Elucidation of Gene-to-Gene and Metabolite-to-Gene Networks in *Arabidopsis* by Integration of Metabolomics and Transcriptomics*[S]

Masami Yokota Hirai,[a,b,c,d] Marion Klein,[d,e] Yuuta Fujikawa,[a] Mitsuru Yano,[a]
Dayan B. Goodenowe,[f] Yasuyo Yamazaki,[f] Shigehiko Kanaya,[g] Yukiko Nakamura,[g,h]
Masahiko Kitayama,[h] Hideyuki Suzuki,[i] Nozomu Sakurai,[i] Daisuke Shibata,[i] Jim Tokuhisa,[j]
Michael Reichelt,[j] Jonathan Gershenzon,[j] Jutta Papenbrock,[e] and Kazuki Saito[a,b,c,k]

From the [a]Department of Molecular Biology and Biotechnology, Graduate School of Pharmaceutical Sciences, Chiba University, Chiba 263-8522, Japan, [b]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, 4-1-8, Saitama 332-0012, Japan, [c]RIKEN Plant Science Center, 1-7-22, Kanagawa 230-0045, Japan, [e]Institute for Botany, University of Hannover, Herrenhäuserstrasse 2, D-30419 Hannover, Germany, [f]Phenomenome Discoveries Inc., 204-407 Downey Road, Saskatoon, Saskatchewan S7N 4L8, Canada, [g]Department of Bioinformatics and Genomics, Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5, Nara 630-0101, Japan, [h]Ehime Women's College, Baba 421, Ehime 798-0025, Japan, [i]Kazusa DNA Research Institute, 2-6-7, Chiba 292-0818, Japan, and [j]Max-Planck-Institute for Chemical Ecology, Hans-Knöll-Strasse 8, D-07745 Jena, Germany

Since the completion of genome sequences of model organisms, functional identification of unknown genes has become a principal challenge in biology. Post-genomics sciences such as transcriptomics, proteomics, and metabolomics are expected to discover gene functions. This report outlines the elucidation of gene-to-gene and metabolite-to-gene networks via integration of metabolomics with transcriptomics and presents a strategy for the identification of novel gene functions. Metabolomics and transcriptomics data of *Arabidopsis* grown under sulfur deficiency were combined and analyzed by batch-learning self-organizing mapping. A group of metabolites/genes regulated by the same mechanism clustered together. The metabolism of glucosinolates was shown to be coordinately regulated. Three uncharacterized putative sulfotransferase genes clustering together with known glucosinolate biosynthesis genes were candidates for involvement in biosynthesis. *In vitro* enzymatic assays of the recombinant gene products confirmed their functions as desulfoglucosinolate sulfotransferases. Several genes involved in sulfur assimilation clustered with *O*-acetylserine, which is considered a positive regulator of these genes. The genes involved in anthocyanin biosynthesis clustered with the gene encoding a transcriptional factor that up-regulates specifically anthocyanin biosynthesis genes. These results suggested that regulatory metabolites and transcriptional factor genes can be identified by this approach, based on the assumption that they cluster with the downstream genes they regulate. This strategy is applicable not only to plant but also to other organisms for functional elucidation of unknown genes.

In the era of post-genomics, a systematic and comprehensive understanding of the complex events of life is a great concern in biology. The first step in this process is to identify all gene functions and gene-to-gene networks as the components of the system, the whole events of life. The metabolome is the final product of a series of gene actions. Hence, metabolomics has a potential to elucidate gene functions and networks, especially when integrated with transcriptomics. A promising approach is pair-wise transcript-metabolite correlation analysis, which can reveal unexpected correlations and bring to light candidate genes for modifying the metabolite content (1). Gene functions involved in the specific pathway of interest have been identified by the integration of transcript and targeted metabolic profiling in experimental systems where the pathway was activated (2–6). However, up to now, no gene function has been identified by non-targeted analyses of the transcriptome and metabolome. In this report, we analyzed the non-targeted metabolome and transcriptome of a model plant *Arabidopsis* under sulfur (S)[1] deficiency whose genome sequencing has been completed. Our strategy for integrated analyses using batch-learning-self-organizing mapping (BL-SOM) (7–9) enabled the identification of gene-to-gene and metabolite-to-gene networks and new gene functions.

EXPERIMENTAL PROCEDURES

Plant Materials—*Arabidopsis thaliana* ecotype Columbia was grown for 21 days on agar-solidified S-sufficient medium following the methodology of Ref. 10. Plants were transferred to S-sufficient or S-deprived medium and grown for up to 1 week (168 h). Rosette leaves and roots were harvested at 3, 6, 12, 24, 48, and 168 h after transfer, immediately frozen with liquid nitrogen, and stored at −80 °C until use.

Metabolome Analyses—Fourier-transform ion cyclotron resonance mass spectrometry analysis was used to conduct non-targeted metabolomic profiling (6). Targeted metabolic profiling of amino acids, *O*-acetyl-L-serine (OAS), anions, organic acids, and sugars was performed by high performance liquid chromatography and capillary electrophoresis as described (9). Extraction of metabolites was conducted in tripli-

[1] The abbreviations used are: S, sulfur; BL-SOM, batch-learning-self-organizing mapping; OAS, *O*-acetyl-L-serine; GLS(s), glucosinolate(s); PAPS, 3′-phosphoadenosine-5′-phosphosulfate; GST, glutathione *S*-transferase; APS, ATP sulfurylase; Serat, serine acetyltransferase.

cate from each sample. Among 2,123 metabolites detected by targeted and non-targeted analyses, 84 metabolites whose coefficient of variation was greater than 0.9 were eliminated. For each metabolite the logarithm of the ratio of the average signal intensity of S-starved samples to that of the control samples was calculated.

Transcriptome Analyses—The transcriptomes were analyzed using the Agilent Arabidopsis 2 microarray (Agilent Technologies, Palo Alto, CA), which carries 21,500 *Arabidopsis* genes, according to the manufacturer's specifications. Data were acquired using Agilent Feature Extraction software. Normalization of log ratio of expression intensity between S-starved and control samples was carried out based on MA plot (11, 12). Initially, log ratio M_i $[=\log(R_i/G_i)]$ and average of logarithmic intensity A_i $[=(\log R_i + \log G_i)/2]$ were calculated for ith gene. Here, R_i and G_i are differences between mean signal and mean background intensities for Cy5 dye (S-starved sample) and for Cy3 dye (control sample), respectively, obtained by Agilent Feature Extraction software. Normalized log ratio M_i'' was estimated as the difference between M_i and baseline M_i'. Here, using a relation between M_i and A_i, $(M_i = f(A_i) + \epsilon_i, \epsilon_i$ was the difference between M_i and $f(A_i)$ for gene i) by MA plot; the baseline for ith gene was estimated by $M_I' = f(A_i)$. The genes whose signal intensity was regarded as zero were eliminated in the present analysis.

BL-SOM Analyses—BL-SOM analyses were conducted according to Ref. 9. The metabolites and transcripts whose maximum log ratio through time course was <0.1 and minimum log ratio was >−0.1 were eliminated. For ~1,000 metabolites and ~10,000 transcripts left after the elimination, the sum of the square of the 6 log ratio values at 6 time points was set equal to 1 to give relative log ratio values, and all data were combined into a single matrix to be subjected to BL-SOM. ~1,000 metabolites and ~10,000 genes were classified into 40 × 29 cells in the lattice formed by BL-SOM based on the time-dependent pattern of accumulation/expression in leaves in response to −S (Fig. 1A). In the same way, ~1,000 metabolites and ~10,000 genes were classified into 40 × 24 cells in the lattice based on the time-dependent pattern of accumulation/expression in roots in response to −S (Fig. 1B). Each cell contained ~10 metabolites and/or genes on the average. Each cell was colored according to the relative log ratio values of metabolites/genes in it. When all of the relative log ratio values of metabolites/genes in the cell were greater or smaller than the average, the cell was colored in pink or pale blue, respectively. Red and blue indicated that at least one of the relative log ratio values was greater than the average + S.D. or smaller than the average − S.D., respectively.

DNA Cloning Techniques—The gene encoding sulfotransferase 18 (At1g74090) from *A. thaliana* (*AtSOT18*) (for nomenclature, see Ref. 13) does not contain any introns. A 1053-bp cDNA encoding *AtSOT18* was amplified from the λEMBL3 genomic library of *A. thaliana* ecotype Columbia (Clontech) with primer 5′-GGA TCC GAA TCA GAA ACC CTA-3′ extended by a BamHI restriction site and primer 5′-AAG CTT TTT ACC ATG TTC AAG C-3′ extended by a HindIII restriction site. The PCR contained 0.2 mM dNTPs (Roth, Karlsruhe, Germany), 0.4 μM each primer (MWG, Ebersberg, Germany), 1 mM MgCl₂, 0.75 μl of Red *Taq* DNA polymerase (Sigma), and 1 μg of template DNA in a final volume of 50 μl. Before starting the first PCR cycle, the DNA was denatured for 180 s at 94 °C followed by 28 PCR cycles conducted for 60 s at 94 °C, 60 s at 48 °C, and 60 s at 72 °C. The process was finished with an elongation phase of 420 s at 72 °C. The amplified PCR fragment was ligated into the expression vector pQE-30 (Qiagen, Hilden, Germany), and the vector was introduced into the *Escherichia coli* strain XL1-blue.

Expression and Purification of the AtSOT18 Protein in E. coli—The recombinant protein was expressed according to the following protocol. After growth of the *E. coli* culture at 37 °C to an A_{600} of 0.6 in Luria Bertani medium containing ampicillin (100 μg ml⁻¹) (AppliChem, Darmstadt, Germany), induction was carried out for 2 h at 30 °C with 1 mM (final concentration) isopropyl-β-D-1-thiogalactopyranoside (AppliChem). Cell lysis was carried out by adding lysozyme (final concentration 1 mg ml⁻¹) (Roth) and vigorously homogenizing using a glass homogenizer and a pestle. The recombinant AtSOT18 protein was purified under non-denaturing conditions by affinity chromatography with the nickel affinity resin according to the manufacturer's instructions (Qiagen), dialyzed overnight, and used for enzyme activity measurements. The purity of the protein preparation was checked by SDS-polyacrylamide gel electrophoresis and subsequent Coomassie brilliant blue and silver staining.

Substrate Preparation and Enzyme Activity Measurement—The desulfo form of the intact allyl GLS, sinigrin (Sigma), was prepared according to Graser *et al.* (14, 15). The enzyme assay with recombinant AtSOT18 protein was set up in the manner of Glendening and Poulton
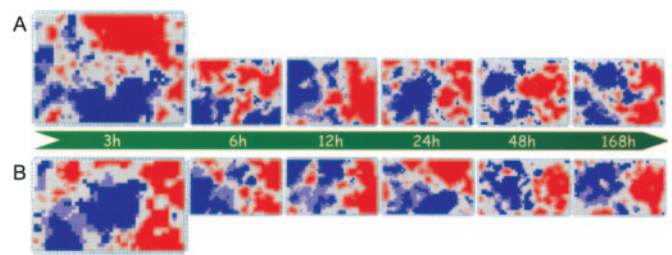

FIG. 1. **BL-SOM analysis of the leaf (A) and root (B) samples.** The map is a lattice comprised of 40 × 29 (A) and 40 × 24 (B) cells. The metabolites and genes were classified into each cell according to their time-dependent pattern of changes in accumulation and expression. The color of each cell indicates the level of induction/repression of the metabolites and genes under S deficiency at the given time points. When all of the relative log ratio values of metabolites/genes in the cell were greater or smaller than the average, the cell was colored in *pink* or *pale blue*, respectively. *Red* and *blue* indicated that at least one of the relative log ratio values was greater than the average + S.D. or smaller than the average − S.D., respectively.

(16). 15 μg of purified recombinant protein was used in each reaction. The 300-μl assays contained: 83 mM Tris/HCl, pH 9.0, 9.2 mM MgCl₂, and 58 μM 3′-phosphoadenosine-5′-phosphosulfate (PAPS; Calbiochem), and 6.2 mM desulfo-allyl GLS. The reaction was started by the addition of PAPS, incubated for 30 min at 30 °C, and stopped by incubation for 20 min at 95 °C. The formation of intact allyl GLS was analyzed by high performance liquid chromatography according to Mellon *et al.* (17). Separation was achieved with a gradient of 0.1% trifluoroacetic acid and acetonitrile with 0.1% trifluoroacetic acid on an RP-C18 column (Supelcosil LC18-DB, 250 × 4.6 mm, 5 μm). The product was identified by its retention time, absorption spectrum, and mass spectrum as described by Mellon *et al.* (17). Quantification of the product formation was done with a calibration curve prepared using authentic sinigrin.

RESULTS AND DISCUSSION

Time-dependent changes in the metabolome and the transcriptome were analyzed in a non-targeted way after *Arabidopsis* plants were subjected to S deprivation for up to 168 h. ~2,000 metabolites detected by targeted and non-targeted analyses and 21,500 transcripts by DNA microarray were used for calculation, and each was expressed as a vector in six-dimensional space, where six components of the vector were six log ratio values of signal intensities under S deficiency compared with those under control condition at six time points.

For integrated analysis of metabolome and transcriptome, we used BL-SOM, which classified the metabolites and the transcripts according to their time-dependent pattern of changes in accumulation and expression. BL-SOM is a sophisticated form of multivariate analyses that can classify metabolites and transcripts into cells on a two-dimensional lattice; those showing similar patterns are clustered into the same or the neighboring cells. After elimination of the metabolites and the transcripts exhibiting almost no change (see "Experimental Procedures"), the sum of the square of 6 log ratio values was set to 1. This procedure made it possible to classify the metabolites and the transcripts by the shape of the time-dependent changing pattern alone and not by the absolute value of the degree of change. All vectors (corresponding to the metabolites and the transcripts) left after elimination were combined into a single matrix to be subjected to BL-SOM. The results are shown in Fig. 1, A (the changes in leaf) and B (in root). By this analysis, sets of metabolites and transcripts with strong correlations were elucidated.

Genes Involved in the Same Metabolic Pathway—Six glucosinolates (GLSs) and four isothiocyanates, which are the degradation products of GLSs, clustered into the cells in the specific regions, suggesting that GLS metabolism is coordinately regulated (Fig. 2, A and B). GLSs are sulfur- and nitrogen-containing secondary metabolites found mainly in the Capparales
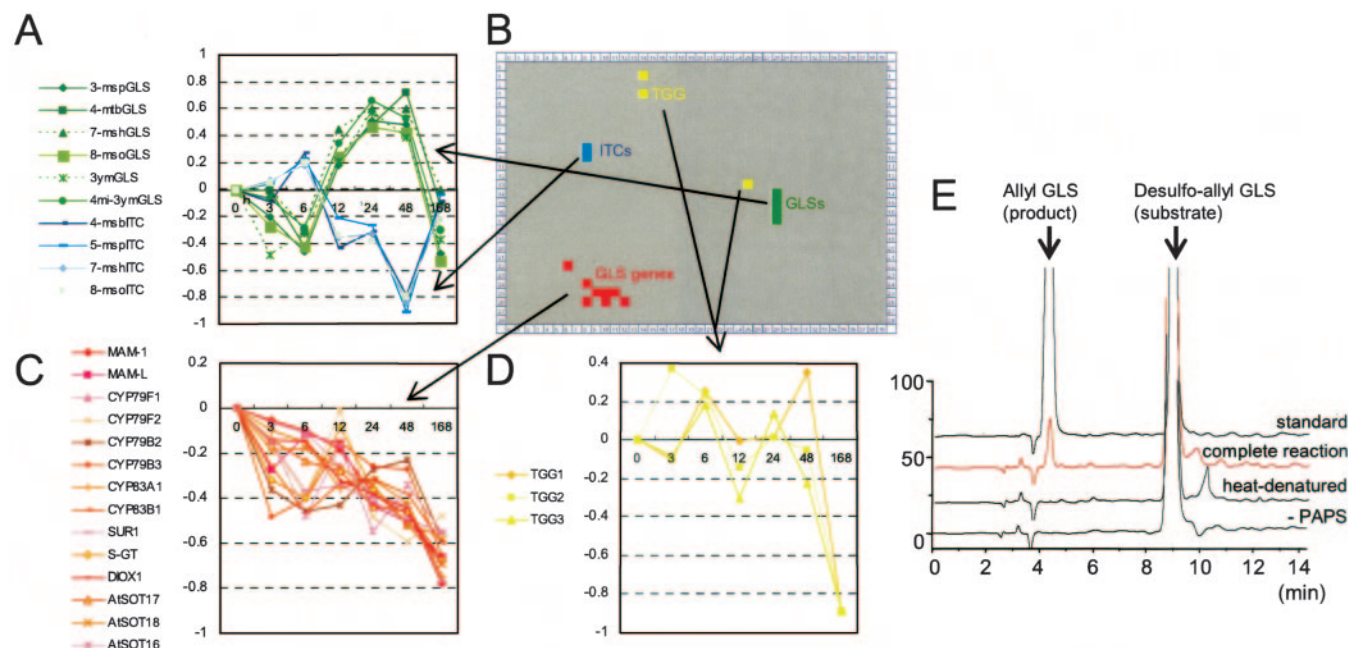
Integrating Omics Sciences for Functional Genomics

FIG. 2. *A*, time-dependent changes in accumulation of GLSs (*green*) and isothiocyanates (*blue*) in leaves under S deficiency. The ordinate scale indicates the relative log ratio values after the sum of squares of 6 log ratio values were set to 1. *3-msp*, 3-methylsulfinylpropyl; *4-mtb*, 4-methylthiolbutyl; *7-msh*, 7-methylsulfinylheptyl; *8-mso*, 8-methylsulfinyloctyl; *3ym*, indol-3-ylmethyl; *4mi-3ym*, 4-methoxyindol-3-ylmethyl; *4-msb*, 4-methylsulfinylbutyl; *5-msp*, 5-(methylsulfinyl)pentyl. *B*, unified map of the leaf samples (the same as Fig. 1*A*) showing the clustering of GLSs (*green*), isothiocyanates (*blue*), GLS biosynthesis genes (*red*), and the degrading enzyme myrosinase genes (*yellow*) into regions. *C* and *D*, time-dependent changes in expression of GLSs biosynthesis genes (*C*) and myrosinase genes *TGG 1–3* (*D*) under S deficiency. The ordinate scale indicates the relative log ratio values. DIOX1 is the gene involved in side chain modification of GLS. *E*, high performance liquid chromatography trace of product of *in vitro* enzymatic assay of AtSOT18 recombinant protein. *Standard*, standards of desulfo-allyl GLS (substrate) and intact allyl GLS (product); *complete reaction*, product formation in the complete reaction mixture; *heat-denatured*, reaction mixture with heat-denatured AtSOT18 protein; −PAPS, reaction mixture with AtSOT18 protein but without PAPS.

order and are important as defense compounds to pathogens and herbivores and as S storage sources (18, 19). They are synthesized from amino acids such as tryptophan, tyrosine, and methionine homologs with elongated side chains. The core biosynthetic pathway of GLS has been elucidated, and the *Arabidopsis* genes encoding most of the enzymes have been identified or at least assumed, except for those encoding the sulfotransferases (20–22) (Fig. 3).

The genes encoding the GLS biosynthesis enzymes, including those encoding the *MAM*, *CYP79*, and *CYP83* families and *SUR1* were clustered into the cells in the same region (Fig. 2, *B* and *C*). It is notable that three putative sulfotransferase genes (At1g74100, At1g18590, and At1g74100 named *At-SOT16*, *AtSOT17*, and *AtSOT18*, respectively, by Klein and Papenbrock (13) were classified into the same region where the known GLS biosynthetic genes were clustered (Fig. 2, *B* and *C*). This result strongly suggested that these putative sulfotransferase genes are also involved in GLS biosynthesis. To confirm the function of these putative sulfotransferase gene products, *in vitro* sulfotransferase assays were conducted using the respective recombinant proteins. The recombinant AtSOT18 protein could convert desulfo-allyl GLS to intact, sulfonated allyl GLS in the presence of PAPS (Fig. 2*E*). The identity of the product was confirmed by mass spectrometry (data not shown). The activity of the gene products of the two other putative sulfotransferase genes (*AtSOT16* and *AtSOT17*) has been analyzed in the same way, giving comparable results with desulfo-allyl GLS (data not shown). These results confirmed that the three sulfotransferase-like genes, which clustered with known GLS biosynthesis genes after BL-SOM analysis, actually encode the proteins catalyzing PAPS-dependent sulfation of desulfo-GLSs to GLSs. During the reviewing process of this publication, Piotrowski *et al.* (23) reached the same conclusion by quite different strategy to ours and reported that these three

genes encode desulfo-GLS sulfotransferase. It was also the case with At1g24100, which was assumed to encode S-glucosyltransferase involved in GLS biosynthesis (24). This gene (Figs. 2 and 3, *S-GT*) was clustered with known GLS biosynthesis genes, supporting the assumption by Petersen *et al.* (24). Recently, the function of this gene was confirmed by Douglas Grubb *et al.* (25). These facts proved that our strategy is effective to elucidate gene functions in the same metabolic pathway at once.

In a similar way, based on the results of BL-SOM the functions of other candidate genes for GLS biosynthesis enzymes (C-S lyase and glutathione *S*-transferase) were also putatively identified. A putative tyrosine aminotransferase gene (At5g36160) and two putative glutathione *S*-transferase (GST) genes (At3g03190 and At1g78370) were also clustered together with the known GLS biosynthesis genes. The putative tyrosine aminotransferase gene At5g36160 may represent a C-S lyase gene involved in GLS biosynthesis. The *SUR1* gene (Fig. 3), whose gene product is the C-S lyase of the core GLS biosynthetic pathway (21), had also been originally misannotated as a tyrosine aminotransferase (26). The *sur1* mutant does not accumulate GLS, at least under normal conditions, suggesting no apparent functional redundancy of C-S lyase of GLS biosynthesis (21). However, this mutant occasionally accumulated a trace level of indol-3-ylmethyl GLS (21), and so the product of At5g36160 might also function as C-S lyase in GLS biosynthesis under different environmental or developmental conditions.

As for the putative GST genes clustering with the other GLS biosynthesis genes, it has been suggested that GST-type enzymes may be involved in an enzyme complex formed by CYP83s and C-S lyase. The S-alkylthiohydroximate formed after CYP83-catalyzed aldoxime oxidation and spontaneous conjugation to cysteine (Fig. 3) is cyclized *in vitro* to form a dead-end product. Hence, metabolic channeling aided by GST-
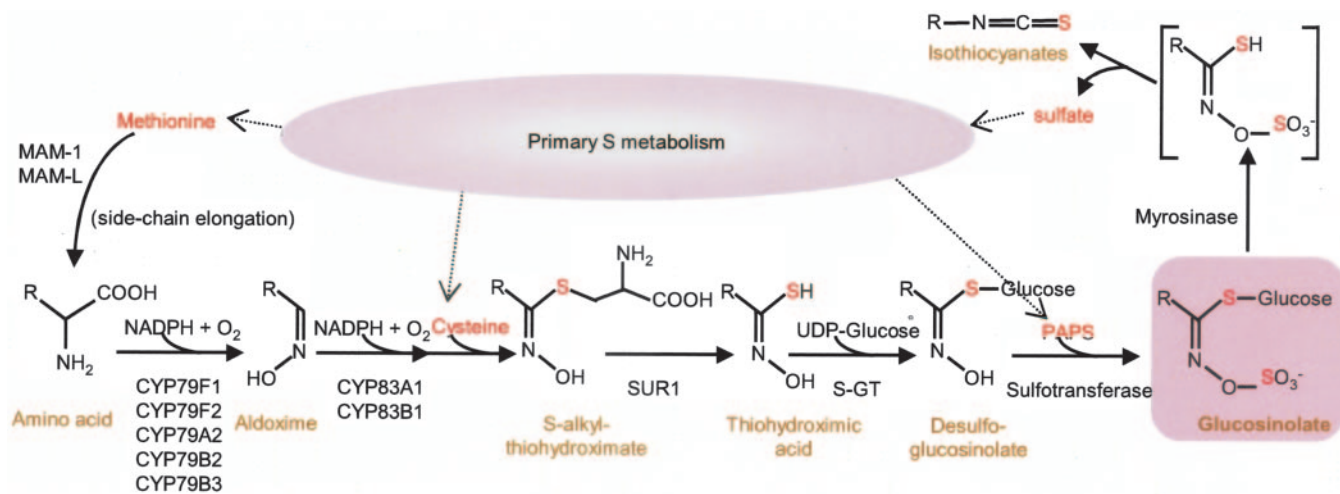
FIG. 3. **Outline of GLS biosynthesis and degradation.** *MAM*, methylthioalkylmalate synthase; *CYP*, cytochrome P450; *SUR1*, a C-S lyase encoded by *SUPERROOT1* gene; *S-GT*, S-glucosyltransferase; *PAPS*, 3′-phosphoadenosine 5′-phosphosulfate. In *Arabidopsis*, methionine and tryptophan are the major precursors of GLSs, but the side chain of methionine is first elongated by a cycle involving MAM-1 and MAM-L. Isoforms of CYP79s and CYP83s catalyze the initial reactions of the core biosynthetic pathway, followed by SUR1, S-GT, and sulfotransferase.
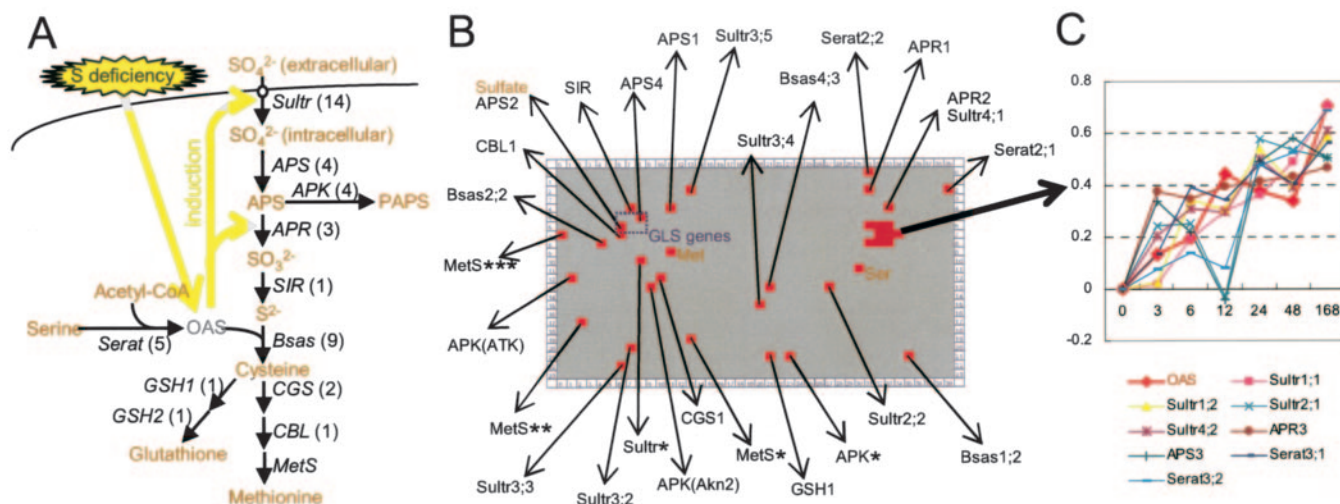

FIG. 4. *A*, outline of primary S metabolism. *Sultr*, sulfate transporter; *APS*, ATP sulfurylase; *APR*, APS reductase; *APK*, APS kinase; *SIR*, sulfite reductase; *Serat*, serine acetyltransferase; *Bsas*, O-acetylserine(thiol)-lyase (a member of the β-substituted alanine synthase family); *CGS*, cystathionine γ synthase; *CBL*, cystathionine β lyase; MetS, methionine synthase; *GSH1*, γ glutamylcysteine synthetase; *GSH2*, glutathione synthetase. The numbers in parentheses indicates the numbers of isoform genes. *B*, unified map of the root samples (the same as Fig. 1*B*) showing clustering of genes encoding enzymes of primary S metabolism. Gene names are shown except for Sultr*, At1g80310; APK*, At5g67520; MetS*, At5g17920; MetS**, At5g20980; and, MetS***, At3g03780. *Blue rectangle* indicates the region where known GLS biosynthesis genes are clustered. *C*, time-dependent changes in OAS accumulation and gene expression. The ordinate scale indicates the relative log ratio values.

type enzymes is postulated *in vivo* to avoid this consequence (21). The two putative GST genes (At3g03190 and At1g78370) could be candidates coding for such an activity. We are now examining the genes clustered with GLSs and isothiocyanates for identification of new factors regulating GLS metabolism.

Regulatory Metabolite O-Acetylserine and Genes under Its Regulation—Primary S metabolism is regulated by modulation of gene expression (see below) and of enzymatic activity (27). For example, the activities of serine acetyltransferases (Fig. 4*A*, *Serat*) are regulated by two mechanisms in an isoform-specific manner: by allosteric feedback inhibition by cysteine and by reversible formation of a protein complex with OAS-(thiol)-lyase (Fig. 4*A*, *Bsas*). The enzymes involved in primary S metabolism are encoded by gene families in *Arabidopsis* (Fig. 4*A*). These genes were scattered on the map (Fig. 4*B*), which was consistent with the fact that primary S metabolism is regulated not only by mRNA accumulation but also by enzymatic activity (27). Among them, however, several genes were clustered together with OAS (Fig. 4, *B* and *C*). Sulfate trans-

porter (*Sultr*) and adenosine phosphosulfate reductase (*APR*) genes are known to be induced by S deficiency. OAS, whose content increases under −S, is considered a positive regulator of this induction mechanism (Fig. 4*A*) (27). The induction of *Sultr* and *APR* genes enables more sulfate ions to be assimilated into cysteine. By BL-SOM, these genes and OAS were clustered into the same region (Fig. 4, *B* and *C*), confirming the previous findings. Among the isoforms of ATP sulfurylase (*APS*) and *Serat* genes, *APS3*, *Serat 3;1* and *Serat3;2* were clustered with OAS (Fig. 4, *B* and *C* and supplemental Fig. 1), suggesting their specific functions under S deficiency (see below).

Besides S assimilation genes, the 12-oxophytodienoate reductase 1 gene and the nitrilase 3 gene are known to be induced under S deficiency and by OAS (10, 28, 29). They were also in the same region where OAS clustered (data not shown). Two putative thioglucosidase genes, which are known to be induced by S deficiency (30), were also in the same region, suggesting that they are coordinately regulated with OAS.

FIG. 5. *A*, outline of anthocyanin bio-synthesis. *PAL*, phenylalanine ammonia lyase; *C4H*, cinnamate 4-hydroxylase; *4CL*, 4-coumarate-CoA ligase; *TT4/CHS*, chalcone synthase; *TT5/CHI*, chalcone isomerase; *TT6/F3H*, flavanone 3-hydroxylase; *TT7/F3'H*, flavonoid 3'-hydroxylase; *TT3/DFR*, dihydroflavonol 4-reductase; *ANS*, anthocyanidin synthase; *AGT 1–3*, anthocyanin glucosyltransferases 1 (At4g14090) (6), 2 (At5g17050) (6), and 3, identified by T. Tohge, Y. Nishiyama, M. Yamazaki, and K. Saito, manuscript in preparation; *AAT*, anthocyanin acyltransferase identified by Y. Nishiyama, T. Tohge, Y. Tanaka, M. Yamazaki, and K. Saito, manuscript in preparation; *TT19/AtGST12*, glutathione *S*-transferase. *B*, unified map of the root samples (the same as Fig. 1*B*) showing clustering of the *PAP1* transcription factor gene and anthocyanin biosynthesis genes. *C*, time-dependent changes in gene expression. The ordinate scale indicates the relative log ratio values TT8/bHLH, At4g09820.
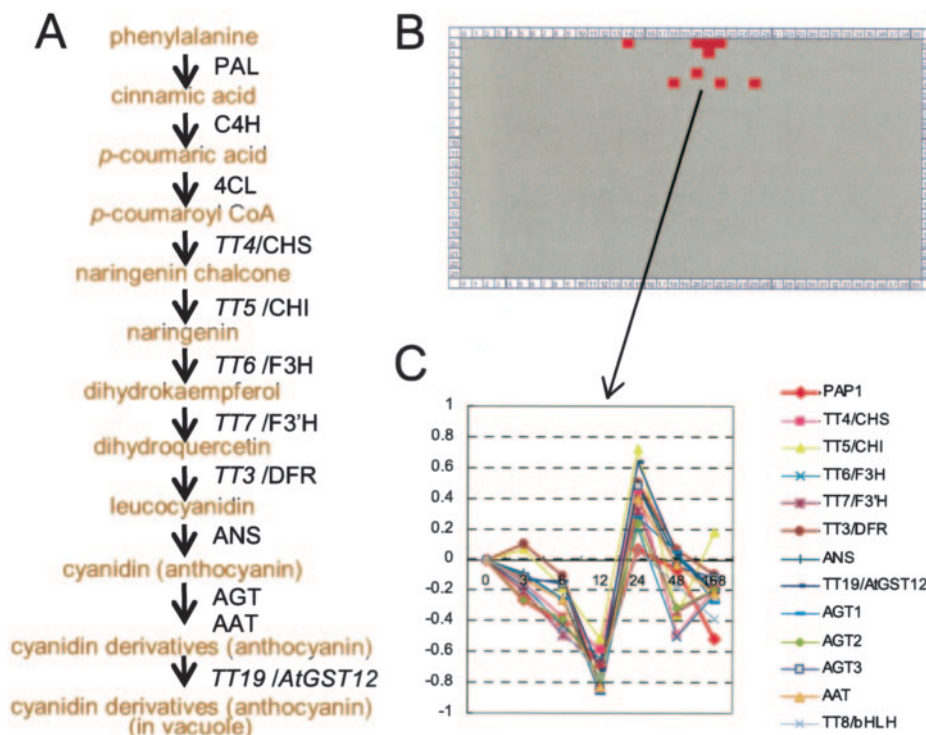
Moreover, the specific function of each member of a gene family could be estimated by this analysis. The ATP sulfurylase isoforms *APS2* and *APS4* among four members of this gene family were clustered with GLS biosynthesis genes (Fig. 4*B* and supplemental Fig. 1), suggesting that these isoforms may be specialized for the biosynthesis of PAPS required for GLS biosynthesis. On the other hand, as mentioned above, *APS3*, which clustered with OAS, may be involved in the enhancement of cysteine biosynthesis under −S.

Transcription Factor and Downstream Genes—PAP1 is a Myb transcription factor that activates phenylpropanoid/flavonoid biosynthesis (31). In an activation-tagged mutant of the *PAP1* gene that overaccumulates red-purple pigment anthocyanins, the biosynthesis genes of flavonoids such as phenylalanine ammonia lyase, chalcone synthase, and dehydroflavonol 4-reductase were induced (31). We clarified that in this mutant line anthocyanin biosynthesis genes were specifically induced among flavonoid biosynthesis genes, suggesting that PAP1 is a transcription factor of anthocyanin biosynthesis genes (6). In this study *PAP1* and the anthocyanin biosynthesis genes were clustered together (Fig. 5, *B* and *C*), indicating that transcription factors and the downstream genes regulated by them can be elucidated by these analyses. Our plant materials subjected to S deficiency did not turn red, indicating anthocyanin biosynthesis was not apparently induced. Nevertheless, *PAP1* and the anthocyanin biosynthesis genes were clustered together. This fact implies that a new regulatory network can be clarified by our strategy even when we do not focus on a specific pathway.

For GLS biosynthesis, several putative transcription factor genes were clustered with GLS biosynthesis genes (data not shown). These genes are the candidate genes controlling GLS biosynthesis.

In the present study, we could find gene-to-gene and metabolite-to-gene networks and could identify a new gene function through integrated analysis of metabolomics and transcriptomics. Our strategy may be useful especially in cases where the gene of interest is functionally redundant and thus no visible changes are observed in knocked-out lines of the gene (32, 33).

This approach is generally applicable not only to plants but also to other organisms, including bacteria and animals, for the identification of novel gene functions.

REFERENCES

1. Urbanczyk-Wochniak, E., Luedemann, A., Kopka, J., Selbig, J., Roessner-Tunali, U., Willmitzer, L., and Fernie, A. R. (2003) *EMBO Rep.* **4,** 989–993
2. Aharoni, A., Keizer, L. C., Bouwmeester, H. J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H. A., Blaas, J., van Houwelingen, A. M., De Vos, R. C., van der Voet, H., Jansen, R. C., Guis, M., Mol, J., Davis, R. W., Schena, M., van Tunen, A. J., and O'Connell, A. P. (2000) *Plant Cell* **12,** 647–662
3. Guterman, I., Shalit, M., Menda, N., Piestun, D., Dafny-Yelin, M., Shalev, G., Bar, E., Davydov, O., Ovadis, M., Emanuel, M., Wang, J., Adam, Z., Pichersky, E., Lewinsohn, E., Zamir, D., Vainstein, A., and Weiss, D. (2002) *Plant Cell* **14,** 2325–2538
4. Goossens, A., Hakkinen, S. T., Laakso, I., Seppanen-Laakso, T., Biondi, S., De Sutter, V., Lammertyn, F., Nuutila, A. M., Soderlund, H., Zabeau, M., Inze, D., and Oksman-Caldentey, K. M. (2003) *Proc. Natl. Acad. Sci. U. S. A.* **100,** 8595–8600
5. Mercke, P., Kappers, I. F., Verstappen, F. W., Vorst, O., Dicke, M., and Bouwmeester, H. J. (2004) *Plant Physiol.* **135,** 2012–2024
6. Tohge, T., Nishiyama, Y., Hirai, M. Y., Yano, M., Nakajima, J., Awazuhara, M., Inoue, E., Takahashi, H., Goodenowe, D. B., Kitayama, M., Noji, M., Yamazaki, M., and Saito, K. (2005) *Plant J.* **42,** 218–235
7. Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T., Mori, H., and Ikemura, T. (2001) *Gene* **276,** 89–99
8. Abe, T., Kanaya, S., Kinouchi, M., Ichiba, Y., Kozuki, T., and Ikemura, T. (2003) *Genome Res.* **13,** 693–702
9. Hirai, M. Y., Yano, M., Goodenowe, D. B., Kanaya, S., Kimura, T., Awazuhara, M., Arita, M., Fujiwara, T., and Saito, K. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 10205–10210
10. Hirai, M. Y., Fujiwara, T., Awazuhara, M., Kimura, T., Noji, M., and Saito, K. (2003) *Plant J.* **33,** 651–663
11. Dudoit, S., Fridlyand, J., and Speed, T. P. (2002) *J. Am. Stat. Assoc.* **97,** 77–87
12. Quackenbush, J. (2002) *Nat. Genet.* **32,** (suppl.) 496–501
13. Klein, M., and Papenbrock, J. (2004) *J. Exp. Bot.* **55,** 1809–1820
14. Graser, G., Schneider, B., Oldham, N. J., and Gershenzon, J. (2000) *Arch. Biochem. Biophys.* **378,** 411–419
15. Graser, G., Oldham, N. J., Brown, P. D., Temp, U., and Gershenzon, J. (2001) *Phytochemistry* **57,** 23–32
16. Glendening, T. M., and Poulton, J. E. (1990) *Plant Physiol.* **94,** 811–818
17. Mellon, F. A., Bennett, R. N., Holst, B., and Williamson, G. (2002) *Anal. Biochem.* **306,** 83–91
18. Wittstock, U., and Gershenzon, J. (2002) *Curr. Opin. Plant Biol.* **5,** 1–9
19. Wittstock, U., and Halkier, B. A. (2002) *Trends Plant Sci.* **7,** 263–270
20. Mikkelsen, M. D., Petersen, B. L., Olsen, C. E., and Halkier, B. A. (2002) *Amino Acids* **22,** 279–295
21. Mikkelsen, M. D., Naur, P., and Halkier, B. A. (2004) *Plant J.* **37,** 770–777
22. Field, B., Cardon, G., Traka, M., Botterman, J., Vancanneyt, G., and Mithen,

R. (2004) *Plant Physiol.* **135,** 828–839

23. Piotrowski, M., Schemenewitz, A., Lopukhina, A., Muller, A., Janowitz, T., Weiler, E. W., and Oecking, C. (2004) *J. Biol. Chem.* **279,** 50717–50725

24. Petersen, B. L., Andreasson, E., Bak, S., Agerbirk, N., and Halkier, B. A. (2001) *Planta* **212,** 612–618

25. Douglas Grubb, C., Zipp, B. J., Ludwig-Muller, J., Masuno, M. N., Molinski, T. F., and Abel, S. (2004) *Plant J.* **40,** 893–908

26. Jones, P. R., Manabe, T., Awazuhara, M., and Saito, K. (2003) *J. Biol. Chem.* **278,** 10291–10296

27. Saito, K. (2004) *Plant Physiol.* **136,** 2443–2450

28. Nikiforova, V., Freitag, J., Kempa, S., Adamik, M., Hesse, H., and Hoefgen, R. (2003) *Plant J.* **33,** 633–650

29. Kutz, A., Muller, A., Hennig, P., Kaiser, W. M., Piotrowski, M., and Weiler, E. W. (2002) *Plant J.* **30,** 95–106

30. Maruyama-Nakashita, A., Inoue, E., Watanabe-Takahashi, A., Yamaya, T., and Takahashi, H. (2003) *Plant Physiol.* **132,** 597–605

31. Borevitz, J. O., Xia, Y., Blount, J., Dixon, R. A., and Lamb, C. (2000) *Plant Cell* **12,** 2383–2394

32. Bino, R. J., Hall, R. D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B. J., Mendes, P., Roessner-Tunali, U., Beale, M. H., Trethewey, R. N., Lange, B. M., Wurtele, E. S., and Sumner, L. W. (2004) *Trends Plant Sci.* **9,** 418–425

33. Weckwerth, W., Loureiro, M. E., Wenzel, K., and Fiehn, O. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 7809–7814