which in turn is generated from naringenin.

These metabotype patterns demonstrate that metabolic changes of the perturbed genes are not random, but can be closely associated with the enzymatic functions of the perturbed genes in the metabolic pathways. These observations prompted us to examine whether the changed metabolites are closely linked to each other and how the changed metabolites are linked to the perturbed genes in the genome-scale metabolic network.

Patterns of Changed Metabolites in the Reconstructed Metabolic Network

To investigate the effects caused by perturbing the expression of a metabolic gene on the metabolic system as a whole, we assessed metabolic changes in the context of a metabolic network. A metabolic network consists of a collection of metabolic pathways composed of biochemical reactions where metabolites are used as substrates to generate products. Therefore, the metabolic system can be represented by a network where a metabolite and a reaction are connected if the metabolite is a substrate or a product of the reaction in a pathway. We reconstructed a metabolic network of Arabidopsis using the pathway database AraCyc version 8.0 (Mueller et al., 2003; Zhang et al., 2010). The network consists of 2,689 reactions and 2,625 metabolites connected by 7,455 metabolite-reaction links (Supplemental Table S4; Supplemental Fig. S10).

To determine whether the changed metabolites in the mutants are catalyzed by reactions in closely related pathways or just randomly associated, we assessed the distance (number of enzymatic steps) between the changed metabolites in the network. The distances among the changed metabolites were significantly shorter than expected by chance ($P = 1.0 \times 10^{-4}$, Student's *t* test) and the distance among the unchanged metabolites ($P = 9.15 \times 10^{-8}$, Student's *t* test) across the mutants (Fig. 5A; Supplemental Fig. S11). As an example, the average distance among the changed metabolites of the mutant *ats1* (SALK_069657C, AT1G32200) in the network was 7.68 links compared with 9.52 links from random expectation (Fig. 5B). These results suggest that the changed metabolites from gene perturbations are not randomly associated but closely linked to each other in related pathways or in the same pathway.

The Number of Isozymes Affects the Distance between the Perturbed Reactions and Changed Metabolites

Metabolite concentrations depend on the rate of the reactions that consume or produce them. Thus, changing the enzymes' abundance by perturbing the expression of the encoding genes would change the concentration of the metabolites that are directly or closely linked to the reactions catalyzed by these enzymes. This somewhat obvious hypothesis has not been tested systematically. To test this hypothesis, we asked whether the changed metabolites were closely associated to the perturbed reactions in the network by

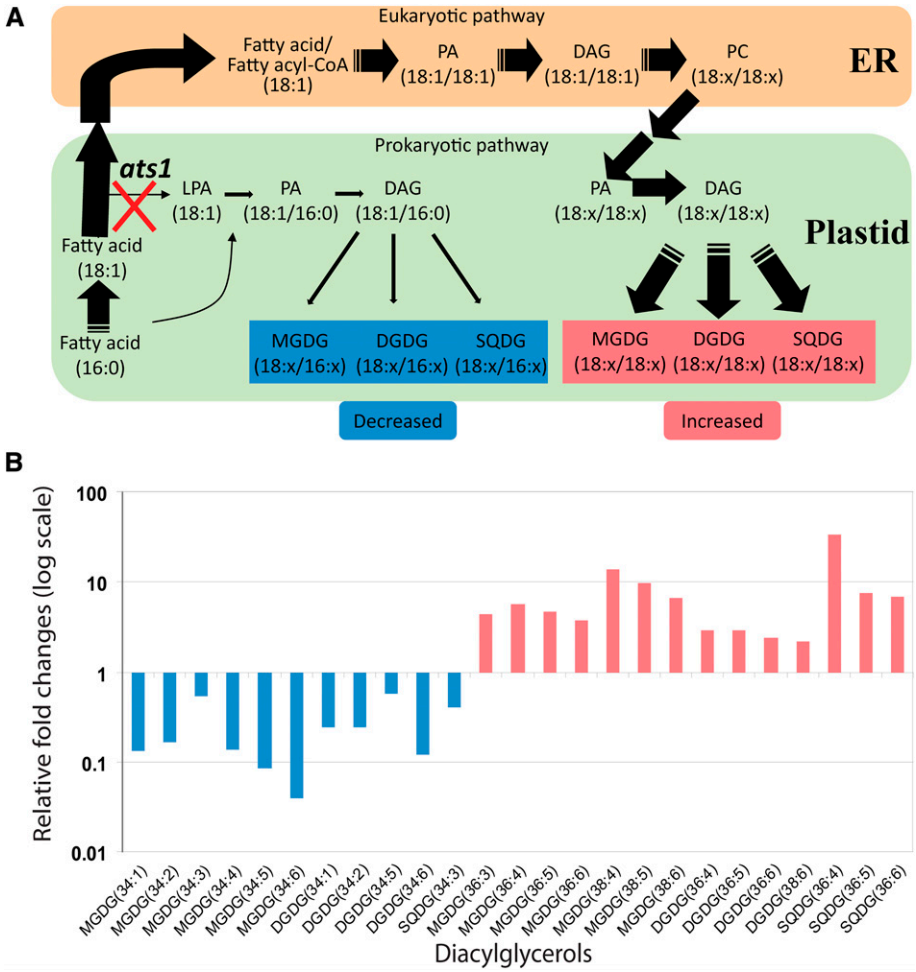

Figure 4. The metabotype of *ats1* reflects the enzymatic function of the gene. A, Simplified DAG biosynthesis pathway. Most DAGs are produced by the eukaryotic pathway located in the ER; in this pathway, an 18-carbon fatty acid is added to the sn-2 position of glycerol-3-P. The red X indicates the perturbed reaction in the *ats1* mutation. Arrows show enzymatic reactions, and thickness of the arrows indicates the flux of enzymatic reactions. B, Relative fold-changes of DAGs in the metabotype of *ats1* compared with the wild type. In the *ats1* mutant, DAGs with 16-carbon fatty acids (34:x) are significantly decreased whereas DAGs with only 18-carbon fatty acids (36:x) are significantly increased. DGDG, Digalactosyl DAG; LPA, lysophosphatidic acid; MGDG, monogalactosyl DAG; PA, phosphatidic acid; PC, phosphatidylcholine; SQDG, sulfoquinovosyl DAG.

measuring the distance between the changed metabolites and the perturbed reactions catalyzed by well-characterized enzymes. Twenty-one out of the 29 well-characterized enzymatic mutants were evaluated because eight mutants did not have changed metabolites in the network. In 13 perturbed reactions of seven mutants, the changed metabolites were significantly closer to the perturbed reactions than the same number of randomly chosen compounds among the detected metabolites ($P < 0.05$, Student's t test). In the remaining 26 reactions in 14 mutants, the changed metabolites were not closer to the perturbed reactions than expected by chance (Fig. 6A). This unexpected result indicates that perturbing a reaction has a more complex effect on the metabolic network than previously assumed.

To understand the causes that underpin the difference in the distance between the changed metabolites and site of perturbation in the network, we grouped the genes into two categories: (1) genes that cause metabotypes that are close to the mutated reaction than expected by chance (proximal metabotype) and (2) genes that cause metabotypes that are not close to the mutated reaction than expected by chance (nonproximal metabotype). We compared biological and network characteristics of the perturbed genes and reactions between the two groups of genes. At the enzyme-coding gene level, we evaluated: (1) the expression levels in 2-week-old rosette leaves and (2) the existence of paralogs. At the reaction level, we evaluated: (1) the number of changed metabolites upon a reaction perturbation, (2) the average distance among the metabolites affected by the changes in the reaction, (3) the number of metabolites connected to a perturbed reaction, (4) the catalytic speed of a perturbed reaction (a rate-limiting reaction), (5) the uniqueness of a perturbed reaction on consuming or producing a metabolite, (6) the number of predicted isozymes catalyzing a perturbed reaction, and (7) the location of the perturbed reaction in the network.

Among the nine properties tested, only the number of isozymes of a perturbed reaction was significantly smaller in the proximal metabotype than the nonproximal metabotype group ($P = 1.15 \times 10^{-4}$, Student's t test; Fig. 6B; Supplemental Fig. S12). Furthermore, the number of predicted isozymes of the perturbed reactions positively correlated with the distance between the changed metabolites and the perturbed reaction

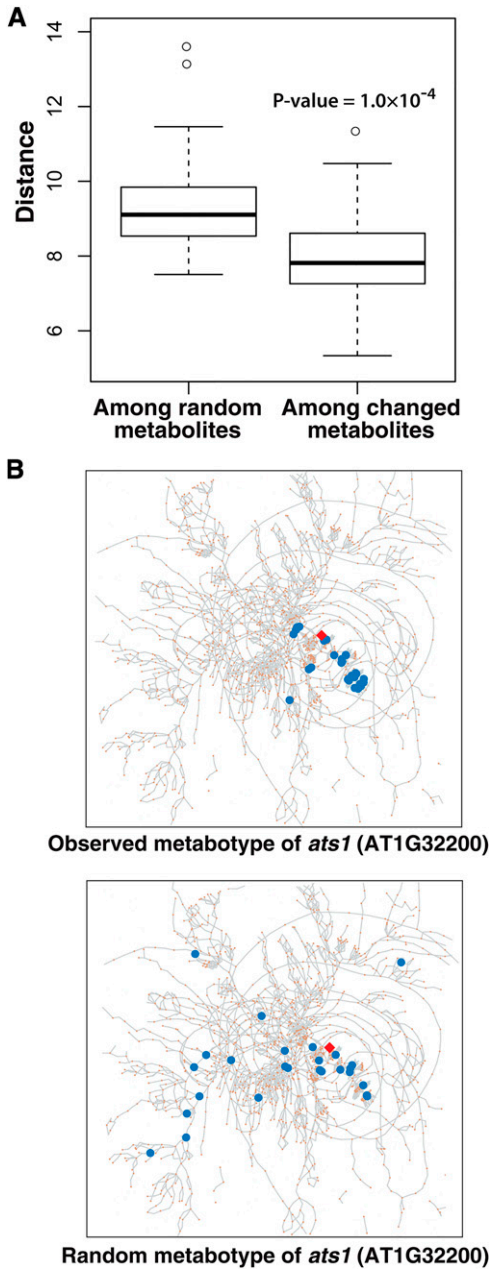

Figure 5. Distance of changed metabolites to each other across the mutants in the metabolic network. A, Distribution of the distances among changed metabolites compared with distances among the same number of randomly selected metabolites from the same metabolic domains across all genes. B, An illustrative example of metabotypes in a reconstructed metabolic network, showing changed metabolites in the knockout mutant SALK_069657C compared with the same number of randomly selected metabolites. Red diamonds are perturbed reactions, and blue dots are the differentially accumulated metabolites (top) and randomly selected metabolites (bottom).

(R = 0.41, R^2 = 0.17, P = 0.01, Fisher's z test; Fig. 6C). These results indicate that metabolite abundance is affected not simply by the adjacency to the perturbed reaction, but by the characteristics of the associated enzyme-encoding genes.

DISCUSSION

This study assessed patterns caused by single gene perturbations by combining genetic, metabolomic, and network analysis. By identifying significantly changed metabolites across 136 single gene mutants, we observed that changed metabolites varied quantitatively and qualitatively. In about 80% of the mutants, perturbing the expression of a gene affected the concentration of less than 10 metabolites. Only two mutants affected more than 50 metabolites. This observation supports the concept that the overall metabolic network is highly robust against single gene mutations (Ishii et al., 2007).

To examine the global properties of the changed metabolites across the 136 mutants, we reconstructed a metabolic network for Arabidopsis using the AraCyc database. We removed the 24 currency metabolites (proton, water, oxygen molecule, $NADP^+$, NADPH, ATP, diphosphate, carbon dioxide, phosphate, ADP, coA, UDP, NAD^+, NADH, AMP, ammonia, hydrogen peroxide, oxidized electron acceptor, reduced electron acceptor, 3-5-ADP, GDP, carbon monoxide, GTP, and FAD; Supplemental Table S5) from the network because they are associated with many reactions in the cell, thus creating many biologically unrealistic shortcuts in the metabolic network (Ma and Zeng, 2003a, 2003b; Verkhedkar et al., 2007). For example, water is an essential compound involved in most biochemical reactions. The biological paths among reactions created by water could be neither functionally associated nor biologically meaningful in metabolic pathway analyses because water is not functionally specific but ubiquitous in metabolism. To analyze the impact of removing the 24 currency metabolites in the network, we measured three topological properties of the network before and after removing the currency metabolites. We compared the average distance of all possible shortest paths, the fraction of reactions in connected components, and the number of links before and after removing the currency metabolites. On average, the shortest path between any two nodes before removing currency metabolites (4.29) was much shorter than the average distance after removing currency metabolites (9.00). In addition, more than 40% of the links (5,236/12,691) were created by the 24 currency metabolites. About 90% of the reactions (2,400/2,689) were included in the largest component of the network after deleting the currency metabolites. These analyses indicate that the deletion of the currency metabolites did not disrupt the structure of the overall network but increased the possibility of identifying functionally and biologically meaningful metabolic paths in the network.

We found that changed metabolites were closer to each other than the same number of randomly selected detected metabolites in the network. Because metabolites are connected to reactions in the network when the metabolite is a substrate or a product of the reaction, the proximity of changed metabolites implies that they are likely participating in the same pathway or closely related pathways. As an example, the changed metabolites of *ats1* involved in DAG biosynthesis were

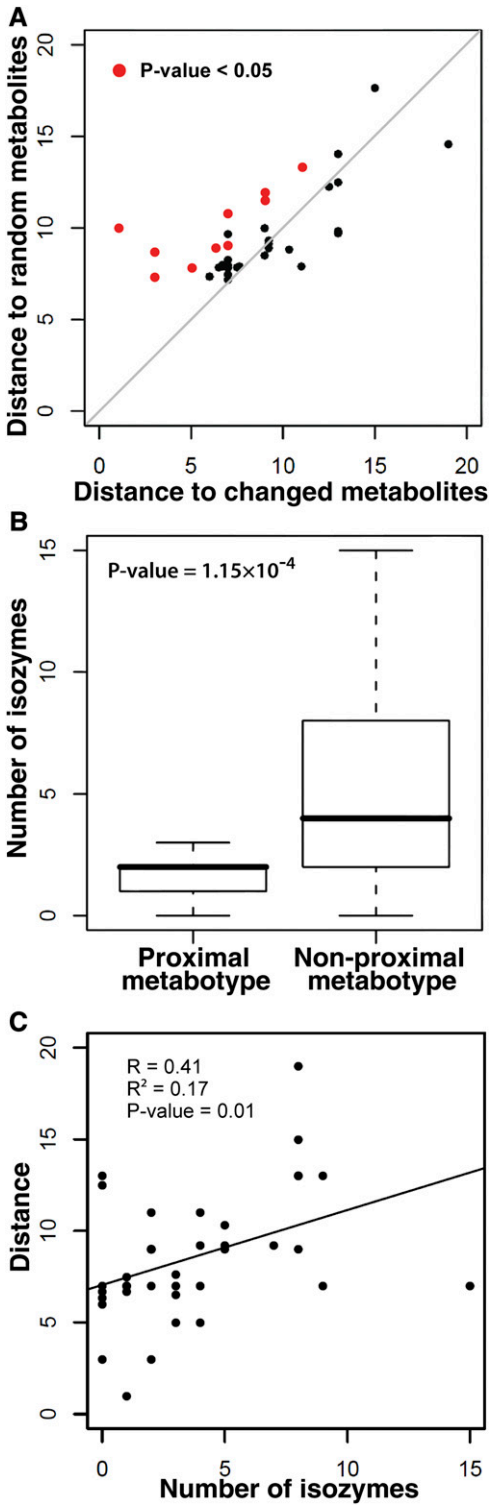

Figure 6. Effect of the number of predicted isozymes on the distance between changed metabolites and perturbed reactions in the metabolic network. A, Pairwise tests on the distance between changed metabolites and perturbed reactions for each well-characterized mutant. Dots represent the average distance from the perturbed reactions to the changed metabolites and the same number of randomly selected metabolites. Red dots are the perturbed reactions with significant closeness to the changed metabolites. B, Distribution of the number of isozymes between the mutants with proximal metabotype (changed metabolites are closer to perturbed reactions than expected by chance) and the mutants with nonproximal metabotype (changed metabolites are not closer to the perturbed reactions than expected by chance). C, Correlation between the number of isozymes of the perturbed reactions and the distance between the changed metabolites and the perturbed reaction.

enriched in lipids and were significantly closer to each other than among all lipids ($P = 0.01$, Student's t test) and among randomly sampled detected compounds ($P = 3.61 \times 10^{-15}$, Student's t test; Figs. 4 and 5B; Supplemental Table S2). However, we also observed dispersed patterns of metabotypes in the network, suggesting that enzymes encoded by these genes could be involved in multiple reactions or regulatory pathways rather than in specific enzymatic functions, which could affect metabolism more globally when perturbed. Another possibility is that these enzymes catalyze the so called hub metabolites, metabolites highly associated with other metabolites in the metabolic network; thus, disturbing these enzymes would likely have global effects. Moing et al. (2011), using a network-based analysis to obtain a global overview of the associations among 715 metabolites in three ripening stages of melon (*Cucumis melo*) fruit, identified that 13% (96/715) of these metabolites were hub metabolites. Alternatively, missing reactions in the network could increase the path lengths among the changed metabolites. Although the changed metabolites were close to each other in the network, they were not immediately adjacent to each other in many of the mutants. This could reflect the plasticity of a metabolic network against gene mutations by rerouting or rewiring biochemical pipelines (Harrison et al., 2007; Hanada et al., 2011). Fendt et al., 2010 also found that, in yeast, a perturbation in central carbon metabolism enzymes increased the local (but not necessarily immediate) reaction substrate concentrations, probably as a passive mechanism to minimize differences in flux upon metabolic disturbance. Better characterization of the flux control coefficients and other kinetic properties of the enzymes and reactions could help explain the exact nature of this type of pattern of metabotypes, but such efforts are outside the scope of this study.

We also found that changed metabolites were not always close to the perturbed reactions in mutants. In only seven out of the 21 well-characterized mutants, the changed metabolites were close to the perturbed reactions (proximal metabotypes). To analyze which properties of the mutated genes or perturbed reactions are responsible for the distance between the changed metabolites and the site of perturbation in the network, we investigated six biological and three metabolic network properties of the mutated genes or perturbed reactions. Among them, only the number of predicted isozymes was correlated with the proximity of the changed

metabolites to the site of perturbation (Fig. 6, B and C; Supplemental Fig. S12). When there were fewer isozymes annotated to a perturbed reaction, the changed metabolites were closer to the perturbed reaction in the metabolic network. Possibly, the absence of isozymes for a reaction prevents the substitution or compensation of the enzymatic function of the perturbed reaction, resulting in the changed metabolites in the vicinity of the corresponding reactions. On the other hand, the presence of several isozymes might allow the divergence of some of them to perform new biochemical functions, resulting in catalyzing a different (distant) reaction in the metabolic network (Schmidt et al., 2003; Fani and Fondi, 2009). Additionally, the isozymes could be expressed in different cell types, and the observed metabotype could represent an average of several cell type-specific metabotypes. In these scenarios, perturbing a gene with many isozymes is likely to cause metabolic changes in functionally different parts of the metabolic network. With our current experimental design focusing on whole leaves, we were not able to examine this possibility.

While growing evidence indicates that metabolomics is valuable for understanding the biological roles of genes (Fiehn et al., 2000; Saito and Matsuda, 2010), there are still challenges to be overcome (Mendes, 2006). The first fundamental issue is the large number of unidentified metabolites (Fig. 2A). About 70% of the detected compounds in our study are not annotated due either to the complexity of the chemical structure of the compounds or the lack of available standards. Second, metabolic changes compared with the wild type are not always detectable for single gene mutations. For example, we could not detect any changes in metabolites in 23% of the mutants (Supplemental Fig. S6, A and B). One possibility is the variability of the levels of detected metabolites arising from biological and technical noise. Perhaps, transient metabolites would not be detected in the current setting. Alternatively, the robustness of the metabolic system caused by genetic or network-level redundancy could also prevent the detection of metabolic changes in the single-gene mutants (Cornelius et al., 2011). Third, detecting changed metabolites in a single time point and from different organs may cause distinct metabotypes even though the same mutants are studied. By surveying other metabolomics studies (Yonekura-Sakakibara et al., 2008; Kusano et al., 2011) of *tt4* and *tt5* mutants, the metabotypes from the previous studies were similar, but not identical, to the metabotypes from our study. We identified flavonoid naringenin and naringenin derivatives including genistein, genistin, luteolin-3′-7-*O*-glucoside, naringenin-7-*O*-glucoside, and naringin (Lapcik et al., 2006), whereas the previous studies identified kaempferol 3-*O*-rhamnoside 7-*O*-rhamnoside, kaempferol 3-*O*-glucoside 7-*O*-rhamnoside, and quercetin 3-*O*-glucoside 7-*O*-rhamnoside as changed metabolites in *tt4* and *tt5* mutants. Because flavonoids are known to be involved in responding to different ultraviolet light conditions and toxic compounds in soil such as aluminum, it is possible that differences in growth conditions could affect the metabolic changes of flavonoids in different studies (Winkel-Shirley, 2002). Nonetheless, both naringenin and kaempferol are known to be associated with the catalytic activity of the enzymes encoded by *TT4* and *TT5*; thus, the metabolic changes from the previous studies and this study were both associated with the metabolic functions of the mutated genes.

In addition to the limitation of metabolite identification, different metabolic network models represent different sets of reactions. To minimize the impact of this limitation, we reconstructed the metabolic network using the AraCyc database, which is extensively curated with experimentally validated data. However, about 38% of the chemically identified metabolites (509/1,348) were still not present in the database, which may reduce the power of network analysis on the metabolic response caused by gene perturbation (Supplemental Fig. S2). Detecting and identifying all of the metabolites in a metabolic network in untargeted studies using incomplete reference compound libraries can be challenging (Hegeman, 2010). Because we evaluated metabolic changes of gene perturbations based on only a fraction of the metabolites in the network, the conclusions we made may change if there were more metabolites detected and identified. In addition, orphan reactions (reactions that are not connected to any other reactions) in pathway databases may be indicators of missing metabolites and enzymatic activities as well as the incompleteness of a metabolic network (Mueller et al., 2003), which could also limit the ability to link metabotypes to the perturbed genes using metabolic network analysis.

Our study reveals characteristics of metabotypes caused by single gene mutations in a genome-scale metabolic network. By examining biological and topological properties of metabolic changes in the context of a genome-scale network, metabolomics can be used as a tool to identify the relationships between changed metabolites and their potential impact on the metabolic system and biology of the organism. This type of analysis could be useful in assessing the effects of GM on crops. For example, we found that perturbations generated by mutating single genes could affect metabolites that are proximal or nonproximal to the perturbation sites. By targeting genes that show only proximal metabotypes, undesired and unknown effects could potentially be minimized and, in parallel, more direct responses could be achieved. In addition, metabolomics combined with metabolic network analysis has the potential to systematically evaluate the substantial equivalence between GM and non-GM organisms. For example, the amount of essential or toxic compounds in GM plants can be quantified by comparing the metabolic profiles with the non-GM plants (Smith et al., 2010; van Rijssen et al., 2013). Developing tools that enable comprehensive evaluation of the transgene's effects upon the GM crops will be important to fully understand the impact of the GM (Ricroch et al., 2011). As experimental methods and data resolution improve, our understanding of the metabolome and the relationship between genotypes

and metabolic phenotypes should become clearer (Fernie, 2007). Emerging tools and resources such as genome-scale metabolic networks, quantitative network modeling, and metabolomics may help assess the effects of GM on metabolism and may facilitate rational assessment of unintended effects of GM on metabolism.

MATERIALS AND METHODS

Selecting Uncharacterized Genes Using AraNet

To select candidates from the pool of mutants perturbed in the expression of previously uncharacterized genes, we predicted functions of the corresponding gene products based on their functional similarity to known enzymes using AraNet (Lee et al., 2010; Hwang et al., 2011). We queried AraNet using well-characterized genes for each metabolic pathway from AraCyc 8.0 (Mueller et al., 2003) and identified candidate genes encoding missing enzymes in the pathway. The gene function predictability of AraNet across 11 functional categories of metabolism is shown in Supplemental Figure S13. Out of the top 200 candidates for each pathway with area-under-the-curve scores greater than 0.7, we chose genes with the highest log likelihood score that met the following conditions: (1) gene expression in 2-week-old leaves was greater than 100 in AtGenExpress data (Schmid et al., 2005), (2) there were no recently duplicated genes (Blanc and Wolfe, 2004), (3) there were no known functional annotations (Lamesch et al., 2012), and (4) there were no visible phenotype (Supplemental Table S1).

Plant Growth Condition

Wild-type Col-0 and mutant seedlings of the 136 selected genes (129 homozygous lines carrying knockout alleles and seven overexpressing lines using the 35S promoter) were grown under consistent conditions to minimize environmental effects as described in Quanbeck et al. (2012). Seeds were sown on sterile Murashige and Skoog basal salt mixture supplemented with 0.1% (w/v) Suc and 1× liquid vitamin solution in 100-mm × 100-mm × 15-mm-square petri dishes. Seeds were arranged in a single horizontal line, and each dish contained between 18 and 20 seeds. After sowing the seeds, the plates were wrapped with Micropore tape and then stored horizontally for 4 d at 4°C, with illumination of 1 μE m^{-2} s^{-1} to break seed dormancy. On the 5th day, plates were moved to the growth room and held in a vertical position in wire rack holders for 16 d under a 24-h regimen of constant fluorescent illumination of 50 μE m^{-2} s^{-1} and temperature of 23°C to 25°C. On the twentieth day after sowing the seeds, dishes were opened, and the aerial portions of the plants were harvested immediately upon plate opening and were frozen in liquid nitrogen. Details of the plant growth condition for each batch are available on PlantMetabolomics.org (http://www.plantmetabolomics.org) and also described in Bais et al. (2010, 2012) and Quanbeck et al. (2012). All of the plants were grown and harvested at Iowa State University and sent out to collaborating laboratories for metabolite extraction and profiling.

Metabolomics Analytical Platforms and Metabolite Profiling

We used 11 analytic platforms (seven targeted and four nontargeted) to profile the concentrations of 1,348 metabolites. We first distinguished 4,482 peaks for putative compounds from the leaves of 2-week-old seedlings using 11 mass spectrometric platforms, and we then identified 1,348 metabolites based on the mass fragmentation patterns by comparison with those of authentic standards using the National Institute of Standards and Technology library. The targeted platforms detected amino acids using GC with a flame ionization detector, isoprenoids using HPLC with diode array detection and MS, lipids using electrospray ionization-tandem MS, and ceramides, cuticle waxes, fatty acids, and phytosterols using GC-MS. The nontargeted platforms included capillary electrophoresis-MS, LC-MS, GC-time-of-flight-MS, and ultra-performance LC-quadrupole time-of-flight-MS. These nontargeted platforms captured various metabolites including amino acids, fatty acids, alcohols, carbohydrates, nucleosides, chalcones, flavonoids, glucosinolates, and terpenoids. Metabolic profiling data are available at http://plantmetabolomics.vrac.iastate.edu/ver2/datasets/overview.php (Bais et al., 2010), and details of the extraction protocols for each platform are available at http://plantmetabolomics.vrac.iastate.edu/ver2/tutorials/protocols.php (Bais et al., 2010, 2012).

Classifying Detected Metabolites

The chemically defined compounds were classified into nine categories (fatty acid and lipid, specialized metabolism, carbohydrate, amino acid, nucleotide, cofactor, plant hormone, amine and polyamine, and others) based on the AraCyc compound ontology (Mueller et al., 2003). We wrote a Java program to map profiled metabolite names to common names and synonyms in AraCyc using text matching. We then manually classified the unmapped metabolites based on the classification from chemical entities of biological interest (de Matos et al., 2010), Kyoto Encyclopedia of Genes and Genomes (Kanehisa et al., 2012), PubChem (Wang et al., 2012), Wikipedia, and Google searches. We exercised caution in adding unmapped metabolites to the database. In cases where there was clear additional evidence for the metabolite's existence in Arabidopsis (*Arabidopsis thaliana*), for example, based on literature citations stored in the Knapsack database (Afendi et al., 2012), the compound was added to AraCyc. However, in cases where the compound was typically associated with bacteria, such as aurantimycin, we used it for metabotype analysis but did not add it to the AraCyc database. Because the network analysis required compounds to be connected to reactions, we did not introduce unmapped compounds into the network that were not linked to a reaction in AraCyc.

Identifying Significantly Changed Metabolites

To identify metabolites that were significantly changed in the mutants compared with wild-type plants, we processed the metabolomics data through several filtering steps (Supplemental Fig. S3). First, we removed metabolites that were below detection limit. To check the consistency of the metabolic concentrations within biological replicates, we evaluated the correlation of three replicates for the amino acid- and ceramide-targeted platforms and capillary electrophoresis-MS and LC-MS-based untargeted platforms; three to six replicates for fatty acid-, cuticle wax-, phytosterol-, plastidial isoprenoid-, and lipid-targeted platforms and the ultra-performance LC-quadrupole time-of-flight-MS-based untargeted platform; or six replicates for the GC-time-of-flight-MS-based untargeted platform and removed the replicate that weakly correlated with the other replicates (correlation threshold < 0.7). About 0.7% of the profiles were removed due to weak correlation among biological replicates, indicating that more than 99% of the profiles were highly reproducible. Plants were grown in several experimental batches. In each batch for each analytic platform, metabolites were normalized using the median concentration of log-transformed values ($\log_2$) across all genotypes within a platform for each metabolite to minimize possible technical differences among the batches (Quackenbush, 2002). We conducted three independent FDR-adjusted Student's t tests (Benjamin and Hochberg, 1993) between mutants and two sets of wild-type (Col-0) plants to identify metabolites that were significantly changed in the mutants (Supplemental Fig. S3). A metabolite was deemed significantly changed in the mutant only if the FDR-adjusted Student's t tests were significant in the relative concentrations between a mutant and two independent sets of wild-type plants (FDR-adjusted Student's t test < 0.10) and an FDR-adjusted Student's t test was not significant between the two sets of wild-type plants (FDR-adjusted Student's t test ≥ 0.10). Significantly changed metabolites for the mutants are listed in Supplemental Table S2.

Enrichment Test for Changed Metabolites across Functional Categories of Genes

To check whether the changed metabolites were enriched in a functional category, we conducted a hypergeometric test on the functional category of metabolites for each functional category of the genes. First, we counted the number of changed metabolites by metabolite functional category for each mutant and then examined whether the changed metabolites belonging to each category was enriched within a mutant by comparing them with the background frequency (the number of all identified metabolites in each metabolite category). Second, we tested whether the number of mutants in a functional category had an enriched category of metabolites using a hypergeometric test.

Reconstructing a Genome-Scale Metabolic Network

We extracted reactions, enzymes, genes, and compounds from the AraCyc database (version 8.0; Mueller et al., 2003) and converted them into a bipartite

metabolic network with two classes of disjoint nodes containing 2,689 reactions and 2,625 metabolites (Supplemental Table S4; Supplemental Fig. S10). Nodes represent either metabolites or reactions, and edges (links) represent the association between metabolites and reactions. Because both a reaction-centric network (a node is a reaction and an edge is a metabolite) and a metabolite-centric network (a node is a metabolite and an edge is a reaction) could not specify the links of metabolites and reactions in the networks, respectively, we reconstructed a genome-scale metabolic network as an undirected bipartite network. We did not consider the directionality of metabolic reactions because only five reactions have directionality supported by experimental evidences in AraCyc. A metabolite was connected to a reaction if the metabolite was a substrate or a product of the reaction. The constructed metabolic network is a bipartite network because nodes were divided into two disjoint sets, U (metabolites) and V (reactions), such that every edge connected a node in U to one in V; that is, U and V were independent sets (Diestel, 2005). We then removed 24 currency metabolites (proton, water, oxygen molecule, $NADP^+$, NADPH, ATP, diphosphate, carbon dioxide, phosphate, ADP, CoA, UDP, NAD^+, NADH, AMP, ammonia, hydrogen peroxide, oxidized electron acceptor, reduced electron acceptor, 3-5-ADP, GDP, carbon monoxide, GTP, and FAD) from the reconstructed metabolic network (Supplemental Table S5). Because they take part in many reactions in the cell, thereby creating many biologically unrealistic shortcuts in the metabolic network, any path-based measurement of the network could be biased (Ma and Zeng, 2003a, 2003b; Verkhedkar et al., 2007). To select currency metabolites, we prioritized metabolites by the connectivity (degree) of metabolites in the reconstructed metabolic network and then manually checked the links created by them by evaluating whether they were functionally meaningful.

Distance Measurement in the Metabolic Network

With the reconstructed metabolic network, the shortest distance among the metabolites and reactions was calculated. The distance of two nodes is 1 when they are directly connected in the network, and the distance of two nodes was measured only if they were connected in the network. We wrote a Java program to calculate a shortest distance based on the breadth-first search algorithm (Russell and Norvig, 2009). A metabolic path was ignored if the path was constructed from reactants to reactants or from products to products in a metabolic reaction.

Using the distance calculation described above, we measured adjacency of changed metabolites by calculating shortest distances among changed metabolites. To calculate the average shortest distance among changed metabolites, we determined pairwise shortest distances among all changed metabolites and averaged them for each mutant. We also measured the distance between a changed metabolite and the perturbed reactions in the network. We determined the shortest distances between changed metabolites and a perturbed reaction and then averaged them for each perturbed reaction of each mutant. Because the reconstructed metabolic network is an undirected network, the directionality of a reaction is not taken into consideration for the distance calculation.

Statistical Test for Network Distance Analyses

To check the statistical significance for the distance analysis, we compared the average shortest distance among the changed metabolites with the average shortest distance among randomly selected metabolites in the reconstructed metabolic network. Because the changed metabolites associated with the mutants are a subset of detected metabolites in the network, we randomly selected the same number of detected metabolites as the number of changed metabolites from each mutant and calculated shortest distances among the randomly sampled metabolites. One hundred independent selections and calculations were made to determine the average shortest distances of the random expectation from each mutant. Only the chemically identified (named) metabolites in the metabolic network were taken into consideration for random expectation of the distance of changed metabolites. Average distance among metabolites was measured over only the connected metabolites. To determine whether the distance among the changed metabolites was significantly different from random expectation, we performed Student's t test between the average shortest paths among the changed metabolites and the average shortest paths among the randomly selected compounds in the 136 mutants (Fig. 5A). In addition, the average distance between changed metabolites was compared with the average distance between unchanged metabolites using Student's t test (Supplemental Fig. S11). To determine whether the distance between the changed metabolites and the perturbed reaction was significantly different from random expectation, we performed one-sample Student's t test between the distance of the average shortest path of the changed metabolites to the perturbed reaction and the average shortest paths of randomly selected metabolites to the perturbed reaction (100 samples) for each well-characterized mutant (Fig. 6A).

Statistical Test of Biological and Topological Properties of Genes

We tested six biological and three topological properties that could distinguish between the two groups of mutants associated with the well-characterized genes with different proximity to the changed metabolites. The number of predicted isozymes was determined by counting the number of enzymes associated with a reaction. The rate limitability of a reaction was curated by PMN (2014). The catalytic irreversibility of a reaction and the catalytic uniqueness of a reaction were obtained from AraCyc (Mueller et al., 2003). Expression levels of the genes in 2-week-old rosette leaves were taken from the AtGenExpress database (Schmid et al., 2005). The number of significantly changed metabolites was determined in this study. Sets of paralogous genes were obtained from Thomas et al. (2006). The degree of a reaction is the number of links from the reaction to directly connected neighboring reactions through shared metabolites in the reconstructed network. The betweenness centrality of a reaction is the number of shortest paths from all nodes that pass through that reaction. These two topological properties were calculated by using the open source Java Universal Network/Graph framework (Madadhain et al., 2005). Average shortest distance was calculated using a Java program as described in the previous section. To evaluate statistical difference between the two groups, we conducted Student's t tests or Fisher's exact tests depending on the types of data ($P < 0.05$; Fig. 6B; Supplemental Fig. S12).

Sequence data from this article can be found in Supplemental Table S1.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Criteria used for selecting target genes.

Supplemental Figure S2. Composition of the detected compounds that are named and mapped in AraCyc.

Supplemental Figure S3. Procedure used to identify significantly changed metabolites in the mutants compared with the wild type.

Supplemental Figure S4. Diversity of changed metabolites within the four types of the selected genes.

Supplemental Figure S5. Specificity of changed metabolites within the four types of the selected genes.

Supplemental Figure S6. Composition and number of the changed metabolites across the mutants.

Supplemental Figure S7. Comparison between the proportion of each type of compounds and the number of changed metabolites across the mutants.

Supplemental Figure S8. Proportions of the changed metabolites in four metabolic domains across well-characterized mutants when three biological replicates were used to identify changed metabolites.

Supplemental Figure S9. Illustration of the flavonoid biosynthesis pathway.

Supplemental Figure S10. Visualization of the reconstructed metabolic network based on AraCyc.

Supplemental Figure S11. The distance among changed metabolites across all the mutants and the well-characterized mutants.

Supplemental Figure S12. Comparison of biological and topological properties of metabolic phenotypes between proximal and nonproximal metabotype groups of mutants of known genes.

Supplemental Figure S13. Gene function predictability of metabolic pathways in AraNet.

Supplemental Table S1. List of 136 selected mutants for this study.

Supplemental Table S2. List of significantly changed metabolites between the wild type and mutants.

Supplemental Table S3. The number of changed metabolites across 136 genes.

Supplemental Table S4. Reconstructed bipartite metabolic network from AraCyc version 8.0 used in this study.

Supplemental Table S5. Twenty-four currency metabolites that were removed from the metabolic network.

ACKNOWLEDGMENTS

We thank Geng Ding, Huanan Jin, Joel Schmidt, Libuse Brachova, Li Xu, Yuqin Jin, Brenda Winkel, David Oliver, Edward Cahoon, Chuck Dietrich, Eran Pichersky, Ling Li, Micheline N. Ngaki, John Browse, Jyoti Shah, Katie Dehesh, Martha James, Alan Myers, and Ron Mitler for donating mutant seeds; Damian Priamurskiy, Ricardo Leitao, Michael Ahn, Caryn Johansen, and Tom Walk for assistance in curating compounds; Kun He for conducting initial statistical analyses of the metabolite profiling data; and Lee Chae for insightful discussion.

Received October 22, 2014; accepted February 6, 2015; published February 10, 2015.

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

Allen J, Davey HM, Broadhurst D, Heald JK, Rowland JJ, Oliver SG, Kell DB (2003) High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. Nat Biotechnol **21:** 692–696

Arbona V, Manzi M, Ollas Cd, Gómez-Cadenas A (2013) Metabolomics as a tool to investigate abiotic stress tolerance in plants. Int J Mol Sci **14:** 4885–4911

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

Bassel GW, Gaudinier A, Brady SM, Hennig L, Rhee SY, De Smet I (2012) Systems analysis of plant functional, transcriptional, physical interaction, and metabolic networks. Plant Cell **24:** 3859–3875

Benjamin Y, Hochberg Y (1993) Controlling the false discovery rate: a practical and powerful approach to multiple testing. JR Stat Soc **57:** 289–300

Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry, Ed 5, Vol 30.1. W.H. Freeman, New York

Blanc G, Wolfe KH (2004) Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. Plant Cell **16:** 1679–1691

Brotman Y, Riewe D, Lisec J, Meyer RC, Willmitzer L, Altmann T (2011) Identification of enzymatic and regulatory genes of plant metabolism through QTL analysis in Arabidopsis. J Plant Physiol **168:** 1387–1394

Browse J, Warwick N, Somerville CR, Slack CR (1986) Fluxes through the prokaryotic and eukaryotic pathways of lipid synthesis in the '16:3' plant *Arabidopsis thaliana*. Biochem J **235:** 25–31

Cascante M, Boros LG, Comin-Anduix B, de Atauri P, Centelles JJ, Lee PW (2002) Metabolic control analysis in drug discovery and disease. Nat Biotechnol **20:** 243–249

Chae L, Kim T, Nilo-Poyanco R, Rhee SY (2014) Genomic signatures of specialized metabolism in plants. Science **344:** 510–513

Chang RL, Ghamsari L, Manichaikul A, Hom EF, Balaji S, Fu W, Shen Y, Hao T, Palsson BO, Salehi-Ashtiani K, et al (2011) Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. Mol Syst Biol **7:** 518

Cornelius SP, Lee JS, Motter AE (2011) Dispensability of *Escherichia coli*'s latent pathways. Proc Natl Acad Sci USA **108:** 3124–3129

Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010) C4GEM, a genome-scale metabolic model to study C4 plant metabolism. Plant Physiol **154:** 1871–1885

Dal'Molin CG, Quek LE, Palfreyman RW, Nielsen LK (2011) AlgaGEM: a genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. BMC Genomics **12:** S5

Dana CD, Bevan DR, Winkel BS (2006) Molecular modeling of the effects of mutant alleles on chalcone synthase protein structure. J Mol Model **12:** 905–914

Davis MC, Fiehn O, Durnford DG (2013) Metabolic acclimation to excess light intensity in *Chlamydomonas reinhardtii*. Plant Cell Environ **36:** 1391–1405

de Matos P, Alcántara R, Dekker A, Ennis M, Hastings J, Haug K, Spiteri I, Turner S, Steinbeck C (2010) Chemical entities of biological interest: an update. Nucleic Acids Res **38:** D249–D254

de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumbley SM, Nielsen LK (2010) AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol **152:** 579–589

DeBerardinis RJ, Thompson CB (2012) Cellular metabolism and disease: what do metabolic outliers teach us? Cell **148:** 1132–1144

Dharmawardhana P, Ren L, Amarasinghe V, Monaco M, Thomason J, Ravenscroft D, McCouch S, Ware D, Jaiswal P (2013) A genome scale metabolic network for rice and accompanying analysis of tryptophan, auxin and serotonin biosynthesis regulation under biotic stress. Rice (N Y) **6:** 15

Diestel R (2005) Graph Theory, **Vol 173**. Springer, Verlag, Heidelberg

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

Ebbels TM, Holmes E, Lindon JC, Nicholson JK (2004) Evaluation of metabolic variation in normal rat strains from a statistical analysis of 1H NMR spectra of urine. J Pharm Biomed Anal **36:** 823–833

Erxleben A, Gessler A, Vervliet-Scheebaum M, Reski R (2012) Metabolite profiling of the moss *Physcomitrella patens* reveals evolutionary conservation of osmoprotective substances. Plant Cell Rep **31:** 427–436

Fani R, Fondi M (2009) Origin and evolution of metabolic pathways. Phys Life Rev **6:** 23–52

Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BO (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol **7:** 129–143

Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. Nat Biotechnol **26:** 659–667

Fendt SM, Buescher JM, Rudroff F, Picotti P, Zamboni N, Sauer U (2010) Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. Mol Syst Biol **6:** 356

Fernie AR (2007) The future of metabolic phytochemistry: larger numbers of metabolites, higher resolution, greater understanding. Phytochemistry **68:** 2861–2880

Fernie AR, Obata T, Allen AE, Araújo WL, Bowler C (2012) Leveraging metabolomics for functional investigations in sequenced marine diatoms. Trends Plant Sci **17:** 395–403

Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol **18:** 1157–1161

Field KJ, Lake JA (2011) Environmental metabolomics links genotype to phenotype and predicts genotype abundance in wild plant populations. Physiol Plant **142:** 352–360

Fukushima A, Kusano M, Mejia RF, Iwasa M, Kobayashi M, Hayashi N, Watanabe-Takahashi A, Narisawa T, Tohge T, Hur M, et al (2014) Metabolomic characterization of knockout mutants in Arabidopsis: development of a metabolite profiling database for knockout mutants in Arabidopsis. Plant Physiol **165:** 948–961

Hanada K, Sawada Y, Kuromori T, Klausnitzer R, Saito K, Toyoda T, Shinozaki K, Li WH, Hirai MY (2011) Functional compensation of primary and secondary metabolites by duplicate genes in *Arabidopsis thaliana*. Mol Biol Evol **28:** 377–382

Harrison R, Papp B, Pál C, Oliver SG, Delneri D (2007) Plasticity of genetic interactions in metabolic networks of yeast. Proc Natl Acad Sci USA **104:** 2307–2312

Hegeman AD (2010) Plant metabolomics: meeting the analytical challenges of comprehensive metabolite analysis. Brief Funct Genomics **9:** 139–148

Heinz E, Roughan PG (1983) Similarities and differences in lipid metabolism of chloroplasts isolated from 18:3 and 16:3 plants. Plant Physiol **72:** 273–279

Holmes E, Wilson ID, Nicholson JK (2008) Metabolic phenotyping in health and disease. Cell **134:** 714–717

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

Hwang S, Rhee SY, Marcotte EM, Lee I (2011) Systematic prediction of gene function in *Arabidopsis thaliana* using a probabilistic functional gene network. Nat Protoc **6:** 1429–1442

Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, et al (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science **316:** 593–597

Jung S, Ficklin SP, Lee T, Cheng CH, Blenda A, Zheng P, Yu J, Bombarely A, Cho I, Ru S, et al (2014) The Genome Database for Rosaceae (GDR): year 10 update. Nucleic Acids Res **42:** D1237–D1244

Kanehisa M, Goto S (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res **28:** 27–30

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

Karp PD, Riley M, Paley SM, Pellegrini-Toole A (2002) The MetaCyc Database. Nucleic Acids Res **30:** 59–61

Keasling JD (2010) Manufacturing molecules through metabolic engineering. Science **330:** 1355–1358

Kell DB (2006) Systems biology, metabolic modelling and metabolomics in drug discovery and development. Drug Discov Today **11:** 1085–1092

Kim TY, Sohn SB, Kim YB, Kim WJ, Lee SY (2012) Recent advances in reconstruction and applications of genome-scale metabolic models. Curr Opin Biotechnol **23:** 617–623

Kunst L, Browse J, Somerville C (1988) Altered regulation of lipid biosynthesis in a mutant of Arabidopsis deficient in chloroplast glycerol-3-phosphate acyltransferase activity. Proc Natl Acad Sci USA **85:** 4143–4147

Kusano M, Tohge T, Fukushima A, Kobayashi M, Hayashi N, Otsuki H, Kondou Y, Goto H, Kawashima M, Matsuda F, et al (2011) 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

Lapcik O, Honys D, Koblovska R, Mackova Z, Vitkova M, Klejdus B (2006) Isoflavonoids are present in *Arabidopsis thaliana* despite the absence of any homologue to known isoflavonoid synthases. Plant Physiol Biochem **44:** 106–114

Lee I, Ambaru B, Thakkar P, Marcotte EM, Rhee SY (2010) Rational association of genes with traits using a genome-scale gene network for *Arabidopsis thaliana*. Nat Biotechnol **28:** 149–156

Lee JW, Kim TY, Jang YS, Choi S, Lee SY (2011) Systems metabolic engineering for chemicals and materials. Trends Biotechnol **29:** 370–378

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

Ma H, Zeng AP (2003a) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics **19:** 270–277

Ma HW, Zeng AP (2003b) The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics **19:** 1423–1430

Madadhain J, Fisher D, Smyth P, White S, Boey YB (2005) Analysis and visualization of network data using JUNG. J Stat Softw **1:** 3–17

May P, Christian JO, Kempa S, Walther D (2009) ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii*. BMC Genomics **10:** 209

Mendes P (2006) Metabolomics and the challenges ahead. Brief Bioinform **7:** 127

Merico D, Gfeller D, Bader GD (2009) How to visually interpret biological data using networks. Nat Biotechnol **27:** 921–924

Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T (2012) Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. Proc Natl Acad Sci USA **109:** 339–344

Moing A, Aharoni A, Biais B, Rogachev I, Meir S, Brodsky L, Allwood JW, Erban A, Dunn WB, Kay L, et al (2011) Extensive metabolic cross-talk in melon fruit revealed by spatial and developmental combinatorial metabolomics. New Phytol **190:** 683–696

Monaco M, Sen TZ, Dharmawardhana P, Ren L, Schaeffer M, Naithani S, Amarasinghe V, Thomason J, Harper L, Gardiner J, et al (2013) Maize Metabolic Network Construction and Transcriptome Analysis. Plant Genome **6:** 1–6

Monaco MK, Stein J, Naithani S, Wei S, Dharmawardhana P, Kumari S, Amarasinghe V, Youens-Clark K, Thomason J, Preece J, et al (2014) Gramene 2013: comparative plant genomics resources. Nucleic Acids Res **42:** D1193–D1199

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, et al (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol **138:** 1310–1317

Mueller LA, Zhang P, Rhee SY (2003) AraCyc: a biochemical pathway database for Arabidopsis. Plant Physiol **132:** 453–460

Nicholson JK, Lindon JC (2008) Systems biology: metabonomics. Nature **455:** 1054–1056

Nikolau BJ, Wurtele ES (2007) Concepts in Plant Metabolomics. Springer, New York

Nishida I, Tasaka Y, Shiraishi H, Murata N (1993) The gene and the RNA for the precursor to the plastid-located glycerol-3-phosphate acyltransferase of *Arabidopsis thaliana*. Plant Mol Biol **21:** 267–277

Okazaki Y, Kamide Y, Hirai MY, Saito K (2013) Plant lipidomics based on hydrophilic interaction chromatography coupled to ion trap time-of-flight mass spectrometry. Metabolomics **9:** 121–131

Plant Metabolic Network (2014) Plant Metabolic Network. http://www.plantcyc.org (October 17, 2013)

Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A genome-scale metabolic model of Arabidopsis and some of its properties. Plant Physiol **151:** 1570–1581

Quackenbush J (2002) Microarray data normalization and transformation. Nat Genet **32:** 496–501

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

Raamsdonk LM, Teusink B, Broadhurst D, Zhang N, Hayes A, Walsh MC, Berden JA, Brindle KM, Kell DB, Rowland JJ, et al (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. Nat Biotechnol **19:** 45–50

Radrich K, Tsuruoka Y, Dobson P, Gevorgyan A, Swainston N, Baart G, Schwartz JM (2010) Integration of metabolic databases for the reconstruction of genome-scale metabolic networks. BMC Syst Biol **4:** 114

Ricroch AE, Bergé JB, Kuntz M (2011) Evaluation of genetically engineered crops using transcriptomic, proteomic, and metabolomic profiling techniques. Plant Physiol **155:** 1752–1761

Riedelsheimer C, Lisec J, Czedik-Eysenberg A, Sulpice R, Flis A, Grieder C, Altmann T, Stitt M, Willmitzer L, Melchinger AE (2012) Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize. Proc Natl Acad Sci USA **109:** 8872–8877

Rischer H, Oresic M, Seppänen-Laakso T, Katajamaa M, Lammertyn F, Ardiles-Diaz W, Van Montagu MC, Inzé D, Oksman-Caldentey KM, Goossens A (2006) Gene-to-metabolite networks for terpenoid indole alkaloid biosynthesis in *Catharanthus roseus* cells. Proc Natl Acad Sci USA **103:** 5614–5619

Roessner U, Bowne J (2009) What is metabolomics all about? Biotechniques **46:** 363–365

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

Russell S, Norvig P (2009) Artificial Intelligence: A Modern Approach, Ed 3. Prentice Hall, Upper Saddle River, New Jersey

Saha R, Suthers PF, Maranas CD (2011) *Zea mays* iRS1563: a comprehensive genome-scale metabolic reconstruction of maize metabolism. PLoS ONE **6:** e21784

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

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

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nat Genet **37:** 501–506

Schmidt S, Sunyaev S, Bork P, Dandekar T (2003) Metabolites: a helping hand for pathway evolution? Trends Biochem Sci **28:** 336–341

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

Smith MD, Asche F, Guttormsen AG, Wiener JB (2010) Food safety. Genetically modified salmon and full impact assessment. Science **330:** 1052–1053

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

Sweetlove LJ, Fell D, Fernie AR (2008) Getting to grips with the plant metabolic network. Biochem J **409:** 27–41

Terzer M, Maynard ND, Covert MW, Stelling J (2009) Genome-scale metabolic networks. Wiley Interdiscip Rev Syst Biol Med **1:** 285–297

Thomas BC, Pedersen B, Freeling M (2006) Following tetraploidy in an Arabidopsis ancestor, genes were removed preferentially from one homeolog leaving clusters enriched in dose-sensitive genes. Genome Res **16:** 934–946

Tikunov Y, Lommen A, de Vos CH, Verhoeven HA, Bino RJ, Hall RD, Bovy AG (2005) A novel approach for nontargeted data analysis for metabolomics: Large-scale profiling of tomato fruit volatiles. Plant Physiol **139:** 1125–1137

Urbanczyk-Wochniak E, Sumner LW (2007) MedicCyc: a biochemical pathway database for *Medicago truncatula*. Bioinformatics **23:** 1418–1423

Van Moerkercke A, Fabris M, Pollier J, Baart GJ, Rombauts S, Hasnain G, Rischer H, Memelink J, Oksman-Caldentey KM, Goossens A (2013) CathaCyc, a metabolic pathway database built from *Catharanthus roseus* RNA-Seq data. Plant Cell Physiol **54:** 673–685

van Rijssen FW, Morris EJ, Eloff JN (2013) Food safety: importance of composition for assessing genetically modified cassava (*Manihot esculenta* Crantz). J Agric Food Chem **61:** 8333–8339

Verkhedkar KD, Raman K, Chandra NR, Vishveshwara S (2007) Metabolome based reaction graphs of *M. tuberculosis* and *M. leprae*: a comparative network analysis. PLoS ONE **2:** e881

Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Zhou Z, Han L, Karapetyan K, Dracheva S, Shoemaker BA, et al (2012) PubChem's BioAssay Database. Nucleic Acids Res **40:** D400–D412

Winkel-Shirley B (2002) Biosynthesis of flavonoids and effects of stress. Curr Opin Plant Biol **5:** 218–223

Wurtele ES, Chappell J, Jones AD, Celiz MD, Ransom N, Hur M, Rizshsky L, Crispin M, Dixon P, Liu J, et al (2012) Medicinal plants: a public resource for metabolomics and hypothesis development. Metabolites **2:** 1031–1059

Xu C, Yu B, Cornish AJ, Froehlich JE, Benning C (2006) Phosphatidylglycerol biosynthesis in chloroplasts of Arabidopsis mutants deficient in acyl-ACP glycerol-3-phosphate acyltransferase. Plant J **47:** 296–309

Yobi A, Wone BW, Xu W, Alexander DC, Guo L, Ryals JA, Oliver MJ, Cushman JC (2012) Comparative metabolic profiling between desiccation-sensitive and desiccation-tolerant species of *Selaginella* reveals insights into the resurrection trait. Plant J **72:** 983–999

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in *Arabidopsis*. Plant Cell **20:** 2160–2176

Zhang P, Dreher K, Karthikeyan A, Chi A, Pujar A, Caspi R, Karp P, Kirkup V, Latendresse M, Lee C, et al (2010) Creation of a genome-wide metabolic pathway database for *Populus trichocarpa* using a new approach for reconstruction and curation of metabolic pathways for plants. Plant Physiol **153:** 1479–1491

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