

Toward genome-wide metabolotyping and elucidation of metabolic system: metabolic profiling of large-scale bioresources

Masami Yokota Hirai · Yuji Sawada · Shigehiko Kanaya · Takashi Kuromori · Masatomo Kobayashi · Romy Klausnitzer · Kosuke Hanada · Kenji Akiyama · Tetsuya Sakurai · Kazuki Saito · Kazuo Shinozaki

Received: 14 March 2010 / Accepted: 18 March 2010 / Published online: 6 April 2010
© The Botanical Society of Japan and Springer 2010

Abstract An improvement in plant production is increasingly important for a sustainable human society. For this purpose, understanding the mechanism of plant production, that is, the plant metabolic system, is an immediate necessity. After the sequencing of the Arabidopsis genome, it has become possible to obtain a bird's eye view

of its metabolism by means of omics such as transcriptomics and proteomics. Availability of thousands of transcriptome data points in the public domain has resulted in great advances in the methodology of functional genomics. Metabolome data can be a “gold mine” of biological findings. However, as the total throughput of metabolomics is far lower than that of transcriptomics due to technical difficulties, there is currently no publicly available large-scale metabolome dataset that is comparable in size to the transcriptome dataset. Recently, we established a novel methodology, termed widely targeted metabolomics, which can generate thousands of metabolome data points in a high-throughput manner. We previously conducted a targeted metabolite analysis of large-scale Arabidopsis bioresources, namely transposon-tagged mutants and accessions, to make a smaller dataset of metabolite accumulation. In this paper, we release approximately 3,000 metabolic profiles obtained by targeted analysis for 36 metabolites and discuss the possible regulation of amino acid accumulation.

M. Y. Hirai and Y. Sawada contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10265-010-0337-2) contains supplementary material, which is available to authorized users.

M. Y. Hirai (✉) · Y. Sawada · T. Kuromori · R. Klausnitzer · K. Hanada · K. Akiyama · T. Sakurai · K. Saito · K. Shinozaki
RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
e-mail: myhirai@psc.riken.jp

M. Y. Hirai · Y. Sawada
JST, CREST, 4-1-8 Hon-chou, Kawaguchi, Saitama 332-0012, Japan

M. Y. Hirai
Graduate School of Bioagricultural Sciences,
Nagoya University, Furo-cho, Chikusa-ku,
Nagoya 464-8601, Japan

S. Kanaya
Graduate School of Information Science,
Nara Institute of Science and Technology, 8916-5 Takayamacho,
Ikoma, Nara 630-0192, Japan

M. Kobayashi
RIKEN BioResource Center, 3-1-1 Koyadai,
Tsukuba 305-0074, Japan

K. Saito
Graduate School of Pharmaceutical Sciences, Chiba University,
1-33 Yayoi-cho, Inage-ku, Chiba, Chiba 263-8522, Japan

Keywords Bioresources · Co-accumulation · High-throughput analysis · Large-scale analysis · Metabolome · Mutant

Introduction

Plants provide humans with food, feed, bioenergy, and phytochemicals such as medicinal compounds and biorefinery. Since improvements in plant production are increasingly important for sustaining the human society, understanding the mechanism of plant production, that is, plant metabolic system, is an immediate necessity of plant science.

To understand the regulatory network of metabolism, we should view metabolism as a whole rather than focus on a specific metabolic pathway, enzyme, or gene. In terms of application study, genetic modification of a specific gene without understanding the whole regulatory mechanism often results in an unexpected metabolic change. To achieve expected results, for example, hyper-accumulation of a useful metabolite—we should consider that the plants' responses possibly resulted from metabolic perturbation caused by genetic modification. After the sequencing of the Arabidopsis genome (Arabidopsis Genome Initiative 2000), it has become possible to obtain a bird's eye view of its metabolism via omics such as transcriptomics and proteomics. In addition to omics data contributed by individual researchers, thousands of microarray data that were systematically acquired and deposited in the public domain (Craigon et al. 2004; Goda et al. 2008; Kilian et al. 2007; Schmid et al. 2005) have enabled us to conduct quantitative and statistical analyses in silico and generate data-driven hypothesis without conducting "wet" experiments on a given object. Good examples are found in coexpression analysis-based functional genomics of Arabidopsis. The functions of unknown genes can often be predicted by using genomic sequences (i.e., functional annotation) and coexpression patterns with known genes, even without biological knowledge on a given biological aspect (Saito et al. 2008). Thus, coexpression analysis utilizing a large-scale transcriptome dataset has become a powerful tool for functional genomics and has led to the release of many web-based applications for coexpression analysis in the past 5 years (Akiyama et al. 2008; Goda et al. 2008; Horan et al. 2008; Jen et al. 2006; Manfield et al. 2006; Mutwil et al. 2008; Obayashi et al. 2009; Rawat et al. 2008; Srinivasasainagendra et al. 2008; Steinhauser et al. 2004; Tokimatsu et al. 2005; Toufighi et al. 2005; Zimmermann et al. 2004).

Both metabolome and transcriptome data can be gold mines of information that provide clues to understanding life phenomena. Unfortunately, however, total throughput of metabolome analysis is far lower due to technical difficulties than that of transcriptome analysis in which commercial microarrays and analytical methodologies have almost been established. Metabolites, the targets of metabolomics, are chemically diverse, so no single analytical technique can detect and quantify all metabolites. The Golm Metabolome Database provides public access to custom mass spectral libraries, metabolite profiling experiments from gas chromatography–mass spectrometry (GC–MS) as well as additional information and tools (Kopka et al. 2005). PlantMetabolomic.org was recently developed to provide a portal for accessing metabolomics data generated by different analytical platforms [capillary

electrophoresis–mass spectrometry and different types of GC–MS and liquid chromatography–mass spectrometry (LC–MS)] from multiple laboratories along with the key visualization tools (Bais et al. 2010). AtMetExpress development project, focusing on accumulation patterns of LC–MS-detectable secondary metabolites across 36 Arabidopsis tissues, has been just launched (Matsuda et al. 2010). However, there is no large-scale metabolome dataset currently available in the public domain that is comparable in size to the transcriptome dataset.

Our technical goal is to generate large-scale metabolome datasets comparable in size to the transcriptome dataset and to develop a bioinformatics methodology to mine them for novel findings. As is the case with transcriptome data, metabolome data obtained in various plant organs, at various developmental stages, and under various growth conditions will be informative and give clues to overview of metabolism. We recently established a novel methodology, widely targeted metabolomics, which can generate thousands of metabolic profile data for 734 compounds (as of February 2010) in a high-throughput manner (Sawada et al. 2009a). We have started to acquire metabolic profile data of publicly available large-scale bioresources. Earlier, we conducted targeted metabolite analysis of 2,656 Arabidopsis transposon-tagged mutants and 225 Arabidopsis accessions as a case study in order to make a smaller dataset of metabolite accumulation and learn how to analyze such a dataset. In this paper, we report the results obtained by the targeted metabolite analyses of the two Arabidopsis bioresources. We also introduce widely targeted metabolomics and its possible application.

Materials and methods

Plant materials used for large-scale metabolic profiling

Seeds of transposon-tagged mutants were analyzed for metabolic profiling. From the *Ds* transposon single-copy insertion lines that we previously established (Kuro-mori et al. 2006), we used 2,656 knockout mutants in which the *Ds* transposon was homozygously inserted into the coding regions. For each gene mutant, seeds were harvested from an F3 individual plant described in Kuro-mori et al. (2006). Briefly, F3 seeds were sown on agar-solidified MS medium containing 3% w/v sucrose and 20 $\mu\text{g mL}^{-1}$ hygromycin. After stratification at 4°C for 1 week, plants were grown for 3 weeks under a 16-h light/8-h dark cycle at 22°C, transferred to soil and grown under the same conditions with application of 2,000-fold-diluted Hyponex (Hyponex Japan Co Ltd., Osaka, Japan) until seed maturity.

Seeds of 225 accessions of *Arabidopsis* were also analyzed for metabolic profiling. Seeds of each accession were sown on soil and germinated plants were grown in a greenhouse under natural light at 22°C with application of 2,000-fold-diluted Hyponex until flowering finished. Then seeds were desiccated on plants for 1.5–2 months, harvested, and then stored at 4°C and 20% humidity until use.

Transposon-tagged mutants and accessions used in this study are available at RIKEN BioResource Center (<http://www.brc.riken.jp/inf/en/index.shtml>) through the National Bio-Resource project of the MEXT, Japan.

Construction of *CAM9*-overexpressing plants

The coding region of *CAM9* (At3g51920) was amplified by polymerase chain reaction with forward (5'-CAC-CATGGCGGATGCTTTCACAGA-3') and reverse (5'-CTAATAAGAGGCAGCAATCATC-3') primers using a cDNA library prepared from 2-week-old rosette leaves of *Arabidopsis* Columbia-0 as a template. An entry vector *CAM9*cds-pENTR was prepared by using a pENTR/D-TOPO Cloning Kit (Life Technologies Japan Ltd., Tokyo, Japan) with the amplified *CAM9* coding region (*CAM9*cds). A binary vector *CAM9*cds-pB2GW7 harboring the *CAM9*cds driven by cauliflower mosaic virus (CaMV) 35S RNA promoter was constructed by LR reaction under Gateway[®] technology with *CAM9*cds-pENTR and pB2GW7. A transposon-tagged line 13-3943-1 and its parental line *Ds13* were transformed with *CAM9*cds-pB2GW7 by the floral dip method (Clough and Bent 1998) using *Agrobacterium tumefaciens* strain GV3101. The transformants were selected on agar-solidified 1/2 MS medium containing 50 µM glufosinate. Transgenic plants (T1) were transferred to soil and grown at 23°C under natural and fluorescent light (16-h light/8-h dark cycle) (Hirai et al. 2007). T2 seeds (segregating population) were used for metabolite analysis by means of widely targeted metabolomics (Sawada et al. 2009a).

High-throughput amino acid and glucosinolate analysis

A total of 200 seeds of each independent mutant and accession were homogenized using a mixer mill MM 200 (Retsch) in 80 µL of extraction buffer [40% acetonitrile in H₂O with 25 µM hydroxyphenyl-glucosinolate (GSL) and 50 µM norleucine as internal standards]. The extracts were diluted with 500 µL of LC-MS grade H₂O and centrifuged (1,000 g) for 5 min. The supernatants were filtered through a CAPTIVA 0.45 µm filter (Varian) and subjected to ultra performance liquid chromatography (UPLC) (Waters)-quadrupole MS analysis. Thirty-six metabolites (17 amino acids, 18 GSLs, and 1 flavonoid; Table S1) were separated on UPLC through a reverse phase column (50 × 2.1 mm,

HSS T3 1.8 µm; Waters) (Table S1) and detected using ZQ mass spectrometers (Waters) with an electrospray ionization (ESI) interface (positive mode for amino acid analysis; negative mode for glucosinolates and flavonoid analysis; capillary voltage, +3.0 and -3.0 keV; cone voltage, +25 and -40 V; source temperature, 120°C; desolvation temperature, 350°C; cone gas flow, 50 L/h; desolvation gas flow, 600 L/h). MassLynx software version 4.0 (Waters) was used to control all instruments and calculate peak areas. The analytical error of this analysis using UPLC-quadrupole MS is less than 10% (data not shown); thus, we did not perform an analytical replicate in this study.

Data analysis using R

The peak area of metabolites was divided by that of the internal standards (norleucine for amino acids; hydroxyphenyl-GSL for glucosinolates and flavonoid) after missing values were replaced with 0. Spearman's rank-correlation coefficient (SCC) was calculated by R (<http://www.r-project.org>). The log₂-transformed data were used for drawing histograms by R.

Results

Analysis of amino acids and GSLs in seeds of transposon-tagged mutants and accessions

As a case study, we chose to clarify accumulation profiles of amino acids and GSLs as examples of primary and secondary metabolites, respectively. GSLs, a group of sulfur- and nitrogen-containing secondary metabolites found mainly in Brassicaceae plants, are synthesized from 1 of 8 amino acids including Met and Trp (Grubb and Abel 2006; Halkier and Gershenzon 2006). In *Arabidopsis*, the biosynthetic pathways of Met- and Trp-derived GSLs are well studied. In this study, we analyzed 17 protein amino acids, 15 Met-derived GSLs, 3 Trp-derived GSLs and 1 flavonoid (Table S1) in mature dry seeds of 2,656 transposon-tagged mutants and 225 *Arabidopsis* accessions by means of LC-MS-based high-throughput targeted metabolite analyses (see “Materials and methods”).

Transposon-tagged mutants have been established as public bioresources and are utilized for generating a phenome database in which visible phenotypic changes are described (Kuromori et al. 2004, 2006). These mutants originated from 1 of 5 parental transgenic lines (*Ds11*, *Ds12*, *Ds13*, *Ds15*, and *Ds16*) carrying the *Ds* element. In the process of establishing each mutant line, *Ds* has been transposed into a single locus on the genome and fixed homozygously in successive generations (Kuromori et al. 2004).

Arabidopsis accessions were collected from around the world by a coordinated effort of Arabidopsis researchers and are now maintained at the RIKEN BioResource Center (<http://sassc.epd.brc.riken.jp/sassc/top.php?mode=general>).

Metabolite accumulation data obtained from both bio-resources are included in Tables S2 and S3. In this study, we further analyzed the data obtained from transposon-tagged lines.

Global profiles of amino acid and GSL accumulation in transposon-tagged mutants

Transposon-tagged genes covered all 23 main functional categories defined in our classification system, Arabidopsis Gene Classifier vol. 1.4 (<http://kanaya.naist.jp/GeneClassifier/top.jsp?fn=arabidopsis>) (Takahashi et al. 2009).

To understand the global tendency of metabolite accumulation patterns, we first analyzed the distribution of each metabolite's content across all 2,656 transposon-tagged mutants (Fig. S1). In the case of amino acids, a single peak was observed, indicating that the amino acid content varied continuously. On the other hand, in the case of several GSLs such as MSc8, two (or three) peaks were observed. However, when the distribution was analyzed across each series of mutants derived from the same parental transgenic line, a single peak was observed (Fig. S1). The seeds of transposon-tagged mutants have been collected under almost the same growth conditions but in different growth chambers depending on their parental line; namely, the mutants that originated from some parent lines were grown in a growth chamber, whereas the mutants from the other parental lines were grown in another growth chamber. We consider that bi- or trimodal distribution of some of the GSL contents resulted from the slight difference in growth conditions under which the seeds were harvested since GSL content varies easily depending on growth conditions (data not shown). Another possibility is that the mutants have different features originated from the parental lines. Then, for further statistical analyses, the mutants that originated from different parental lines were treated separately.

To understand the trend of metabolite accumulation patterns, correlation between metabolite contents was analyzed by calculating Spearman's correlation coefficient (SCC) (Fig. 1). As mentioned above, since some metabolites exhibited bimodal distribution depending on which parent the mutants were derived from, the mutants were classified into five subgroups according to their parental lines, and the correlation coefficient was separately calculated within each subgroup. Then five SCC values were averaged for each metabolite combination. Among 36 metabolites analyzed, the content of BOc4, 1MOI3M, and 4MOI3M (for abbreviations, see Table S1)

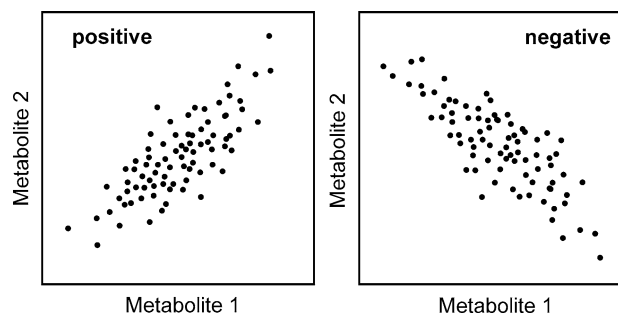


Fig. 1 Schematic diagram of the co-accumulation relationship between metabolite pairs. *Horizontal* and *vertical* axes indicate metabolite content. Each *dot* represents a mutant with a given genotype (this study) or an individual plant with the same genotype (Kusano et al. 2007; Weckwerth et al. 2004). The left and right panels indicate positive (SCC > 0) and negative (SCC < 0) correlations, respectively

Table 1 Metabolite combinations co-accumulated in seeds of transposon-tagged mutants

Metabolite 1	Metabolite 2	Average SCC	RSD (%)
Ile	Val	0.91	1.8
Leu	Val	0.90	3.3
Leu	Ile	0.89	1.4
MTc6	MTc5	0.89	5.4
Arg	Lys	0.78	6.9
His	Val	0.77	9.9
His	Ile	0.76	14.7
His	Leu	0.75	11.5
MSc3	MTc3	0.75	14.9
MSc6	MSc4	0.74	19.5
His	Lys	0.72	15.4
Arg	His	0.70	13.0
MSc7	MSc4	0.70	10.3
Thr	Pro	0.70	7.3
Tyr	His	0.69	11.8
Tyr	Lys	0.68	10.3
Thr	Gly	0.66	17.8
Lys	Val	0.66	18.7

The average Spearman's correlation coefficient (SCC) of five subgroups was ranked. Metabolite combinations ranking in the top 5% among 528 combinations with relative standard deviation (RSD) less than 20% were shown

was very low and often fell below detection limits. These metabolites were then omitted from further analyses and 528 combinations among 33 metabolites were analyzed. All SCC values were summarized in Table S4. The metabolite combinations that exhibited higher positive correlations, ranked in the top 5%, and had relative standard deviation (RSD) values of less than 20% are listed in Table 1.

An example of mutants which exhibited hyper-accumulation of amino acids

Each transposon-tagged mutant we analyzed had a transposon in the coding region of a given gene. If a gene involved in metabolism was knocked out or knocked down by insertion of a transposon, the accumulation pattern of the amino acids and/or GSLs might be changed by direct or indirect effects. We analyzed 13-3943-1 as a model mutant that exhibited drastic changes in metabolite accumulation. This mutant exhibited hyper-accumulation of branched-chain amino acids (Leu, Ile, and Val) in seeds (Fig. 2). Hyper-accumulation of these amino acids was not observed in rosette leaves (data not shown). In this mutant, the transposon was inserted in At3g51920 (*CALMODULIN 9*, *CAM9*). To confirm that the metabolic change was caused by knocking out *CAM9*, 13-3943-1 (hereafter referred to as *cam9*) was complemented with the *CAM9* gene driven by CaMV 35S RNA promoter. In complemented lines, the contents of Leu, Ile, and Val in seeds were recovered to similar levels to a parental plant, Ds13 (Fig. 2). In addition, when *CAM9* was overexpressed in a parental line Ds13, accumulation of Leu, Ile, and Val slightly decreased (0.2- to 0.8-fold) (Fig. 2). These results indicated that knocking out *CAM9* actually resulted in hyper-accumulation

of these amino acids, suggesting that an unknown function of *CAM9* is involved in amino acid accumulation.

Discussion

Analysis of metabolic profiles in seeds of transposon-tagged mutants

In this study, we analyzed the accumulation profiles of 36 metabolites in seeds of 2,656 transposon-tagged lines and 225 *Arabidopsis* accessions. Our main purpose was to get an overview of the accumulation pattern of amino acids and GSLs as the first step to large-scale metabolome analysis, which is followed by statistical analysis and mathematical modeling to understand metabolism as a system. We selected mature seeds as the first target of metabolite analysis since the seed is the most important sink of metabolites as nutrient reservoir for the next generation to grow during infancy. In addition, the seed is a major organ in terms of GSL accumulation.

Although metabolite accumulation is genetically regulated, there is fluctuation in the absolute content of each metabolite among individual plants of the same genotype, even under the same growth conditions. However, when paying attention to specific metabolite pairs, positive or negative correlation between their contents can often be found in the data obtained from a number of individual plants grown under the same growth conditions (Kusano et al. 2007; Weckwerth et al. 2004) (Fig. 1, each dot represents an individual plant with the same genotype). The correlation relationship between a given metabolite pair could change in different genotypes (Kusano et al. 2007; Weckwerth et al. 2004). For example, a positive correlation between the contents of malate and fumarate in leaves was observed in wild-type and *transparent testa 4* (*tt4*) mutant of *Arabidopsis* but not in *methionine over-accumulation 1* (*mtol*) mutant (Kusano et al. 2007). In this study, we observed a correlation relationship between metabolite pairs in seeds conserved throughout genotypes (Fig. 1). In addition, although the maternal plants were grown under almost the same growth conditions, the seeds analyzed were not always harvested side by side at the same time. Therefore, the correlation relationship observed in our analysis was conserved irrespective of changes in the microenvironment.

Higher positive correlation was observed between amino acids and between Met-GSLs (Table S4). Co-accumulation of Met-GSLs could be explained by the fact that most of the genes involved in Met-GSLs biosynthesis are coordinately regulated by specific transcription factors, Myb28, Myb29, and Myb76 (Beekwilder et al. 2008; Gigolashvili et al. 2007a, b; Hirai et al. 2007; Malitsky

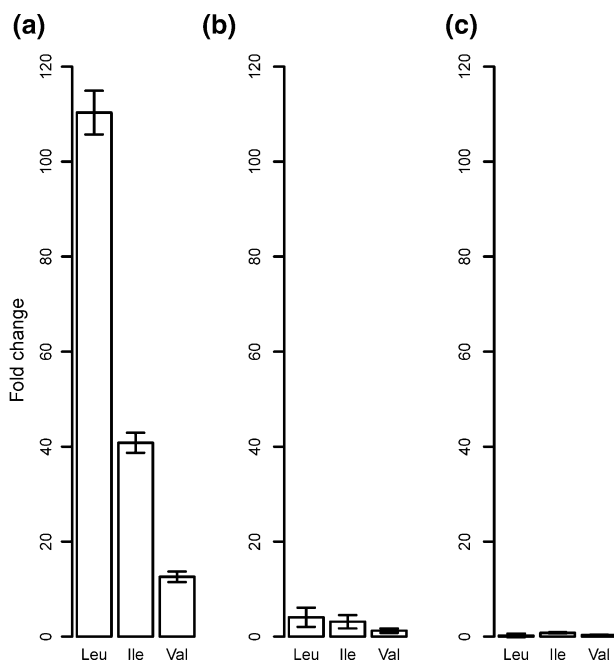


Fig. 2 Accumulation of branched-chain amino acids in seeds. The average contents of Leu, Ile, and Val in a parental line Ds13 ($n = 4$) were set to 1. Vertical axes indicate relative contents of these amino acids in **a** a knockout mutant *cam9* ($n = 5$), **b** complementary lines (35S::*CAM9*/*cam9*), and **c** *CAM9*-overexpressing lines (35S::*CAM9*/Ds13). For **b** and **c**, amino acid contents in the seeds harvested from three independent T1 plants were separately analyzed and averaged

et al. 2008; Sønderby et al. 2007). On the other hand, co-accumulation of amino acids cannot be simply explained from the viewpoint of metabolic pathway, i.e., metabolite pairs that were not directly connected in the metabolic pathway (e.g., His and Tyr) exhibited a co-accumulation relationship. We then further analyzed our data focusing on amino acids. Figure S3 shows the SCC values for all amino acid pairs. Co-accumulation relationships are schematically indicated in Fig. S4. These data suggest that accumulations of Val, Ile, Leu, Arg, Lys, Tyr, and His are positively correlated to each other while accumulations of Met, Asp, Ala, Trp, and Glu are independent of those of other amino acids in these series of mutants. In our previous integration analysis of transcriptome and metabolome data obtained from nutrient-starved Arabidopsis, we observed that changes in amino acid accumulation were not simply explained by the changes in expression of their biosynthetic genes (Hirai et al. 2004). Co-accumulation of some amino acids observed in this study might be achieved by post-translational regulation such as feedback inhibition of biosynthetic enzymes and/or other mechanisms in which unknown genes are involved. Given that some genes with unknown functions play a role in maintaining metabolic balance, knocking out the gene would result in metabolic imbalance. In this study, we found many outliers in which some metabolites accumulated to unexpected levels judging from the frequency distribution of the metabolite. One example is a mutation in *CAM9*, which resulted in hyper-accumulation of branched-chain amino acids in seeds only. *CAM9* encodes a divergent member of calmodulin, an EF-hand family of Ca^{2+} -binding proteins. Recently, it was reported that the lines possessing mutations in *CAM9* presented a hypersensitive response to ABA and enhanced tolerance to salt stress and water deficits (Magnan et al. 2008). Our result suggested an as-yet-unknown relationship between amino acid metabolism and ABA-mediated stress responses. Thus, this study can lead to discovery of genes that play important roles in metabolic homeostasis.

Our current technology and future perspective

MS- and NMR-based metabolomics have been developed over the past decade in order to establish non-targeted analysis procedures (Werner et al. 2008a, b). Due to the diverse nature of the chemical and physical properties of these metabolites, metabolome analysis can only be achieved through the integration of the obtained data using various techniques. Although non-targeted metabolomics enables us to identify the broad metabolite profiles of samples and to find novel metabolites that can be used as biomarkers (Glinski and Weckwerth 2006; Saito and Matsuda 2010), there are still a lot of difficulties to be settled. For example, it is technically difficult and

extremely time-consuming to merge all of the obtained data in different formats by different instruments and to identify the unknown metabolites (Werner et al. 2008b). Hence, the application of non-targeted metabolomics to thousands of biological samples is not yet practical. This is the reason why we need an alternative methodology for targeted but high-throughput metabolomics.

We recently established a highly selective and sensitive procedure that utilizes the multiple reaction monitoring (MRM) mode of UPLC-quadrupole MS. MRM enables high sensitivity, ease of reproducibility, and a broad dynamic range of analysis (Unwin et al. 2005). Briefly, we first optimized the following analytical conditions for each of the 734 authentic compounds (as of February 2010) (Sawada et al. 2009a): retention time (RT) of UPLC separation and combination of polarity, precursor ion mass, fragment (product) ion mass, cone voltage and collision voltage for MS/MS detection, which we call MRM condition. As more than one compound occasionally gave the same optimal condition, we finally obtained 574 sets of RT and MRM conditions (RT–MRM sets) for 734 compounds (as of February 2010). For metabolic profiling of the plant samples, we measured areas of the peaks observed using the abovementioned 574 RT–MRM sets. In the case of Arabidopsis leaves, we observed that 217 out of 574 RT–MRM sets reproducibly gave detectable peaks (Table S5). This analysis, which we call widely targeted metabolomics, was successfully applied to functional genomics of Arabidopsis by analyzing metabolic profiles in knockout mutants of genes involved in GSL biosynthesis (Sawada et al. 2009b, c).

Owing to high-throughput (e.g., 600 Arabidopsis metabolic profiles per week), this methodology has broadened the application field of metabolomics. We are analyzing various kinds of bioresources such as knockout lines, accessions of Arabidopsis, and wild species of crop plants as well as field-grown plants from agronomic and ecological viewpoints (to be published elsewhere). Acquired large-scale datasets would be novel gold mines of biological findings. Our next challenge is to develop a novel strategy for analysis of metabolome data, i.e., a novel tool for treasure hunting.

Acknowledgments This work was supported in part by the Core Research for Evolutional Science and Technology (CREST) of the Japan Science and Technology Agency (Project name: “Elucidation of Amino Acid Metabolism in Plants Based on Integrated Omics Analyses”).

References

Akiyama K, Chikayama E, Yuasa H, Shimada Y, Tohge T, Shinozaki K, Hirai MY, Sakurai T, Kikuchi J, Saito K (2008) PRIME: a

- web site that assembles tools for metabolomics and transcriptomics. *In Silico Biol* 8:339–345
- Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Bais P, Moon SM, He K, Leitao R, Dreher K, Walk T, Sucaet Y, Barkan L, Wohlgenuth G, Roth MR, Wurtele ES, Dixon P, Fiehn O, Lange BM, Shulaev V, Sumner LW, Welti R, Nikolau BJ, Rhee SY, Dickerson JA (2010) PlantMetabolomics.org: a web portal for plant metabolomics experiments. *Plant Physiol*, pp 109.151027
- Beekwilder J, van Leeuwen W, van Dam NM, Bertossi M, Grandi V, Mizzi L, Soloviev M, Szabados L, Molthoff JW, Schipper B, Verbocht H, de Vos RC, Morandini P, Aarts MG, Bovy A (2008) The impact of the absence of aliphatic glucosinolates on insect herbivory in *Arabidopsis*. *PLoS ONE* 3:e2068
- Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J* 16:735–743
- Craigon DJ, James N, Okyere J, Higgins J, Jotham J, May S (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res* 32:D575–D577
- Gigolashvili T, Engqvist M, Yatushevich R, Muller C, Flugge UI (2007a) HAG2/MYB76 and HAG3/MYB29 exert a specific and coordinated control on the regulation of aliphatic glucosinolate biosynthesis in *Arabidopsis thaliana*. *New Phytol* 177:627–642
- Gigolashvili T, Yatushevich R, Berger B, Muller C, Flugge U-I (2007b) The R2R3-MYB transcription factor HAG1/MYB28 is a regulator of methionine-derived glucosinolate biosynthesis in *Arabidopsis thaliana*. *Plant J* 51:247–261
- Gliniski M, Weckwerth W (2006) The role of mass spectrometry in plant systems biology. *Mass Spectrom Rev* 25:173–214
- Goda H, Sasaki E, Akiyama K, Maruyama-Nakashita A, Nakabayashi K, Li W, Ogawa M, Yamauchi Y, Preston J, Aoki K, Kiba T, Takatsuto S, Fujioka S, Asami T, Nakano T, Kato H, Mizuno T, Sakakibara H, Yamaguchi S, Nambara E, Kamiya Y, Takahashi H, Hirai MY, Sakurai T, Shinozaki K, Saito K, Yoshida S, Shimada Y (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J* 55:526–542
- Grubb CD, Abel S (2006) Glucosinolate metabolism and its control. *Trends Plant Sci* 11:89–100
- Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. *Annu Rev Plant Biol* 57:303–333
- Hirai MY, Yano M, Goodenowe DB, Kanaya S, Kimura T, Awazuhara M, Arita M, Fujiwara T, Saito K (2004) Integration of transcriptomics and metabolomics for understanding of global responses to nutritional stresses in *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 101:10205–10210
- Hirai MY, Sugiyama K, Sawada Y, Tohge T, Obayashi T, Suzuki A, Araki R, Sakurai N, Suzuki H, Aoki K, Goda H, Nishizawa OI, Shibata D, Saito K (2007) Omics-based identification of *Arabidopsis* Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc Natl Acad Sci USA* 104:6478–6483
- Horan K, Jang C, Bailey-Serres J, Mittler R, Shelton C, Harper JF, Zhu J-K, Cushman JC, Gollery M, Girke T (2008) Annotating genes of known and unknown function by large-scale coexpression analysis. *Plant Physiol* 147:41–57
- Jen CH, Manfield IW, Michalopoulos I, Pinney JW, Willats WG, Gilmartin PM, Westhead DR (2006) The *Arabidopsis* co-expression tool (ACT): a WWW-based tool and database for microarray-based gene expression analysis. *Plant J* 46:336–348
- Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K (2007) The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. *Plant J* 50:347–363
- Kopka J, Schauer N, Krueger S, Birkemeyer C, Usadel B, Bergmuller E, Dormann P, Weckwerth W, Gibon Y, Stitt M, Willmitzer L, Fernie AR, Steinhauser D (2005) GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21:1635–1638
- Kuromori T, Hirayama T, Kiyosue Y, Takabe H, Mizukado S, Sakurai T, Akiyama K, Kamiya A, Ito T, Shinozaki K (2004) A collection of 11,800 single-copy Ds transposon insertion lines in *Arabidopsis*. *Plant J* 37:897–905
- Kuromori T, Wada T, Kamiya A, Yuguchi M, Yokouchi T, Imura Y, Takabe H, Sakurai T, Akiyama K, Hirayama T, Okada K, Shinozaki K (2006) A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of *Arabidopsis*. *Plant J* 47:640–651
- Kusano M, Fukushima A, Arita M, Jonsson P, Moritz T, Kobayashi M, Hayashi N, Tohge T, Saito K (2007) Unbiased characterization of genotype-dependent metabolic regulations by metabolomic approach in *Arabidopsis thaliana*. *BMC Syst Biol* 1:53
- Magnan F, Ranty B, Charpentreau M, Sotta B, Galaud J-P, Aldon D (2008) Mutations in AtCML9, a calmodulin-like protein from *Arabidopsis thaliana*, alter plant responses to abiotic stress and abscisic acid. *Plant J* 56:575–589
- Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A (2008) The transcript and metabolite networks affected by the two clades of *Arabidopsis* glucosinolate biosynthesis regulators. *Plant Physiol* 148:2021–2049
- Manfield IW, Jen CH, Pinney JW, Michalopoulos I, Bradford JR, Gilmartin PM, Westhead DR (2006) *Arabidopsis* Co-expression Tool (ACT): web server tools for microarray-based gene expression analysis. *Nucleic Acids Res* 34:W504–W509
- Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K (2010) AtMetExpress development: a phytochemical atlas of *Arabidopsis thaliana* development. *Plant Physiol* 152:566–578
- Mutwil M, Obro J, Willats WGT, Persson S (2008) GeneCAT—novel webtools that combine BLAST and co-expression analyses. *Nucl Acids Res* 36:W320–W326
- Obayashi T, Hayashi S, Saeki M, Ohta H, Kinoshita K (2009) ATTED-II provides coexpressed gene networks for *Arabidopsis*. *Nucl Acids Res* 37:D987–D991
- Rawat A, Seifert G, Deng Y (2008) Novel implementation of conditional co-regulation by graph theory to derive co-expressed genes from microarray data. *BMC Bioinformatics* 9:S7
- Saito K, Matsuda F (2010) Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol* 61
- Saito K, Hirai MY, Yonekura-Sakakibara K (2008) Decoding genes with coexpression networks and metabolomics—'majority report by precogs'. *Trends Plant Sci* 13:36–43
- Sawada Y, Akiyama K, Sakata A, Kuwahara A, Otsuki H, Sakurai T, Saito K, Hirai MY (2009a) Widely targeted metabolomics based on large-scale MS/MS data for elucidating metabolite accumulation patterns in plants. *Plant Cell Physiol* 50:37–47
- Sawada Y, Toyooka K, Kuwahara A, Sakata A, Nagano M, Saito K, Hirai MY (2009b) *Arabidopsis* bile acid: sodium symporter family protein 5 is involved in methionine-derived glucosinolate biosynthesis. *Plant Cell Physiol* 50:1579–1586
- Sawada Y, Kuwahara A, Nagano M, Narisawa T, Sakata A, Saito K, Hirai MY (2009c) Omics-based approaches to methionine side chain elongation in *Arabidopsis*: characterization of the genes encoding methylthioalkylmalate isomerase and methylthioalkylmalate dehydrogenase. *Plant Cell Physiol* 50:1181–1190
- Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37:501–506

- Sønderby IE, Hansen BG, Bjarnholt N, Ticconi C, Halkier BA, Kliebenstein DJ (2007) A systems biology approach identifies a R2R3 Myb gene subfamily with distinct and overlapping functions in regulation of aliphatic glucosinolates. *PLoS ONE* 2:e1322
- Srinivasasainagendra V, Page GP, Mehta T, Coulibaly I, Loraine AE (2008) CressExpress: a tool for large-scale mining of expression data from *Arabidopsis*. *Plant Physiol* 147:1004–1016
- Steinhauser D, Usadel B, Luedemann A, Thimm O, Kopka J (2004) CSB.DB: a comprehensive systems-biology database. *Bioinformatics* 20:3647–3651
- Takahashi H, Kawazoe M, Wada M, Hirai A, Nakamura K, Altaf-Ul-Amin M, Sawada Y, Hirai MY, Kanaya S (2009) KNApSAcK gene classification system for *Arabidopsis thaliana*: comparative genomic analysis of unicellular to seed plants. *Plant Biotechnol* 26:509–516
- Tokimatsu T, Sakurai N, Suzuki H, Ohta H, Nishitani K, Koyama T, Umezawa T, Misawa N, Saito K, Shibata D (2005) KaPPA-view: a web-based analysis tool for integration of transcript and metabolite data on plant metabolic pathway maps. *Plant Physiol* 138:1289–1300
- Toufighi K, Brady SM, Austin R, Ly E, Provart NJ (2005) The botany array resource: e-Northerns, expression angling, and promoter analyses. *Plant J* 43:153–163
- Unwin RD, Griffiths JR, Leverenz MK, Grallert A, Hagan IM, Whetton AD (2005) Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics* 4:1134–1144
- Weckwerth W, Loureiro ME, Wenzel K, Fiehn O (2004) Differential metabolic networks unravel the effects of silent plant phenotypes. *Proc Natl Acad Sci USA* 101:7809–7814
- Werner E, Croixmarie V, Umbdenstock T, Ezan E, Chaminade P, Tabet JC, Junot C (2008a) Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. *Anal Chem* 80:4918–4932
- Werner E, Heilier JF, Ducruix C, Ezan E, Junot C, Tabet JC (2008b) Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends. *J Chromatogr B Analyt Technol Biomed Life Sci* 871:143–163
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W (2004) GENEVESTIGATOR. *Arabidopsis* microarray database and analysis toolbox. *Plant Physiol* 136:2621–2632