PCP
PLANT & CELL PHYSIOLOGY

PRIMe Update: Innovative Content for Plant Metabolomics and Integration of Gene Expression and Metabolite Accumulation

Tetsuya Sakurai[1,*], Yutaka Yamada[1], Yuji Sawada[1], Fumio Matsuda[1,2,4], Kenji Akiyama[1], Kazuo Shinozaki[1], Masami Yokota Hirai[1] and Kazuki Saito[1,3]

[1]RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, 230-0045 Japan
[2]Organization of Advanced Sciences and Technology, Kobe University, 1-1 Rokkodaicho, Nada-ku, Kobe, 657-8501 Japan
[3]Graduate School of Pharmaceutical Sciences, Chiba University, Inohana 1-8-1, Chuo-ku, Chiba, 260-8675 Japan
[4]Present address: Department of Bioinformatic Engineering, Osaka University, 1-5 Yamadaoka, Suita, Osaka, 565-0871 Japan
*Corresponding author: E-mail, stetsuya@psc.riken.jp; Fax, +81-45-503-9489.
(Received October 31, 2012; Accepted December 15, 2012)

PRIMe (http://prime.psc.riken.jp/), the Platform for RIKEN Metabolomics, is a website that was designed and implemented to support research and analyses ranging from metabolomics to transcriptomics. To achieve functional genomics and annotation of unknown metabolites, we established the following PRIMe contents: MS2T, a library comprising >1 million entries of untargeted tandem mass spectrometry (MS/MS) data of plant metabolites; AtMetExpress LC-MS, a database of transcriptomics and metabolomics approaches in Arabidopsis developmental stages (AtMetExpress Development LC-MS) and a data set of the composition of secondary metabolites among 20 Arabidopsis ecotypes (AtMetExpress 20 ecotypes LC-MS); and ReSpect, hybrid reference MS/MS data resources (acquisitions and literature). PRIMeLink is a new web application that allows access to the innovative data resources of PRIMe. The MS2T library was generated from a set of MS/MS spectra acquired using the automatic data acquisition function of mass spectrometry. To increase the understanding of mechanisms driving variations in metabolic profiles among plant tissues, we further provided the AtMetExpress Development LC-MS database in PRIMe, facilitating the investigation of relationships between gene expression and metabolite accumulation. This information platform therefore provides an integrative analysis resource by linking Arabidopsis transcriptome and metabolome data. Moreover, we developed the ReSpect database, a plant-specific MS/MS data resource, which allows users to identify candidate structures from the suite of complex phytochemical structures. Finally, we integrated the three databases into PRIMeLink and established a walk-through link between transcriptome and metabolome information. PRIMeLink offers a bi-directional searchable function, from the gene and the metabolite perspective, to search for targets seamlessly and effectively.

Keywords: Arabidopsis • Database • Metabolome • Multiomics • Transcriptome.

Abbreviations: AGI, Arabidopsis Genome Initiative; BL-SOM, batch-learning self-organizing map; GO, Gene Ontology; LC-MS, liquid chromatography–mass spectrometry; *m/z*, mass-to-charge ratio; MS/MS, tandem mass spectrometry; MS2T, MS/MS spectral tag; Q-TOF-MS, quadrupole time-of-flight mass spectrometer; *SHT*, spermidine hydroxycinnamoyl transferase; TAIR, The Arabidopsis Information Resource.

Introduction

A comprehensive understanding of the cell necessarily requires 'omics' analyses (e.g. genomics, transcriptomics, proteomics and metabolomics) due to the complicated nature of its molecular networks (Mochida and Shinozaki 2010). The genomes of various species have been sequenced to date, facilitating such analyses. The genomes of many plant species have been determined and have been annotated and organized in websites such as Phytozome (Goodstein et al. 2012), The Arabidopsis Information Resource (TAIR) (Lamesch et al. 2012) and GRAMENE (Youens-Clark et al. 2011). Moreover, genes involved in relevant biological processes and pathways are often cooperatively expressed to ensure proper functioning, and, thus, information on such co-expression is key to understanding biological systems at the molecular level (Eisen et al. 1998). Many transcriptome experiments in plants have revealed valuable information about gene expression related to various conditions, tissues and developmental stages, which enables gene co-expression analyses. Similarly, comparisons of the metabolite composition of plant tissues [e.g. in Arabidopsis

Plant Cell Physiol. 54(2): e5(1–8) (2013) doi:10.1093/pcp/pcs184, available online at www.pcp.oxfordjournals.org

(Brown et al. 2003) and *Lotus japonicus* (Desbrosses et al. 2005)] have revealed that metabolic systems have evolved in a tissue-dependent manner in order to produce a variety of metabolites, which ultimately increases plants' adaptive potential. Recent progress in phytochemical genomics with the model plant Arabidopsis has enriched the growing list of functionally identified genes (Hirai et al. 2007, Saito et al. 2008, Yonekura-Sakakibara et al. 2008) with the aid of transcriptome resources (Craigon et al. 2004, Schmid et al. 2005, Kilian et al. 2007, Goda et al. 2008, Obayashi et al. 2011). However, because the phytochemicals produced in Arabidopsis have not been fully identified and characterized, the majority of the functions of metabolic genes and metabolic systems in plants remain unknown.

The Platform for RIKEN Metabolomics (PRIMe) was developed to serve as an integrative analysis resource between Arabidopsis transcriptome and metabolome data (Akiyama et al. 2008). To facilitate the identification of phytochemicals, the MS2T library, AtMetExpress LC-MS and ReSpect databases were constructed and implemented on PRIMe as described below. Metabolome analyses using liquid chromatography–mass spectrometry (LC-MS) have revealed that Arabidopsis produces a huge number of unknown phytochemicals (Bottcher et al. 2008, Farag et al. 2008, Iijima et al. 2008, Matsuda et al. 2009, Sawada et al. 2009). Fragment pattern analysis of tandem mass spectrometry (MS/MS) is useful for the structural characterization of metabolites. Construction of a plant-specific MS/MS data resource and database would further allow for the identification of candidate structures from the catalogue of complex phytochemical structures. Therefore, we developed an MS/MS spectral tag (MS2T) library (Matsuda et al. 2009) and incorporated it in PRIMe. Subsequently, we constructed the AtMetExpress Development LC-MS database (Matsuda et al. 2010). This information platform facilitated insight into the mechanisms behind the variations in metabolic profiles among plant tissues by offering a method to investigate the relationship between gene expression and metabolite accumulation. Similarly, to investigate variations in the composition of secondary metabolites among Arabidopsis ecotypes (accessions), metabolic profile data were obtained from the rosette leaves of 20 accessions of Arabidopsis, which were distributed in the data set designated as AtMetExpress 20 ecotypes LC-MS (Matsuda et al. 2011). In addition, we have developed a web-based database for obtaining the MS/MS data of phytochemicals, termed ReSpect (RIKEN tandem mass spectral database) (Sawada et al. 2012). Seventy-six percent of the metabolites in ReSpect records were derived from data obtained from the literature, and other data were obtained from analyses of standard compounds.

As described above, these three databases offer powerful tools for developing a further understanding of the relationship between gene expression and metabolite accumulation. Here, we report the update of PRIMe implementing these three databases and an integration of the databases in PRIMeLink. Because of the current lack of metabolite annotation in the AtMetExpress Development LC-MS database, information provided by its generated metabolite spectra peaks can now be enhanced using the ReSpect annotation. Therefore, we have integrated and linked information regarding both genes and metabolites and have developed a new integrated database named PRIMeLink. This new database also provides a bi-directional searchable function from either the gene or the metabolite perspective. This integration permits users to browse through relationships between gene expression and metabolite accumulation in order to enhance the annotation of metabolite spectra peaks.

Database contents

PRIMe is a publicly accessible website that provides access to databases and analytical tools without the need for registration. Three new databases, the MS/MS spectral tag (MS2T) library (Matsuda et al. 2009), AtMetExpress LC-MS (Matsuda et al. 2010, Matsuda et al. 2011) and ReSpect (Sawada et al. 2012), were appended on PRIMe since our previous report (Akiyama et al. 2008). Moreover, we linked the three new contents, which contain gene expression and metabolite accumulation and annotations, and developed a new integrated database named PRIMeLink in the PRIMe website. PRIMeLink integrates the above three new databases and provides a bi-directional searchable function from the gene or metabolite perspective (**Fig. 1**). The three contents integrated into PRIMeLink are described below.

MS/MS spectral tag (MS2T) library

The MS2T library-based peak annotation procedure was developed for informative non-targeted metabolic profiling analysis using LC-MS (Matsuda et al. 2009). In order to create the MS2T libraries of plant metabolites, sample extracts derived from various tissues were analyzed using liquid chromatography coupled with electrospray ionization quadrupole time-of-flight MS/MS (LC-ESI-Q-TOF-MS/MS) by running the mass spectrometer in data-dependent acquisition mode (Ishihama 2005, Hernandez et al. 2006). With respect to metabolic profile data for Arabidopsis tissues containing $>1{,}000$ peaks, approximately 50% of the peaks were tagged by MS2Ts, and 90 peaks were identified or tentatively annotated with metabolite information by searching the metabolite databases and manually interpreting the MS2Ts. A comparison of metabolic profiles among Arabidopsis tissues revealed that many unknown metabolites accumulated in a tissue-specific manner, some of which were deduced to be unusual Arabidopsis metabolites on the basis of the MS2T data. Each MS2T ID was given in the format 'ATH06p00004', for example, denoting the fourth spectrum (00004) derived from the sixth library of *Arabidopsis thaliana* (ATH06) extracts obtained in the positive ion mode (p, positive). A web-based tool, named 'MS2T viewer', is provided on PRIMe in order to visualize the MS/MS spectral data of the MS2T accessions. A total of 875,000 MS2T records derived from Arabidopsis are contained in

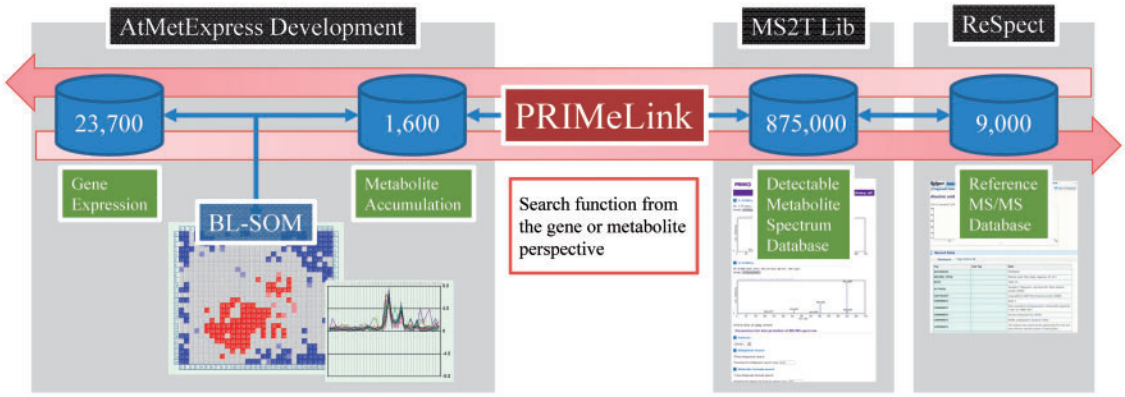

Fig. 1 Relationship diagram showing the integration of PRIMeLink with AtMetExpress Development LC-MS, MS2T library and ReSpect. PRIMeLink integrates the three new databases and provides a bi-directional searchable function from the gene or metabolite perspective. The numbers of each database represent the numbers of the records in each database.

PRIMeLink in order to link data regarding Arabidopsis genes and metabolites.

AtMetExpress LC-MS

In addition to the genome and transcriptome information resources, metabolome information would enrich our knowledge of secondary metabolism in plants. We therefore developed AtMetExpress LC-MS, which is a phytochemical atlas of Arabidopsis. At present, it consists of two data sets: AtMetExpress Development LC-MS and AtMetExpress 20 ecotypes LC-MS. Plants possess many metabolic genes for the production of a wide variety of phytochemicals in a tissue-specific manner. Therefore, we obtained quadruplicate metabolic profiles of 36 distinct Arabidopsis tissues by using LC-ESI-Q-TOF-MS/MS for determining the metabolite levels, in a manner similar to the MS2T library (Matsuda et al. 2009) mentioned above. Based on this data set, we detected 1,589 metabolite signals, from which the structures of 167 metabolites were elucidated. To clarify any possible similarities in the expression patterns of each gene with the accumulation patterns of their associated metabolites, we conducted a clustering analysis using the batch-learning self-organizing map (BL-SOM) method. BL-SOM is an improved and reproducible method of the original SOM (Kanaya et al. 2001) and thus leads to successful prediction of gene functions when applied to an integrated analysis of transcriptomes and metabolomes (Hirai et al. 2004, Hirai et al. 2005). All 1,589 metabolite signals, along with the 10,147 metabolism-related genes (probe sets of the GeneChip Arabidopsis ATH1 array) selected by Gene Ontology (GO) terms (Ashburner et al. 2000) and TAIR9 annotation, were classified into a 30×26 lattice, according to their relative expression levels across 36 tissues. These results suggested that the dynamics and diversity of plants' secondary metabolism depend on a simple mode of regulation from transcript to metabolite, which were organized into AtMetExpress Development LC-MS (Matsuda et al. 2010). To investigate variations in the composition of secondary metabolites among Arabidopsis ecotypes (accessions) for AtMetExpress 20 ecotypes LC-MS (Matsuda et al. 2011), metabolic profile data were obtained from the rosette leaves of 20 ecotypes of Arabidopsis by LC-ESI-Q-TOF-MS/MS analysis. The 20 diverse ecotypes were previously selected by Clark et al. (2007) to investigate genetic variation within Arabidopsis.

ReSpect for phytochemicals

To enhance the annotation rate of complex phytochemical structures, the ReSpect database has three major features that distinguish it from other MS/MS databases (Sawada et al. 2012). The first feature of the ReSpect database is a newly established fragment search system. ReSpect contains 3,341 records derived from data from 163 published studies. The literature offers a valuable resource with respect to MS/MS data of phytochemicals, but only half of them describe the mass to charge ratio (m/z) value of fragment ions. For such data, we have adjusted the relative intensities of all ions to 100% for visualization. Secondly, we enumerated 'MS/MS fragmentation association rules' among ReSpect records (Sawada et al. 2012). The association rule is an algorithm of frequent pattern mining for discovering interesting relationships between variables in the database. We applied this approach to estimate common MS/MS fragment patterns. The third feature of ReSpect is a plant-specific MS/MS data resource. Because thousands of phytochemicals have been reported in the literature, MS/MS data of phytochemicals can be obtained from a search in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) (Sayers et al. 2012) and Google Scholar (http://scholar.google.com) (Butler 2004). Thus, the MS/MS data were manually digitized, and all data obtained were included in the ReSpect database. The addition of this plant-specific data resource may allow the MS/MS data search in ReSpect to result in higher annotation success rates.

Database construction

PRIMe assembles databases and tools for plant transcriptomics and metabolomics. Since reports of its new contents (MS2T

library, AtMetExpress Development LC-MS and ReSpect) have already been published individually, in this section we focus mainly on details of the database construction of PRIMeLink.

Integration of gene expression and metabolite accumulation

AtMetExpress Development LC-MS is one of the most successful analytical platforms for integrating transcriptome and metabolome data while including the relevant MS2T information. Likewise, ReSpect, which houses highly annotated MS/MS data, is a useful tool. However, until now, these two tools were not functionally linked. Therefore, we upgraded the MS spectrum annotation by integrating the two databases and then developing the PRIMeLink database. Moreover, PRIMeLink enables users to search the genes related to metabolite accumulation with an optional metabolite name, as well as to find the metabolites accumulating in conjunction with gene expression patterns.

With respect to the clustering of metabolite signals and metabolism-related genes, we adopted the BL-SOM method, as in AtMetExpress Development LC-MS. We processed the data used in AtMetExpress Development LC-MS, and all 1589 metabolite signals with 10,643 metabolism-related genes selected by GO terms (Ashburner et al. 2000) were clustered into a 30×26 lattice (780 cells) according to their relative expression levels across 36 tissues. With respect to the confidence-level distribution (Sumner et al. 2007) of AtMetExpress Development LC-MS metabolite peaks, the numbers of identified, putatively annotated, putatively characterized and unknown level compounds were 50, 26, 114 and 1,399, respectively.

Assignment of MS/MS spectral tags to compound information

Because the AtMetExpress Development LC-MS data and the MS2T libraries were established using comparable analytical conditions, we obtained MS/MS spectral data of a metabolite signal in AtMetExpress Development LC-MS by identifying the MS2T data acquired from the same metabolite. This means that the metabolite signals in AtMetExpress Development LC-MS were tagged with the corresponding MS2Ts. In this study, we selected 36 Arabidopsis MS2T libraries, which corresponded to the AtMetExpress Development LC-MS samples from all 173 libraries (**Table 1**). As a result, 56,481 MS2Ts were assigned to the metabolite signals in AtMetExpress Development LC-MS, and these MS2Ts were then assigned to the confirmed spectrum data in the ReSpect database. Subsequently, the MS2T spectrum data were assigned to metabolite information by using the spectrum search function of ReSpect, because not all MS2Ts have been well annotated. As a result, 6,454 MS2Ts were assigned to the ReSpect records as metabolite annotations (**Supplementary Table S1**). Finally, we were able to obtain appropriate compound information for 637 of the 1,399 unknown metabolite signals in AtMetExpress Development LC-MS.

Instructions

PRIMeLink provides a bi-directional searchable function from the gene or metabolite perspective, which is equipped to search the database in three ways: BL-SOM result, keyword search and ID search. To switch from a gene to a metabolite search, move to each search page by clicking the appropriate button on the menu area (**Fig. 2**).

BL-SOM result

PRIMeLink houses the BL-SOM result of the gene expression and metabolite accumulation patterns. The BL-SOM result consists of 780 SOM cells, so that the distribution of the cells estimates the strength of the relationship between the metabolite accumulation and gene expression pattern. Users can browse the information of metabolites and genes in each cell by clicking on any cell (**Fig. 2A**).

Keyword search

To find target genes or metabolites, PRIMeLink's keyword search accepts a single word of two or more characters, or multiple words as a query, and searches among the functional descriptions of genes and metabolite names in PRIMeLink's own database (**Fig. 2B**). When SOM cells containing target genes or metabolites are found, the cells are highlighted in pink (**Fig. 2C**). Click a given highlighted cell to show its contents. Users can then browse the information regarding the metabolites and genes in each cell (**Fig. 2D, E**).

Metabolite fragment search

Users can use MS/MS data to query the metabolites by specifying the *m/z* value. The SOM cells containing target metabolites that have a metabolite fragment peak in the range of the user-queried *m/z* are highlighted in pink (**Fig. 2F**).

ID search

Apart from a keyword search, PRIMeLink's ID search accepts single or multiple IDs up to 1,000 as a query and searches Arabidopsis Genome Initiative (AGI) gene locus IDs, AtMetExpress peak IDs, MS2T IDs and ReSpect accessions (**Fig. 2G**). When SOM cells containing target genes or metabolites are found, the cells are highlighted in a manner similar to that in the keyword search (**Fig. 2C**). Click a given highlighted cell to show its contents. Users can then browse the information of the metabolites and genes in each cell (**Fig. 2D, E**).

Metabolite/gene detailed information

Detailed information pages of each SOM cell contain material regarding the metabolite and the gene. First, information of metabolites that were clustered into the same cell by BL-SOM consists of its AtMetExpress peak ID, MS2T ID, ReSpect accession and the hyperlinks to corresponding information on each external database (**Fig. 2D**). Similarly, the information on genes consists of its AGI gene locus ID, a short

Table 1 List of the MS2T libraries corresponding to the AtMetExpress Development LC-MS samples

Library ID	Precursor selection[a]	Tissue	Growth condition	No. of records
ATH06p	Narrow	Buds	4-week-old	10,477
ATH07p	Narrow	Cauline leaves	4-week-old	9,274
ATH08p	Narrow	Flowers	4-week-old	10,879
ATH09p	Narrow	Rosette leaves (first–fifth)	4-week-old	9,867
ATH10p	Narrow	Rosette leaves (fifth–10th)	3-week-old	7,506
ATH11p	Narrow	Silique	4-week-old	9,938
ATH12p	Narrow	Internodes (first and second)	4-week-old	9,305
ATH13p	Narrow	Seeds	4-week-old	12,395
ATH14p	Narrow	Roots	4-week-old	10,439
ATH56p	Wide	Buds	4-week-old	15,153
ATH57p	Wide	Cauline leaves	4-week-old	13,472
ATH58p	Wide	Flowers	4-week-old	15,350
ATH59p	Wide	Rosette leaves (first–fifth)	4-week-old	12,242
ATH60p	Wide	Rosette leaves (fifth–10th)	3-week-old	13,659
ATH61p	Wide	Silique	4-week-old	14,683
ATH62p	Wide	Internodes (first and second)	4-week-old	13,133
ATH63p	Wide	Seeds	4-week-old	17,986
ATH64p	Wide	Roots	4-week-old	14,513
ATH06n	Narrow	Buds	4-week-old	10,092
ATH07n	Narrow	Cauline leaves	4-week-old	11,870
ATH08n	Narrow	Flowers	4-week-old	12,875
ATH09n	Narrow	Rosette leaves (first–fifth)	4-week-old	13,038
ATH10n	Narrow	Rosette leaves (fifth–10th)	3-week-old	10,715
ATH11n	Narrow	Silique	4-week-old	7,022
ATH12n	Narrow	Internodes (first and second)	4-week-old	5,836
ATH13n	Narrow	Seeds	4-week-old	13,255
ATH14n	Narrow	Roots	4-week-old	11,247
ATH56n	Wide	Buds	4-week-old	18,775
ATH57n	Wide	Cauline leaves	4-week-old	19,239
ATH58n	Wide	Flowers	4-week-old	19,647
ATH59n	Wide	Rosette leaves (first–fifth)	4-week-old	17,702
ATH60n	Wide	Rosette leaves (fifth–10th)	3-week-old	17,567
ATH61n	Wide	Silique	4-week-old	16,412
ATH62n	Wide	Internodes (first and second)	4-week-old	15,736
ATH63n	Wide	Seeds	4-week-old	18,979
ATH64n	Wide	Roots	4-week-old	15,752

[a] The width of precursor ion selection was controlled by the LM parameter of Q-T of Premier (Waters). For 'narrow' 1.0-Da width and 'wide' 3.0-Da width modes, LM was set to 4.7 and 15 V, respectively.

MS2Ts (56,481) were assigned to the metabolite signals in AtMetExpress Development LC-MS and were assigned to the confirmed spectrum data in the ReSpect database. *Arabidopsis thaliana* (Col-0) was the plant species used for assignment.

description, a curator summary and a computational description from TAIR 10 (Lamesch et al. 2012). The hyperlinks to corresponding information on TAIR and the Arabidopsis eFP browser (Winter et al. 2007) are also implemented (**Fig. 2E**). To compare metabolite accumulation data by AtMetExpress with gene expression data by the Arabidopsis eFP browser, users can confirm the relationship of gene expression and metabolite accumulation.

Example of analytical procedures

Click a cell of the SOM map to browse genes and metabolites that are similarly expressed. Specific genes and metabolites can be searched by entering keywords. **Fig. 3A** and **B** shows the genes and metabolites obtained from a keyword search using 'spermidine'. We can see that closer cells at the bottom-right area are highlighted in the two maps. Details of the genes and metabolites assigned in the cells are displayed by clicking the cell.

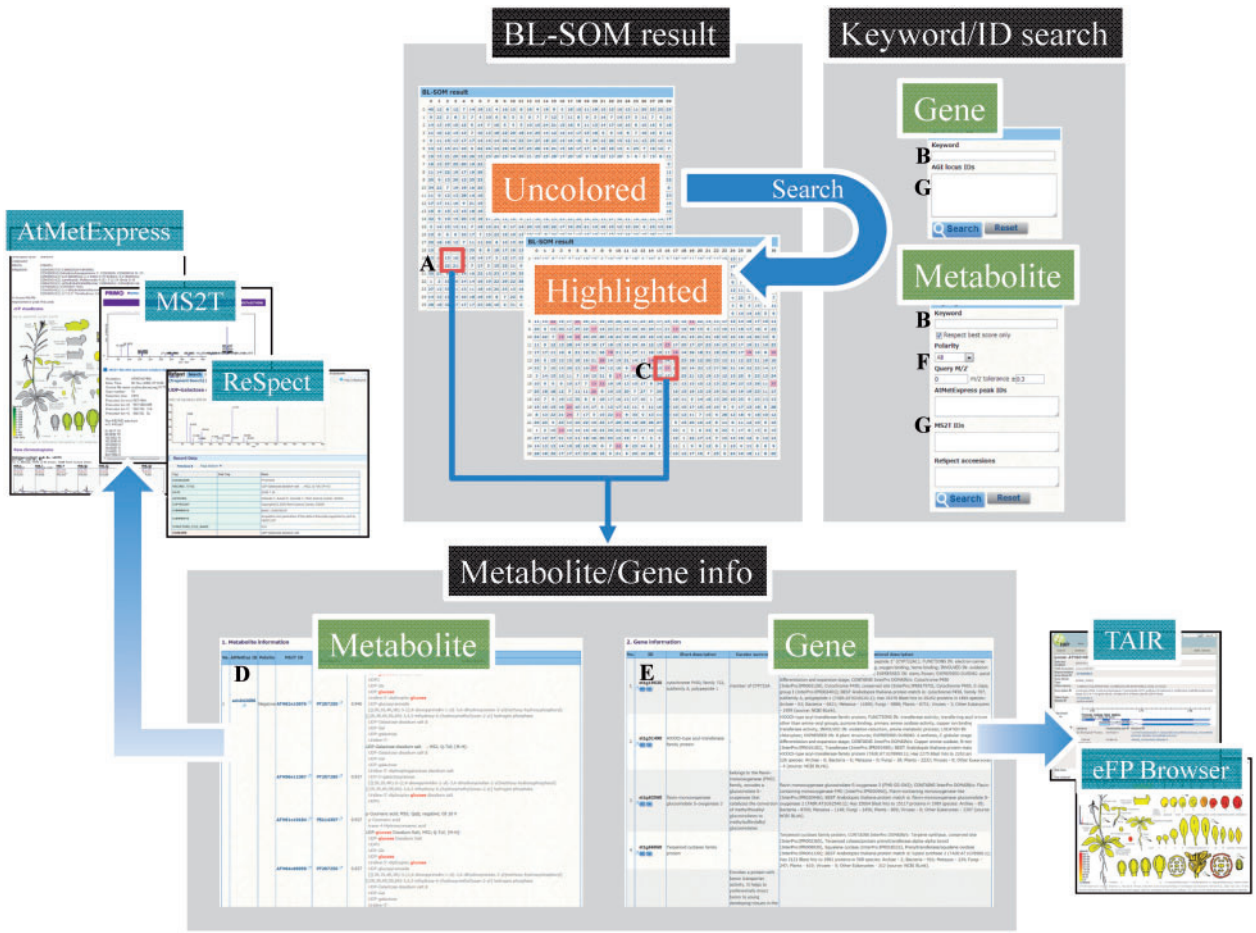

Fig. 2 PRIMeLink Screen transitions. (A) Link to detailed metabolite/gene information for a selected SOM cell. (B) Keyword search functions. (C) The SOM cell that matches with the search request is highlighted. (D) Detailed metabolite information. (E) Detailed gene information. (F) Metabolite fragment search function. (G) Various ID search functions.

The gene spermidine hydroxycinnamoyl transferase (*SHT*) is found in the cell at X:29, Y:25, and spermidine-related metabolites are found in closer cells. Sperimidine-*p*-coumaroyl-feruloyl is found at X:26, Y:25 as reported in Matsuda et al. (2010). Furthermore, spermidine-diferuloyl is found at X:26, Y:25 and X:25, Y:25, and spermidine-diferuloyl-hydroxyferuloyl, spermidine-trihydroxyferuloyl and spermidine-dihydroxyferuloyl-sinapyl at X:24, Y:25. These findings are brought by the enrichment of annotation with ReSpect. *SHT* has also been shown to function in the biosynthesis of spermidine-diferuloyl (Grienenberger et al. 2009). In addition, PRIMeLink also accepts the *m/z* of metabolite peaks as search terms for researchers who have obtained their own metabolome data. Furthermore, MS2T ID search is available if the users attach MS2T IDs of their own peaks in the user interface of MS2T Development. These search functions powerfully support the users to annotate peaks and to find candidates of related genes even when peaks are not well annotated. In addition, users can confirm whether the metabolite accumulation and gene expression patterns are similar by referring to detailed information that can be accessed via the hyperlinks for spermidine-diferuloyl (adp016958 in SOM cell X:26 Y:25) and *SHT* (At2g19070 in SOM cell X:29 Y:25) (**Fig. 3C, D**).

Conclusion

Because transcriptomics and metabolomics studies and their analytical platforms (e.g. higher density arrays, new generation sequencing and MS) will continue advancing and extending our knowledge, PRIMe is an ongoing project. In this study, we appended three new databases on PRIMe, and upgraded it as a more useful information platform. Furthermore, we integrated the three databases into PRIMeLink and established a walk-through link between transcriptome and metabolome information, as well as a well-annotated MS/MS data resource.

Currently, several information resources for MS/MS data are available, such as MassBank (Horai et al. 2010). However, the difficulty in developing novel methodologies remains even now. Our new databases offer powerful tools for the annotation of phytochemicals based on accessible MS/MS data resources and databases. In fact, we were able to obtain appropriate

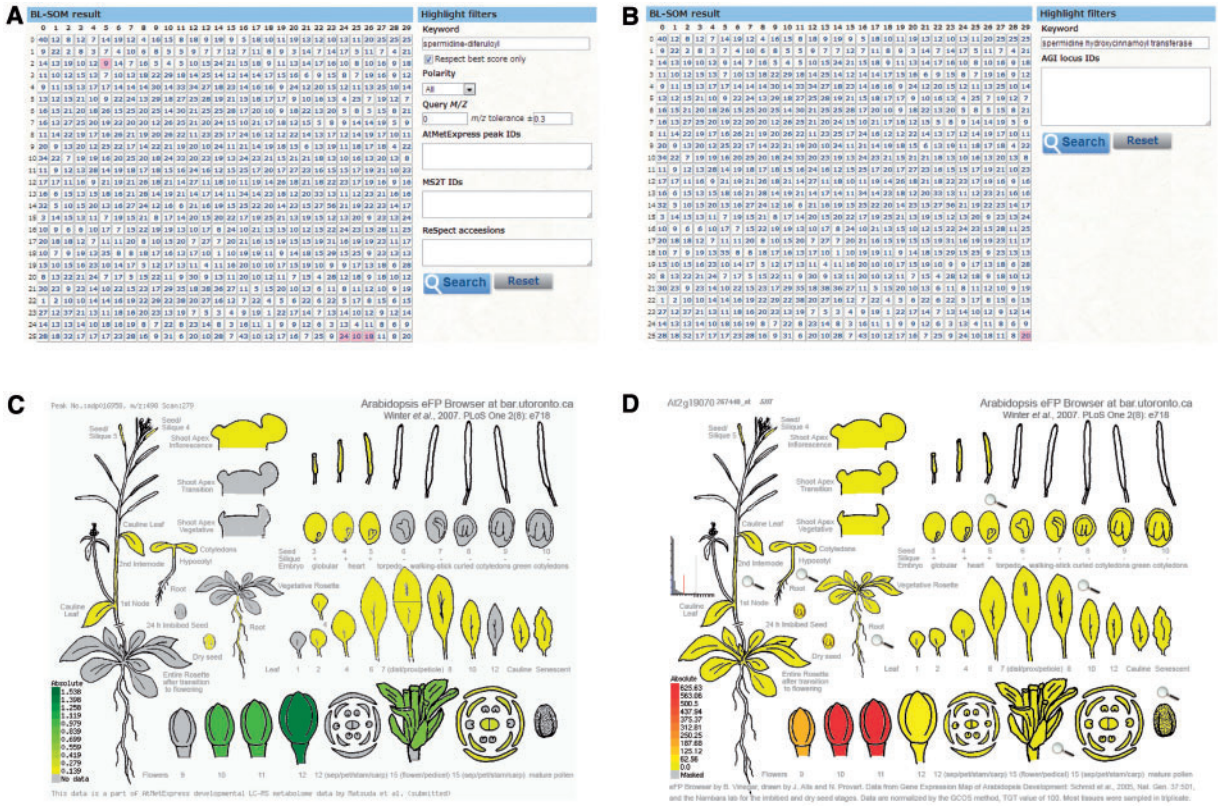

Fig. 3 Example of the relationship between metabolite accumulation and gene expression. (A) A highlighted SOM cell by matching with 'spermidine-diferuloyl' in Metabolite Search function. (B) A highlighted SOM cell by matching with 'spermidine hydroxycinnamoyl transferase' in Gene Search function. (C) The metabolite accumulation pattern of spermidine-diferuloyl (adp016958) on PRIMeLink. (D) The gene expression pattern of spermidine hydroxycinnamoyl transferase (*SHT*; At2g19070) on the Arabidopsis eFP Browser.

compound information from 637 of the 1,399 unknown metabolite signals in AtMetExpress Development LC-MS. This suggests that our updates to PRIMe are epochal, will provide a more useful data resource for studies at larger scales and offer further opportunities for progress in multiomics research.

We further expect that researchers in bioinformatics and metabolomics will be able to develop novel algorithms and analytical methods using our databases and data resources. As methods and research in this field develop, we will continue to construct databases and tools to meet current needs.

Funding

This study was supported by the Japan Society for the Promotion of Science [a Grant-in-Aid for Young Scientists (B) (grant No. 18700106 to T.S.)].

Acknowledgments

We thank Ryo Nakabayashi, Makoto Suzuki and Keiichi Mochida (RIKEN) for their helpful discussions on the database construction and website implementation.

References

Akiyama, K., Chikayama, E., Yuasa, H., Shimada, Y., Tohge, T., Shinozaki, K. et al. (2008) PRIMe: a Web site that assembles tools for metabolomics and transcriptomics. *In Silico Biol.* 8: 339–345.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25: 25–29.

Bottcher, C., von Roepenack-Lahaye, E., Schmidt, J., Schmotz, C., Neumann, S., Scheel, D. et al. (2008) Metabolome analysis of biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in Arabidopsis. *Plant Physiol.* 147: 2107–2120.

Brown, P.D., Tokuhisa, J.G., Reichelt, M. and Gershenzon, J. (2003) Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. *Phytochemistry* 62: 471–481.

Butler, D. (2004) Science searches shift up a gear as Google starts Scholar engine. *Nature* 432: 423.

Clark, R.M., Schweikert, G., Toomajian, C., Ossowski, S., Zeller, G., Shinn, P. et al. (2007) Common sequence polymorphisms shaping genetic diversity in Arabidopsis thaliana. *Science* 317: 338–342.

Craigon, D.J., James, N., Okyere, J., Higgins, J., Jotham, J. and May, S. (2004) NASCArrays: a repository for microarray data generated by NASC's transcriptomics service. *Nucleic Acids Res.* 32: D575–D577.

Desbrosses, G.G., Kopka, J. and Udvardi, M.K. (2005) Lotus japonicus metabolic profiling. Development of gas chromatography–mass spectrometry resources for the study of plant–microbe interactions. *Plant Physiol.* 137: 1302–1318.

Eisen, M.B., Spellman, P.T., Brown, P.O. and Botstein, D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* 95: 14863–14868.

Farag, M.A., Huhman, D.V., Dixon, R.A. and Sumner, L.W. (2008) Metabolomics reveals novel pathways and differential mechanistic and elicitor-specific responses in phenylpropanoid and isoflavonoid biosynthesis in Medicago truncatula cell cultures. *Plant Physiol.* 146: 387–402.

Goda, H., Sasaki, E., Akiyama, K., Maruyama-Nakashita, A., Nakabayashi, K., Li, W. et al. (2008) The AtGenExpress hormone and chemical treatment data set: experimental design, data evaluation, model data analysis and data access. *Plant J.* 55: 526–542.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40: D1178–1186.

Grienenberger, E., Besseau, S., Geoffroy, P., Debayle, D., Heintz, D., Lapierre, C. et al. (2009) A BAHD acyltransferase is expressed in the tapetum of Arabidopsis anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. *Plant J.* 58: 246–259.

Hernandez, P., Muller, M. and Appel, R.D. (2006) Automated protein identification by tandem mass spectrometry: issues and strategies. *Mass Spectrom. Rev.* 25: 235–254.

Hirai, M.Y., Klein, M., Fujikawa, Y., Yano, M., Goodenowe, D.B., Yamazaki, Y. et al. (2005) Elucidation of gene-to-gene and metabolite-to-gene networks in arabidopsis by integration of metabolomics and transcriptomics. *J. Biol. Chem.* 280: 25590–25595.

Hirai, M.Y., Sugiyama, K., Sawada, Y., Tohge, T., Obayashi, T., Suzuki, A. et al. (2007) Omics-based identification of Arabidopsis Myb transcription factors regulating aliphatic glucosinolate biosynthesis. *Proc. Natl Acad. Sci. USA* 104: 6478–6483.

Hirai, M.Y., Yano, M., Goodenowe, D.B., Kanaya, S., Kimura, T., Awazuhara, M. et al. (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.

Iijima, Y., Nakamura, Y., Ogata, Y., Tanaka, K., Sakurai, N., Suda, K. et al. (2008) Metabolite annotations based on the integration of mass spectral information. *Plant J.* 54: 949–962.

Ishihama, Y. (2005) Proteomic LC-MS systems using nanoscale liquid chromatography with tandem mass spectrometry. *J. Chromatogr. A* 1067: 73–83.

Kanaya, S., Kinouchi, M., Abe, T., Kudo, Y., Yamada, Y., Nishi, T. et al. (2001) Analysis of codon usage diversity of bacterial genes with a self-organizing map (SOM): characterization of horizontally transferred genes with emphasis on the *E. coli* O157 genome. *Gene* 276: 89–99.

Kilian, J., Whitehead, D., Horak, J., Wanke, D., Weinl, S., Batistic, O. et al. (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.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R. et al. (2012) The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* 40: D1202–1210.

Matsuda, F., Hirai, M.Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N.J. et al. (2010) AtMetExpress development: a phytochemical atlas of Arabidopsis development. *Plant Physiol.* 152: 566–578.

Matsuda, F., Nakabayashi, R., Sawada, Y., Suzuki, M., Hirai, M.Y., Kanaya, S. et al. (2011) Mass spectra-based framework for automated structural elucidation of metabolome data to explore phytochemical diversity. *Front. Plant Sci.* 2: 40.

Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kuromori, T., Shinozaki, K. and Saito, K. (2009) MS/MS spectral tag-based annotation of non-targeted profile of plant secondary metabolites. *Plant J.* 57: 555–577.

Mochida, K. and Shinozaki, K. (2010) Genomics and bioinformatics resources for crop improvement. *Plant Cell Physiol.* 51: 497–523.

Obayashi, T., Nishida, K., Kasahara, K. and Kinoshita, K. (2011) ATTED-II updates: condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant Cell Physiol.* 52: 213–219.

Saito, K., Hirai, M.Y. and 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. et al. (2009) 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., Nakabayashi, R., Yamada, Y., Suzuki, M., Sato, M. et al. (2012) RIKEN tandem mass spectral database (ReSpect) for phytochemicals: a plant-specific MS/MS-based data resource and database. *Phytochemistry* 82: 38–45.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K. et al. (2012) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 40: D13–D25.

Schmid, M., Davison, T.S., Henz, S.R., Pape, U.J., Demar, M., Vingron, M. et al. (2005) A gene expression map of Arabidopsis thaliana development. *Nat. Genet.* 37: 501–506.

Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A. et al. (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3: 211–221.

Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G.V. and Provart, N.J. (2007) An 'Electronic Fluorescent Pictograph' browser for exploring and analyzing large-scale biological data sets. *PloS One* 2: e718.

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

Youens-Clark, K., Buckler, E., Casstevens, T., Chen, C., Declerck, G., Derwent, P. et al. (2011) Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* 39: D1085–D1094.