Recent progress in the development of metabolome databases for plant systems biology

Atsushi Fukushima1 * and Miyako Kusano1,2
1 RIKEN Plant Science Center, Yokohama, Kanagawa, Japan
2 Department of Genome System Sciences, Graduate School of Nanobioscience, Kihara Institute for Biological Research, Yokohama, Kanagawa, Japan

INTRODUCTION

Metabolomics, i.e., the measurement of the full suite of metabolites in a living tissue, has expanded greatly over the last decade, especially in the context of biochemical phenotyping. Specifically, in plant science, metabolomic approaches are increasingly used for understanding regulatory networks involved in genotype comparison (Roessner et al., 2001; Weckwerth et al., 2004), measurement of diurnal/circadian rhythms (Urbanczyk-Wochniak et al., 2005; Gibon et al., 2006; Fukushima et al., 2009a; Espinoza et al., 2010), evaluation of genetically modified plants (Catchpole et al., 2005; Baker et al., 2006; Kusano et al., 2011a; Ricroch et al., 2011), uncovering relationships between metabolites associated with carbon and nitrogen metabolism (Stitt and Fernie, 2003; Sato et al., 2008; Kusano et al., 2011b), stress responses (Kaplan et al., 2004; Urano et al., 2009; Caldana et al., 2011; Kusano et al., 2011c; Obata and Fernie, 2012), characterization of many bioresources (Meyer et al., 2007; Rowe et al., 2008; Sulpice et al., 2009), and identifying metabolite quantitative trait loci (mQTLs) (Morreel et al., 2006; Schauer et al., 2006; Liseč et al., 2008; Carreno-Quintero et al., 2012; Matsuda et al., 2012).

It is estimated that approximately 200,000 metabolites are produced in the plant kingdom (Fiehn, 2002). There is no single technique suitable for measurement of all metabolites because of the chemical diversity of cellular metabolites and their broad dynamic range, particularly as this pertains to plants (Hall, 2006; Fukushima et al., 2009b; Saito and Matsuda, 2010; Lei et al., 2011; Weckwerth, 2011). Because of this, an array of analytical methods and extraction procedures has been developed for the detection of a broad spectrum of metabolites. Most procedures are based on either mass spectrometry (MS) or nuclear magnetic resonance (NMR). Metabolite data are typically generated through the following processes (Figure 1): (1) Sample preparation, (2) Data acquisition, and (3) Data pre-processing (the first half of the cycle in Figure 1). The resultant data are then subjected to multi-step downstream processes, including (4) Statistical data analysis approaches such as principal component analysis (PCA) and (5) Data interpretation using methods such as pathway analysis, which facilitate (6) The generation of testable hypotheses and the construction of models that best represent the biological phenomenon (the second half of the cycle in Figure 1). Experimental validation (7) of hypotheses and models [(7) in Figure 1] is necessary for closing the systems biology research cycle (Kitano, 2002; Fernie, 2012).

Traditionally, the elucidation of the structure of an unknown natural chemical compound has typically required the study of physicochemical properties, including the accurate mass and chemical shifts in 1H- and 13C-NMR spectra when the compound was first isolated. In contrast, metabolite identification in metabolomics using gas chromatography–mass spectrometry (GC-MS) and liquid chromatography–mass spectrometry (LC-MS) is not unambiguous. There are two venues to the identification of a chromatographic peak: (1) purification and NMR analysis or (2) interpretation of the spectra yielding a putative structure, followed by synthesizing or buying the compound and spiking. Detectable peaks using GC-MS and LC-MS are thought to be abundant and often authentic standards exist to identify them by spiking. Nonetheless, researchers commonly attempt to provisionally identify these peaks by comparing their mass spectra and/or the retention time or retention indices (RIs) with those present in a database build from the data of authentic standards. To validate the metabolite identification rigorously, the Metabolomics Standard Initiative (MSI)(Fiehn et al., 2007; Sansone et al., 2007) recommends different levels of identification (Sumner et al., 2007). Fernie et al. (2011) have also stated additional practical recommendations for reporting large-scale metabolite data.
It is quite evident that databases cataloging mass spectra and compounds give great support to metabolomic studies (for example, see Tohge and Fernie, 2009; Scalbert et al., 2011). Here, we highlight a wide range of metabolome databases, especially those that are widely used in MS-based metabolite profiling for rapid, but accurate, quantification, and identification of metabolites (Fiehn, 2002). We also discuss further steps to develop future databases facilitating metabolomic analyses and to improve bioinformatics tools in plant systems biology.

### Mass Spectrum-Oriented Information

Since non-targeted metabolite profiling using GC-MS for plant extracts was established in the early 2000s (Fiehn, 2001; Lisec et al., 2006), many software packages and databases for electron impact (EI) mass spectra and RIs of compounds analyzed by GC-MS have been created. The NIST/EPA/NIH mass spectral database represents the largest database commercially available for metabolite identification, containing mainly EI mass spectra RIs (Stein, 1999). Recently, the database also stores a set of MS/MS spectra of metabolites, drugs, peptides, and other compounds which are obtained by using ion trap-as well as tandem-MS instruments. The Golm Metabolome Database (GMD) provides GC-EI-MS mass spectral and RI (MSRI) libraries (Kopka et al., 2005; Schauer et al., 2005). It also contains mass spectral tags (MSTs) (Schauer et al., 2005), i.e., MS spectra of putative biological molecules which remain largely unidentified due to the lack of authentic standard compounds. GMD uses both alkanes and fatty acid methyl esters (FAMEs) for RI calculation whereas FiehnLib (Kind et al., 2009), a commercial MSRI library, uses FAMEs rather than alkanes. The Spectral Database for Organic Compounds (SDBS) includes a wide range of mass spectra for organic compounds, such as polysaccharides. MassBase is a mass spectral archive for LC-, GC-, and Capillary Electrophoresis–MS (CE-MS). SetupX and BinBase are a Laboratory Information Management System (LIMS)/database system for automated metabolite annotation and mass spectra, respectively. The Adams library (Adams, 2007), Terpenoids Library, and VocBinBase (Skogerson et al., 2011) are GC-specific MSRI libraries for volatile compounds. The former two are commercially available, while the VocBinBase database is freely available for their provisional identification (Skogerson et al., 2011). For MS data management and data sharing, MetabolomeExpress (Carroll et al., 2010) and MetaboLights (Haug et al., 2012; Steinbeck et al., 2012) were developed. The former is an ftp server that acts as a public data repository and web application for online data pre-processing and meta-analysis of publicly available metabolomic datasets analyzed by GC-MS. The latter is a general metabolomics repository; users can browse publicly available metabolomic datasets, search and see experimental meta-data, and re-use associated data files.

---

1 [http://riodb01.ibase.aist.go.jp/sdbs/](http://riodb01.ibase.aist.go.jp/sdbs/)
2 [http://webs2.kazusa.or.jp/massbase/](http://webs2.kazusa.or.jp/massbase/)
3 [http://code.google.com/p/setupx/](http://code.google.com/p/setupx/)
4 [http://sourceforge.net/projects/binbase/](http://sourceforge.net/projects/binbase/)
5 [http://massfinder.com/wiki/Terpenoids_Library](http://massfinder.com/wiki/Terpenoids_Library)
Table 1 | Metabolome databases involving mass spectra, compounds, metabolic pathways, metabolite profiles, and statistical/mathematical tools.

| Database                  | URL                                          | PMID         | Notes and contents                                                                 |
|---------------------------|----------------------------------------------|--------------|-------------------------------------------------------------------------------------|
| **MASS SPECTRUM-ORIENTED** |                                              |              |                                                                                     |
| AtMetExpress Development  | http://prime.psc.riken.jp/cms/AtMetExpress/   | 20023150     | A phytochemical atlas of *Arabidopsis thaliana* [LC-specific]                        |
| BinBase                   | http://eros.fiehnlab.ucdavis.edu:8080/binbase-compound/ | NA           | A database system for automated metabolite annotation                                |
| Bio-MassBank              | http://bio.massbank.jp/                      | NA           | Mass spectra from biological samples [currently, LC-specific]                        |
| FihenLib                  | http://fiehnlab.ucdavis.edu/projects/FiehnLib/index_html | 19928838     | Mass spectra and RI library based on GC-MS [GC-specific]                             |
| GMD@CSB.DB                | http://gmd.mpimp-golm.mpg.de/Default.aspx    | 15613389     | GC-MS, retention index, profiles [GC-specific]                                      |
| MaConDa                   | http://www.maconda.bham.ac.uk               | 22954629     | A database for mass spectrometry contaminants                                         |
| MassBank                  | http://www.massbank.jp/                     | 20623627     | Mass spectral database for LC-MS, GC-MS, CE-MS, MALDI-MS, MS2, etc                   |
| MassBase                  | http://webs2.kazusa.or.jp/massbase/          | NA           | A mass spectral archive for LC-MS, GC-MS, and CE-MS                                 |
| MetaboLights              | http://www.ebi.ac.uk/metabolights/           | 23060735     | A database for metabolomics data and meta-data.                                      |
| METLIN                    | http://metlin.scripps.edu                   | 16404815     | Mass spectral database for LC-MS, MS2                                               |
| MetaboSearch              | http://omics.georgetown.edu/MetaboSearch.html | 22768229     | A software tool for metabolite identification                                         |
| MS/MS spectral tag (MS2T) | http://prime.psc.riken.jp/cms/ms2tview/ms2tview.html | 18939963     | MS2 collections based on LC-MS [LC-specific]                                         |
| NIST                      | http://www.sisweb.com/software/ms/nist.htm   | NA           | Mass spectra based on GC-MS, LC-MS, MS2, etc                                        |
| ResPect                   | http://spectra.psc.riken.jp/                 | 22867903     | MSn spectrum collection of literature [LC-specific]                                  |
| SDBS                      | http://riodb01.ibase.aist.go.jp/sdbs/cgi-bin/cre_index.cgi?lang=en | NA           | Spectral database for organic compounds                                              |
| SetupX                    | http://fiehnlab.ucdavis.edu/projects/FiehnLib/8080/m1/ | NA           | A management system of mass spectrometric data                                       |
| The glycan mass spectral  | http://riodb.ibase.aist.go.jp/rcmg/glycodb/Ma_ResultSearch | 16053281     | Mass spectral database for glycan [MALDI-specific]                                   |
| database (GMDB)           |                                              |              |                                                                                     |
| The MetabolomeExpress     | https://www.metabolome-express.org/          | 20626915     | Online data repository for GC-MS-based metabolomics [GC-specific]                    |
| Adams library (4th Edition)| ISBN 978-1-932633-21-4                      |              | 2,205 MSRI library for plant essential oils [GC-specific]                            |
| Terpenoids and related    | http://massfinder.com/wiki/Terpenoids.Library | NA           | >2,400 MSRI library for industry and research in the areas of essential oils [GC-specific] |
| constituents of essential|                                              |              |                                                                                     |
| oils                      | http://vocbinbase.fiehnlab.ucdavis.edu/      | 21816034     | MSRI library for volatile compounds in GC-MS [GC-specific]                           |
| **COMPOUND-ORIENTED**     |                                              |              |                                                                                     |
| BKMrACT                   | http://www.cas.org/                          | 21824409     | A biochemical knowledge database                                                     |
| CAS                       | http://www.cas.org/                          | NA           | The world’s authority for chemical information similar to PubChem                     |
| ChEBI                     | http://www.ebi.ac.uk/chebi/                  | 17932057     | A database for chemical structure                                                    |
| Chemspider                | http://www.chemspider.com/                   | NA           | A database for small compounds                                                       |
| KEGG COMPOUND            | http://www.genome.jp/kegg/compound/          | 22080510     |                                                                                     |
| KNApSack                  | http://kanaya.naist.jp/KNApSack/             | 22123792     | A comprehensive species-metabolite relationship database                              |

(Continued)
## Table 1 | Continued

| Database | URL | PMID | Notes and contents |
|----------|-----|------|--------------------|
| LipidBank | http://lipidbank.jp/ | NA | A database for natural lipids |
| LipidMaps | http://www.lipidmaps.org/ | 17584797 | A database for lipid species |
| Metabolomics.jp | http://metabolomics.jp/ | 18822113 | A wiki database for metabolomics |
| MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases | http://metrxn.che.psu.edu/ | 2233419 | A web-based database for metabolites and reactions |
| Plant metabolome database (PMDB) | http://www.sastra.edu/scbt/pmdb/ | NA | Annotated database for metabolites in plants |
| PubChem | http://pubchem.ncbi.nlm.nih.gov | 19498078 | A database of small organic compounds |
| The MMD data | http://dbkgroup.org/MMD/ | 19562197 | A database for endogenous and exogenous metabolites |

### PATHWAY-ORIENTED

| Database | URL | PMID | Notes and contents |
|----------|-----|------|--------------------|
| AraCyc | http://arabidopsis.org/biocyc/ | 15888675 | Metabolic pathways for Arabidopsis |
| AraPath | http://bioinformatics.sdstate.edu/arapath/ | 22760305 | A knowledgebas for molecular pathways in Arabidopsis |
| BioCyc | http://biocyc.org/ | 22102576 | A collection of pathway/genome databases |
| iPath | http://pathways.embl.de | 2154551 | A pathway analysis database |
| KaPPA-View | http://kpv.kazusa.or.jp/en/ | 21097783 | A web-based database for analyzing omics data |
| MapMan | http://mapman.gabipd.org | 19389052 | A stand-alone tool for analyzing omics data |
| MetaCrop | http://metacrop.ipk-gatersleben.de | 22086948 | Metabolic pathways for crops |
| MetaCyc | http://metacyc.org/ | 22102576 | Metabolic pathways for multi-organisms |
| MetPA | http://metpa.metabolomics.ca | 20628077 | A web application for analyzing and visualizing metabolomic data |
| MetScape | http://metscape.ncibi.org/ | 22135418 | A Cytoscape plugin for visualizing metabolomic data |
| Paintomics | http://www.paintomics.org | 21098431 | A web application for the visualization of metabolomic data |
| Pathos | http://motif.gla.ac.uk/Pathos/ | 22002696 | A web-based database for the storage and analysis of metabolomic data |
| Pathvisio | http://www.pathvisio.org | 18817533 | A tool for visualizing biological pathways |
| PlantCyc | http://www.plantcyc.org/ | NA | Metabolic pathways for plants |
| ProMeta | http://www.cebitec.uni-bielefeld.de/groups/bio/software/prometa_info/ | 19698148 | A viewer for multiple omics data |
| SMPDB | http://www.smpdb.ca | 19948758 | A database for small molecule pathways |
| UniPathway | http://www.unipathway.org/ | 22102589 | A manually curated database for metabolic pathways |
| VANTED | http://vanted.ipk-gatersleben.de/ | 16519817 | A stand-alone tool for mapping omics data into metabolic pathways |

### METABOLITE-PROFILING-ORIENTED

| Database | URL | PMID | Notes and contents |
|----------|-----|------|--------------------|
| Plants | http://www.armec.org/ | 20003623 | A database for ESI-MS-based metabolomics including mainly potato |
| ARMeC | | | |
| Chloroplast 2010 Project | http://bioinfo.bch.msu.edu/2010_LIMS | 21224340 | Metabolite profiles of > 10000 SALK lines |
| GMD@CSB.D8: the Golm metabolome database | http://gmd.mpimp-golm.mpg.de/Default.aspx | 15613389 | A metabolome database for plants |
| KOMIC Market | http://webs2.kazusa.or.jp/komics/ | NA | A database for mass spectrometry-based metabolomics |
| McGill Metabolome Database | http://metabolomics.mcgill.ca/ | NA | A metabolome database for crops |

(Continued)
### Table 1 | Continued

| Database | URL | PMID | Notes and contents |
|----------|-----|------|-------------------|
| Medicinal plant metabolomics resource | http://metnetdb.org/mpmr_public/ | doi:10.3390/metabo2041031 | A metabolome database for medicinal plants |
| MeXO@PRIMe | http://prime.psc.riken.jp/meko/ | NA | A web-portal for visualizing metabolomic data of Arabidopsis |
| Moto DB (Metabolome tomato database) | http://appliedbioinformatics.wur.nl/moto/ | 16896233 | A metabolic database for tomato based on LC-MS |
| Plantmetabolomics.org | http://www.plantmetabolomics.org | 22080512 | A web-based database for analyzing and sharing metabolomic data of Arabidopsis |
| SoyMetDB: the soybean metabolome database | http://soymetdb.org/ | NA | A web-based database for soybean metabolomics |
| Animals | | | |
| HMDB | http://www.hmdb.ca/ | 18953024 | Human metabolome database |
| MMsDB | http://mmmsdb.iab.keio.ac.jp | 22139941 | Mouse multiple tissue metabolome database |
| SMDB | http://www.serummetabolome.ca/ | 21359215 | Serum Metabolome database |
| Bacteria | | | |
| ECMBD | http://www.ecmbd.ca/ | 23109553 | Ecoli metabolome database |
| YMDB | http://www.ymdb.ca/ | 22064856 | Yeast metabolome database |
| **TOOLS** | | | |
| Chemical translation service | http://cts.fiehnlab.ucdavis.edu | 20829444 | A web tool for translation of chemical information |
| IMPaLA | http://impaia.molgen.mpg.de/ | 21483477 | A web tool for over-representation and enrichment analysis |
| MBRole | http://csbg.cnb.csic.es/mbrole/ | 21208986 | A web application to perform various types of enrichment analyses |
| Metab2MeSH: annotating compounds with medical subject headings | http://metab2mesh.ncbi.nlm.nih.gov/ | 22492643 | A web application for annotating compounds with MeSH |
| MetaboAnalyst | http://www.metaboanalyst.ca/ | 22553367 | A web application to analyze metabolomic data [multiple functions] |
| MetaGeneAlyse | http://metageneanalyse.mpimp-golm.mpg.de/ | 14630670 | A web application for analyzing omics data [multiple functions] |
| MetaMapp | http://metamapp.fiehnlab.ucdavis.edu | 22591066 | A web application to generate network graph using metabolomics data |
| metaP-Server | http://metabolomics.helmholtz-muenchen.de/metap2/ | 20936179 | A web application to analyze metabolomic data |
| MetiTree | | | |
| MetMask | http://www.MetiTree.nl | 22851531 | A database for mass spectra of small molecules |
| MPEA | http://ekhida.biocenter.helsinki.fi/poxo/mpea/ | 20426876 | Integration tool for chemical identifiers |
| MSEA | http://www.msea.ca/ | 21551139 | Metabolite pathway enrichment analysis |
| MS-MS fragment viewer | http://webs2.kazusa.or.jp/msmsfragmentviewer/ | 20547745 | A web application to perform various types of enrichment analyses |
| SMPDB | http://www.smpdb.ca/ | 19948758 | A database for FT-MS-based metabolomics |
| MehIDB | http://meltdb.cebitec.uni-bielefeld.de | 18785459 | Small molecule pathway database |
| GENOME-SCALE METABOLIC MODELS | | | |
| AraGEM | http://web.aibn.uq.edu.au/cssb/resources/Genomes.html | 20044452 | A genome-scale metabolic reconstruction in Arabidopsis |
| C4GEM | http://web.aibn.uq.edu.au/cssb/resources/Genomes.html | 20974891 | A genome-scale metabolic reconstruction in C4 plants |
### Table 1 | Continued

| Database                      | URL                                    | PMID      | Notes and contents                                                                 |
|-------------------------------|----------------------------------------|-----------|------------------------------------------------------------------------------------|
| Poolman’s model               | http://www.plantphysiol.org/content/suppl/2009/10/08/pp.109.141267/DC1/141267Poolman_etal_Supl.zip | 19755544  | A genome-scale metabolic reconstruction in Arabidopsis                              |
| Radrich’s model               | http://www.biomedcentral.com/1752-0509/4/114/additional                          | 20712863  | A genome-scale metabolic reconstruction in Arabidopsis                              |
| Mintz-Oron model              | http://www.cs.technion.ac.il/~tomersh/methods.html                             | 22184215  | A tissue-specific genome-scale metabolic reconstruction in Arabidopsis              |
| BioModel database             | http://www.ebi.ac.uk/compneuro-srv/biomodels-main/                           | 20587024  | A database for mathematical models of biological pathways                            |
| BiGG                          | http://bigg.ucsd.edu/                                                      | 20426874  | A high-quality curated database for genome-scale metabolic reconstruction            |
| The model SEED                | http://www.theseed.org/models/                                                  | 20802497  | A high-throughput generation system for genome-scale metabolic model                |

It has been demonstrated that metabolite profiling using LC-MS has the potential to reveal secondary metabolites produced by plants, but most of the detected peaks in LC-MS profile data are largely unknown (Moco et al., 2006; De Vos et al., 2007; Iijima et al., 2008; Matsuda et al., 2010). Compared to EI, LC-MS ionization methods such as electrospray ionization (ESI) does hardly fragment the molecular ions. Even if authentic standards do not exist, putative metabolite identification can be done via MS/MS fragmentation and recording of accurate masses using ultra-high resolution MS such as Fourier-transform ion cyclotron resonance mass spectrometry (FT-ICR-MS) (Lenz et al., 2004; Nakabayashi et al., 2013). Different collision energy/analytical conditions cause different fragment patterns in mass spectra and should also be noted. Several databases have been developed for the sharing of ESI mass spectral information. METLIN stores high-resolution MS/MS spectra at four different collision energies (Tautenhahn et al., 2012). MassBank is a publicly available database of ESI-MS/MS spectra of authentic metabolites standards obtained under five collision energies as well as EI spectra (Horai et al., 2010). Bio-MassBank catalogs those obtained from biological samples6. In analogy with the MST spectra obtained from GC-MS data, the MS/MS spectral tag (MS2T) of detectable metabolites using LC-ESI-quadrupole-time-of-flight/MS is also available (Matsuda et al., 2009, 2010; Sakurai et al., 2013). This library contains MS2T obtained from species such as *Arabidopsis thaliana*, rice, soybean, and wheat. Compatible LC-MS settings with those used for the MS2T recording can be used for annotating detected peaks. Based on literature surveys, a plant-specific MS/MS spectra database was constructed by the same group (Sawada et al., 2012). MetaboSearch can be used to simultaneously retrieve mass-based metabolite data from multiple metabolite databases (Zhou et al., 2012). This contains tools to query and/or comprehensively analyze LC-MS-based metabolomic data. Together, these mass spectral databases make an important contribution to metabolite identification and also facilitate the development of bioinformatic tools (e.g., mining unknown metabolites) in metabolomics.

### COMPOUND-ORIENTED INFORMATION AND STRUCTURE CHARACTERIZATION

#### COMPOUND DATABASES

There are also compound databases (Table 1), such as Chemical Abstract Service (CAS)7. CAS is the oldest database of chemical information (e.g., journal abstracts); substances in the CAS registry database are each assigned a unique ID number. The PubChem database in NCBI (Wang et al., 2009), ChEBI (Deltcharenko et al., 2008), and ChemSpider (Pence and Williams, 2010) are freely available and can be used to retrieve chemical structures of small molecules. Well-curated chemical information including compounds and pathways are available in KEGG database (Kanehisa et al., 2012), KNAPSAcK database (Afendi et al., 2012), and Metabolomics.jp8 (Arita and Suwa, 2008). The Plant Metabolome Database (PMDB) is a freely available database of secondary metabolites in plants (Udayakumar et al., 2012). For bioactive lipids, LipidBank (Watanabe et al., 2000) and LipidMaps (Fahy et al., 2007) are available. The Manchester Metabolomics Database (MMD) has been developed to simultaneously utilize genome-scale data from the Human Metabolome Database (HMDB) (Wishart et al., 2009), KEGG, and LipidMaps. Other well-organized databases of biochemical knowledge are also available, such as BKM-react (Lang et al., 2011) and MetRxn (Kumar et al., 2012).

#### STRUCTURAL CHARACTERIZATION

One of the known bottlenecks in metabolomics is in the identification process of unknown metabolites, which can be classified as either “known unknowns” or “unknown unknowns” (Wishart, 2009). The former corresponds to a metabolite that has been previously detected but has not yet been identified, while the latter corresponds to a truly novel metabolite that has never been formally identified. Schymanski et al. (2012) have shown that consensus structure elucidation using a combination of GC-EI-MS, structure generation, and physicochemical properties calculated from

---

7http://www.cas.org/

8http://metabolomics.jp/
unknown compounds may be applicable to the characterization of unknown metabolites. Kumari et al. (2011) evaluated a novel de novo workflow for the annotation of unknown metabolites using accurate mass data, PubChem queries, RI matching, and structure constraints. To predict elemental compositions from accurate mass data collected from high-resolution mass spectrometers, "Seven Golden Rules" (Kind and Fiehn, 2007) and MFSearcher9 are available. Krumsiek et al. (2012) demonstrated that the integration of metabolite profiling with genome-wide association studies (GWAS) on metabolic quantitative traits is very useful for deriving biochemical pathways for unknown metabolites. In addition, several groups have attempted to classify unidentified MSTs using supervised machine learning approaches, including decision tree (Hummel et al., 2010) and soft independent modeling of class analogies (SIMCA)(Tsugawa et al., 2011). For structural characterization, there are recent powerful approaches by comparing mass spectral fragmentation trees (Rasche et al., 2011; Hufsky et al., 2012; Rojas-Cherto et al., 2012)(see also the review by Xiao et al., 2012). To evaluate whether detected peaks are biochemically produced by organisms, an in vivo 13C-labeling system has been used with 13CO2 in metabolite profiling using both GC-MS (Huege et al., 2007) and LC-MS (Giavalisco et al., 2011). The method allows for the rejection of non-biological peaks and improved annotation of elemental composition. Because artificial biological gradients developed by Redestig et al. (2011) can evaluate actual concentration differences of metabolite peaks detected in two different types of samples (e.g., leaves and fruits), this allows to filter out all unavoidable artifacts in MST/MS2T data. Such a method will make it possible to reject analytical artifacts and prioritize unknown candidate metabolites for further characterization.

**STATISTICAL MODELS, PATHWAY INFORMATION, AND DATA INTERPRETATION**

**UNIV- AND MULTIVARIATE ANALYSIS**

To perform extensive data analysis such as PCA, MetaGeneAlyse (Daub et al., 2003) and Metabo Analyst (Xia et al., 2012) are available. Conceptually, these are very similar web-based applications for the analysis of high-throughput omics data. MetaGeneAlyse implements standard normalization/clustering methods, e.g., k-means, and independent component analysis (ICA). Metabo Analyst provides many statistical methods, including t-tests, partial least square discriminant analysis (PLS-DA), pathway enrichment analysis, and additional machine learning methods. Please note that several tools and databases presented in this review have multiple functions. Furthermore, several web-based applications for metabolomic data are available (Table 1), such as MetaMapp (Barupal et al., 2012), metaP-Server (Kastenmuller et al., 2011), MeltDB (Neuweger et al., 2008), and MetiTree (Rojas-Cherto et al., 2012). They cover multiple steps from data pre-processing to biological interpretation.

**METABOLITE PATHWAY ANALYSIS**

Metabolite data, which contain information about metabolite name and changes of metabolite levels/relationships, can be described in pathways or networks. For example, the nodes represent metabolites and the edges represent biochemical reactions. Well-curated database for metabolic pathways in plants are available, such as KEGG (Kanehisa et al., 2012) and AraCyc (Zhang et al., 2005). MetaCrop stores well-curated information for 60 major metabolic pathways in eight crop plants, as well as Arabidopsis (Schreiber et al., 2012). UniPathway (Morgat et al., 2012) and SMPDB (Frolkis et al., 2010) also provide well-curated information about metabolic pathways. Tools involving pathway analysis and enrichment analysis are also available, such as AraPath (Lai et al., 2012), Kappa-view (Tokimatsu et al., 2005; Sakurai et al., 2011), and MapMan (Usadel et al., 2009) (Table 1). For detailed information about these tools, see the excellent review by Chagoyen and Pazos (2012).

**MATHEMATICAL MODEL INFORMATION AND OTHER TOOL GENOME-SCALE METABOLISM RECONSTRUCTION**

Over the past few decades, a significant number of metabolic reconstructions have been performed in many organisms, for example, SEED servers (Aziz et al., 2012). Currently, several genome-scale metabolic models in plants are available for evaluating metabolic behavior based on the alteration of metabolic pathways (Table 1) (Collakova et al., 2012; De Oliveira Dal’molin and Nielsen, 2012; Seaver et al., 2012). Poolman et al. (2009) constructed such a metabolism model in Arabidopsis to characterize possible flux behaviors using flux balance analysis (FBA) (Orth et al., 2010;Sweetlove and Ratcliffe, 2011). Instead of using metabolic flux analysis (MFA)(for example, see the reviews by Libourel and Shachar-Hill, 2008; Allen et al., 2009), this analysis can predict steady-state flux distribution by using a linear programing. AraGEM is also another metabolic reconstruction of Arabidopsis metabolism (De Oliveira Dal’molin et al., 2010). Radrich et al. (2010) semi-automatically integrated multiple databases involving metabolic pathways to reconstruct Arabidopsis metabolism. A compartmentalized, reconstructed metabolic model of Arabidopsis is also currently available (Mintz-Oron et al., 2012). Combinations of theoretical and experimental approaches will pave the way for robust interpretation of metabolomic data and practical metabolic engineering in plants.

**TOOLS FOR METABOLITE IDENTIFIERS**

Managing compound identifiers in metabolomic data analysis is important. MSI also proposed the use of database identifiers for peer-reviewed papers, for example, the most common compound identifiers, including CAS, KEGG COMPOUND, CHEBI, and HMDB. The Chemical Translation Service (CTS) (Wohlgenannt et al., 2010) and MetMask (Redestig et al., 2010) are a conversion tool for chemical identifiers (Table 1). The former is a web-based tool for performing batch conversions of compound identifiers, while the latter is a stand-alone command line program for integrating the most common compound identifiers. Metab2MeSH (Sartor et al., 2012) is a web application for annotating compounds with Medical Subject Headings (MeSH), which is a controlled vocabulary. Controlled vocabulary means well defined index term is used for indexing journal articles. Metab2MeSH links from metabolites to the biomedical research literature, PubChem,
and HMDB. These tools in this subsection are helpful for reporting metabolomic data.

**METABOLITE-PROFILING-ORIENTED INFORMATION**

In addition to mass spectrum and compound databases, several metabolite-profiling databases have also been developed in the past few years (Table 1). Among these, PlantMetabolomics.org (Bais et al., 2010, 2012; Quanbeck et al., 2012) and Medicinal Plant Metabolomics Resource (MPMR) (Wurtele et al., 2012) are one of the most important databases. These contain metabolomic information for >140 Arabidopsis mutants and 14 medicinal plants based on MS data from multiple laboratories (Bais et al., 2012). Their profiling broadly covers a wide range of metabolites relating to amino and fatty acids, organic acids, phytosterols, isoprenoids, lipids, and secondary metabolites. PlantMetabolomics.org and MPMR also provide multiple data analysis tools including data normalization and visualization. Using these tools investigators can generate testable hypotheses with respect to gene functions in Arabidopsis (Quanbeck et al., 2012). Another example is Chloroplast 2010, which contains data related to large-scale phenotypic screening of Arabidopsis chloroplast mutants (Lu et al., 2011), based on assays of amino acids and fatty acids in leaves and seeds using GC-MS and LC-MS (Gu et al., 2007; Bell et al., 2012). Recently, we constructed the MeKO database10 (Fukushima et al., submitted), which is similar in concept to PlantMetabolomics.org. MeKO contains metabolomic information on 50 Arabidopsis mutants, including plants with uncharacterized gene functions. The website also provides MSI-compliant data, experimental meta-data, and the results of statistical data analyses such as differential accumulation compared with wild-type plants (Columbia ecotype). These databases are very useful for functional genomics and make it possible to develop additional bioinformatic tools for pre-processing of metabolomics raw data, extraction of biologically meaningful mass spectra, and reduction/correction of unwanted variation in large-scale metabolomic data.

**CONCLUSION**

In this review, we have highlighted an extensive list of databases that incorporate both MS-based metabolomics, as well as data analysis tools. Clearly, a small, but significant, number of integrated databases, including the full annotation of metabolites, metabolite profiling, and data analysis tools are emerging, such as PlantMetabolomics.org (Quanbeck et al., 2012). In addition to those for plants, metabolome databases for bacteria and animals also exist (see Table 1). Increases in metabolomic data sharing and the improvement of technological capabilities, such as database integration, are likely to play important roles in the future development of plant metabolomics, and facilitate advances plant systems biology.

**ACKNOWLEDGMENTS**

Research activity of Atsushi Fukushima is partly supported by a Grant-in-Aid for Young Scientists (B; grant no. 23700355 to Atsushi Fukushima) from the Ministry of Education, Culture, Sports, Science, and Technology, Japan.

**REFERENCES**

Adams, R. P. (2007). Identification of Essential Oil Components by Gas Chromatography-Quadrupole Mass Spectroscopy, 4th Edn. Carol Stream, IL: Allured Pub Corp.

Afendi, F. M., Okada, T., Yamazaki, M., Hiroi-Morita, A., Nakamura, Y., Nakamura, K., et al. (2012). KNApSack family databases: integrated metabolite-plant species databases for multifaceted plant research. Plant Cell Physiol. 53, el.

Allen, D. K., Libourel, I. G., and Shachar-Hill, Y. (2009). Metabolic flux analysis in plants: coping with complexity. Plant Cell Environ. 32, 1241–1257.

Arita, M., and Sowa, K. (2008). Search extension transforms wki into a relational system: a case for flavonoid database. BioData Min. 1, 7.

Azi, R. K., Devoid, S., Diz, T., Edwards, R. A., Henry, C. S., Olsen, G. I., et al. (2012). SEED serves: high-performance access to the seed genomes, annotations, and metabolic models. PLoS ONE 7:e48053. doi:10.1371/journal.pone.0048053

Bais, P., Moon, S. M., He, K., Leitao, R., Dreher, K., Walk, T., et al. (2010). PlantMetabolomics.org: a web portal for plant metabolomics experiments. Plant Physiol. 152, 1807–1816.

Bais, P., Moon-Quanbeck, S. M., Nikolau, B. J., and Dickerson, J. A. (2011). Plantmetabolomics.org: mass spectrometry-based Arabidopsis metabolomics – database and tools update. Nucleic Acids Res. 40, D1216–D1220.

Baker, J. M., Hawkins, N. D., Ward, J. L., Lovegrove, A., Napier, J. A., Shewry, P. R., et al. (2006). A metabolomic study of substantial equivalence of field-grown genetically modified wheat. Plant Biotechnol. J. 4, 381–392.

Barupal, D. K., Haldiya, P. K., Wohlgemuth, G., Kind, T., Kothari, S. L., Pinkerton, K. E., et al. (2012). MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. BMC Bioinformatics 13:99. doi:10.1186/1471-2105-13-99

Bell, S. M., Burgoon, L. D., and Last, R. L. (2012). MIPHENO: data normalization for high throughput metabolite analysis. BMC Bioinformatics 13:10. doi:10.1186/1471-2105-13-10

Caldana, C., Degenkolbe, T., Cuadros-Inostroza, A., Klie, S., Sulpice, R., Leisse, A., et al. (2011). High-density kinetic analysis of the metabolic and transcriptomic response of Arabidopsis to eight environmental conditions. Plant J. 67, 869–884.

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

Carroll, A. J., Badger, M. R., and Harvey Millar, A. (2010). The Metabolome Express project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics data. BMC Bioinformatics 11:376. doi:10.1186/1471-2105-11-376

Catchpole, G. S., Beckmann, M., Enot, D. P., Mondhe, M., Zwyklicki, B., Taylor, J., et al. (2005). Hierarchical metabolomics demonstrates substantial compositional similarity between genetically modified and conventional potato crops. Proc. Natl. Acad. Sci. U.S.A. 102, 14458–14462.

Chapoyen, M., and Patoz, F. (2012). Tools for the functional interpretation of metabolomic experiments. Brief. Bioinformatics (in press).

Collakova, E., Yen, J. Y., and Senger, R. S. (2012). Are we ready for genome-scale modeling in plants? Plant Sci. 19, 53–70.

Dauh, C. O., Kloska, S., and Selbig, J. (2003). MetagenEyle: analysis of integrated transcriptional and metabolite data. Bioinformatics 19, 2332–2333.

De Oliveira Dal’molin, C. G., and Nielsen, L. K. (2012). Plant genome-scale metabolic reconstruction and modelling. Curr. Opin. Biotechnol. 24, 271–277.

De Oliveira Dal’molin, C. G., Quark, L. E., Palfreyman, R. W., Bramley, S. M., and Nielsen, L. K. (2010). AraGEM, a genome-scale reconstruction of the primary metabolic network in Arabidopsis. Plant Physiol. 152, 579–589.

De Vos, R. C., Moom, S., Lommen, A., Keurentjes, J. J., Bino, R. J., and Hall, R. D. (2007). Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat. Protoc. 2, 778–791.
Degtyarenko, K., De Matos, P., Ennis, M., Hastings, J., Zhbinder, M., McNaught, A., et al. (2008). ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. 36, D344–D350.

Espinoza, C., Degkenolke, T., Caldana, C., Zuther, E., Leisse, A., Willmitzer, L., et al. (2010). Interaction with diurnal and circadian regulation results in dynamic metabolic and transcriptional changes during cold acclimation in Arabidopsis. PLoS ONE 5, e11401. doi:10.1371/journal.pone.0011410

Fabry, E., Sud, M., Cotter, D., and Subramaniyan, S. (2007). LIPID MAPS online tools for lipid research. Nucleic Acids Res. 35, W606–W612.

Fernie, A. R. (2012). Grand challenges in metabolomics research. Plant Physiol. 159, 859–860.

Fernie, A. R., et al. (2010). Interactomics: an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res. 41, D781–D786.

Hail, R. D. (2006). Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol. 169, 453–468.

Hall, K., Sulek, R. M., Conesa, P., Hastings, J., de Matos, P., Rijnbeek, M., et al. (2012). Metabolights – an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res. 41, D781–D786.

Horie, H., Arita, M., and Saito, K. (2009b). Expression atlas of Arabidopsis: interactive and integrated view of gene expression and transcriptional regulation. Nucleic Acids Res. 37, 2529–2533.

Huege, J., Sulpice, R., Gibson, Y., Lisec, J., Koehl, K., and Kopka, J. (2007). GC-TOF-MS analysis of in vivo carbon-partitioning into soluble metabolite pools of higher plants by monitoring isotope dilution after $^{13}$C labelling. Phytochemistry 68, 2258–2272.

Hubsky, F., Rempt, M., Rasche, F., Buchert, G., and Becker, S. (2012). Decision tree supported substructure prediction of metabolites from GC-MS profiles. Metabolomics 6, 322–333.

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.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, D119–D124.

Kasahara, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, D119–D124.

Kaplan, R., Kopka, J., Haskell, D. W., and Schilperoort, R. A., et al. (2004). Exploring the temperature-stress metabolome of Arabidopsis. Plant Biologist 136, 4159–4168.

Kastenmüller, G., Remisch-Margl, W., Wagele, H., Blaimair, E., and Sührer, K. (2011). metaP-server: a web-based metabolomics data analysis tool. J. Biomed. Biotechnol. 2011:839862. doi:10.1155/2011/839862

Kind, T., and Fiehn, O. (2007). Seven golden rules for heuristics filtering of molecular formulae obtained by accurate mass spectrometry. BMC Bioinformatics 8:105. doi:10.1186/1471-2105-8-105

Kobayashi, M., Hayashi, N., Arita, M., and Saito, K. (2009). MS/MS/MS assay for protein amino acids and metabolically related compounds for large-scale screening of metabolic phenotypes. Anal. Chem. 81, 8067–8075.

Kitano, H. (2002). Systems biology: a brief overview. Science 295, 1662–1666.

Kopka, J., Schauer, N., Krueger, S., Birke-meyer, C., Usadel, B., Bergmuller, E., et al. (2005). GMD/CSDB: the Colm metabolome database. Bioinformat 21, 1635–1638.

Kremsiek, J., Sührer, K., Evans, A. M., Mitchell, M. W., Mohney, R. P., Milburn, M. V., et al. (2012). Miming the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genet. 8:e1003005. doi:10.1371/journal.pgen.1003005

Kumar, A., Suthers, P. F., and Maranas, C. D. (2012). MetRx: a knowledge-base of metabolites and reactions spanning metabolic models and databases. BMC Bioinformatics 13:6. doi:10.1186/1471-2105-13-6

Kusano, M., Redestig, H., Hirai, T., Okawa, A., Matsuda, F., Fukushimas, A., et al. (2011a). Covering chemical diversity of genetically-modified tomatoes using metabolomics for objective substantial equivalence assessment. PLoS ONE 6:e16989. doi:10.1371/journal.pone.0016989

Kusano, M., Tabuchi, M., Fukushimas, A., Funayama, K., Diaz, C., Kobayashi, M., et al. (2011b). Metabolomics data reveal a crucial role of cytosolic glutamine synthetase 1;1 in coordinating metabolic balance in rice. Plant J. 66, 456–466.

Kusano, M., Tohge, T., Fukushimas, A., Kobayashi, M., Hayashi, N., Otsuki, H., et al. (2011c). Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light. Plant J. 67, 354–369.

Lai, I., Liberezon, A., Hennessey, L., Jiang, G., Qi, J., Mesirov, J. P., et al. (2012). ArabPath: a knowledgebase for pathway analysis in Arabidopsis. Bioinformat 28, 2291–2292.

Lang, M., Stelzer, M., and Schomburg, D. (2011). BKM-react, an integrated biochemical reaction database. BMC Biochem. 12:42. doi:10.1186/1471-2091-12-42

Lei, Z., Huhman, D. V., and Sumner, L. W. (2011). Mass spectrometry strategies in metabolomics. J. Biol. Chem. 286, 25435–25442.

Lenz, E. M., Bright, J., Knight, R., Wil-son, I., and Major, H. (2004). Cyclespinor A-induced changes in endogenous metabolites in rat urine: a metabolomic investigation using high field 1H NMR spectroscopy. J. Pharm. Biomed. Anal. 35, 599–608.

Libourel, I. G., and Shachar-Hill, Y. (2008). Metabolic flux analysis in plants: from intelligent design to rational engineering. Annu. Rev. Plant Biol. 59, 625–650.

Lisse, J., Meyer, R. C., Steinfath, M., Redestig, H., Becher, M., Wittruch, W., et al. (2008). Identification of metabolic and biomass QTL in Arabidopsis thaliana in a parallel analysis of RIL and F2 populations. Plant J. 53, 960–972.

Lisse, J., Schauer, N., Kopka, J., Willmitzer, L., and Fernie, A. R. (2006). Gas chromatography mass spectrometry-based metabolite profiling in plants. Nat. Protoc. 1, 387–396.

Lu, Y., Savage, L. I., Larson, M. D., Wilkerson, C. G., and Last, R. L. (2011). Chloroplast 2010: a database for large-scale phenotypic screening of Arabidopsis mutants. Plant Physiol. 155, 1589–1600.

Matsuda, F., Hirai, M. Y., Sasaki, E., Akiyama, K., Yonekura-Sakakibara, K., Provart, N. J., et al. (2010). AtMetExpress development: a phylogenetic resource of Arabidopsis metabolic and physiological traits. Nat. Protoc. 5, 407–416.

Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, I., et al. (2012). Dissection of genotype-phenotype associations in rice using mapping metabolome quantitative trait loci database. Plant J. 70, 624–636.

Matsuda, F., Yonekura-Sakakibara, K., Niida, R., Kurimoto, T., Shinozaki, K., and Saito, K. (2009). MS/MS spectral tag-based annotation of metabolites of Arabidopsis thaliana. Plant Physiol. 151, 86–98.
non-targeted profile of plant sec-
dondary metabolites. Plant J. 57, 555–577.
Meyer, R. C., Steinfath, M., Lisej, I., Becher, M., Witucka-Wall, H., Tor-
jek, Ö., et al. (2007). The meta-
bolic signature related to high plant
growth rate in Arabidopsis thaliana.
Proc. Natl. Acad. Sci. U.S.A. 104, 4759–4764.
Mintz-Oron, S., Meir, S., Malitsky, S., Ruppin, E., Aharoni, A., and Shlomi,
T. (2012). Reconstruction of Ara-
bidopsis metabolic network models
accounting for subcellular compart-
mentalization and tissue-specificity.
Proc. Natl. Acad. Sci. U.S.A. 109, 339–344.
Moco, S., Bino, R. J., Vorst, Ö., Verhoeven, H. A., De Groot, J., Van Beek, T. A., et al. (2006). A liquid chromatography-mass spectrometry-based metabolome database for tomato. Plant Physiol. 141, 1205–1218.
Morgat, A., Coissac, E., Coudert, O., Axelsen, K. B., Keller, G., Bairoch, A., et al. (2011). UniPathway: a resource for the exploration and annotation of metabolic pathways. Nucleic Acids Res. 40, D761–D769.
Morreel, K., Goeminne, G., Storme, V., Fukushima and Kusano Plant metabolome databases
Obata, T., and Fernie, A. R. (2012). The van Beek, T. A., et al. (2006).
Verhoeven, H. A., De Groot, J., Ruppin, E., Aharoni, A., and Shlomi, T., et al. (2012). Metabolite identification using automated compari-
nation of high-resolution multistage mass spectral data. Plant Physiol. 155, 1752–1761.
Roessner, U., Luemdemann, A., Brust, D., Fiehn, O., Linke, T., Willmitzer, L., et al. (2001). Metabolic profiling allows comprehensive pheno-
typing of genetically and environmen-
tally modified plant systems. Plant Cell 13, 11–29.
Rojas-Cherto, M., Peironcely, J. F., Kasper, P. T., Van Der Hoof, J. J., De Vos, R. C., Vreeken, R., et al. (2012). Metabolite iden-
fication using automated compara-
nion of high-resolution multistage
mass spectral trees. Anal. Chem. 84, 5524–5534.
Rowe, H. C., Hansen, B. G., Halkier, B. A., and Kliebenstein, D. J. (2008). Biochemical networks and epistasis shape the Arabidopsis thaliana metabolome. Plant Cell 20, 1199–1216.
Saito, K., and Matsuda, F. (2010). Metabolomics for functional genomics, systems biology, and biotechnology. Annu. Rev. Plant Biol. 61, 463–489.
Sakurai, N., Ara, T., Ogata, Y., Sano, R., Ohno, T., Sugiyama, K., et al. (2011). KappaView: a metabolic path-
way database for representation and analysis of correlation networks of gene co-expression and metabolite co-accumulation and omics data. Nucleic Acids Res. 39, D677–D684.
Sakurai, T., Yamada, Y., Sawada, Y., Mat-
suda, F., Akiyama, K., Shionozaki, K., et al. (2013). PRIME update: innova-
tive content for plant metabolomics and integration of gene expression and metabolite accumulation. Plant Cell Physiol. 54, 49.
Sassone, S. A., Fan, T., Goodacre, R., Griffin, J. L., Hardy, N. W., Kaddurah-Daouk, R., et al. (2007). The metabolomics standards initia-
tive. Nat. Biotechnol. 25, 846–848.
Sartor, M. A., Ade, A., Wright, Z., Stakes, D., Ommen, G. S., Athey, B., et al. (2012). Metab2MeSH: annotat-
ing compounds with medical sub-
ject headings. Bioinformatics 28, 1408–1410.
Sato, S., Arita, M., Soga, T., Nishi-
toka, T., and Tomita, M. (2008). Time-resolved metabolomics reveal metabolic modulation in rice folage. BMC Syst. Biol. 2:51.
doi:10.1186/1752-0509-2-51
Sawada, Y., Nakabayashi, R., Yamada, Y., Suzuki, M., Saito, M., Sakata, A., 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.
Scalbert, A., Andres-Lacueva, C., Arita, M., Kroon, P. M., Manach, C., Urpi-
sarda, M., et al. (2011). Databases on food phytochemicals and their uses.
Phytochemistry 72, 4313–4348.
Schauer, N., Semel, Y., Roessner, U., Gur, A., Balbo, I., Carrari, F., et al. (2006). Comprehensive meta-
bolomic profiling and phenotyping of interspecific introgression lines for tomato improvement. Nat. Biotechnol. 24, 447–454.
Schauer, N., Steinhauser, D., Strelkov, S., Schomberg, D., Allison, G., Moritz, T., et al. (2005). GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBs Lett. 579, 1332–1337.
Seifer, B., Colsme, C., Cauzad-
terna, T., Grafahrend-Belau, E., Hansen, A., Junker, A., et al. (2012). MetaCrop 2.0: managing and exploring information about crop plant metabolism. Nucleic Acids Res. 40, D1173–D1177.
Schemansky, E. L., Gallampoos, C. M., Krauss, M., Meringer, M., Neumann, S., Schulze, T., et al. (2012). Con-
sensus structure elucidation com-
ing GC/MS-MS, structure genera-
tion, and calculated properties. Anal. Chem. 84, 3287–3295.
Seaver, S. M., Henry, C. S., and Han-
son, A. D. (2012). Frontiers in meta-
bolite reconstruction and modeling of plant genomes. J. Exp. Bot. 63, 2247–2258.
Skogerson, K., Wohlgemuth, G., Baru-
pal, D. K., and Fiehn, Ö. (2011). The volatile compound BinBase mass spectral database. BMC Bioinformat-
is 12:212. doi:10.1186/1471-2105-
12-321
Stein, S. E. (1999). An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J. Am. Soc. Mass Spectrom. 10, 770–781.
Steinbeck, C., Conesa, P., Haug, K., Mahendraker, T., Williams, M., Maguire, E., et al. (2012). Meta-
BioLigths: towards a new COSMOS of metabolomics data management. Metabolomics 8, 757–760.
Stitt, M., and Fernie, A. R. (2003). From measurement of metabo-
lites to metabolomics: an ‘on the fly’ perspective illustrated by recent studies of carbon-nitrogen interac-
tions. Curr. Opin. Biotechnol. 14, 136–144.
Sulpice, R., Pyl, E. T., Ishihara, H., Trenkamp, S., Steinfath, M., Witucka-Wall, H., et al. (2009). Starch as a major integrator in the regulation of plant growth. Proc. Natl. Acad. Sci. U.S.A. 106, 10348–10353.
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.
Sweetlove, L. J., and Ratcliffe, R. G. (2011). Flux-balance modeling of plant metabolism. Front. Plant Sci. 2:338. doi:10.3389/fpls.2011.00038
Tautenhahn, R., Cho, K., Uribein-
thai, W., Zhu, Z., Patti, G. J., and Suizdak, G. (2012). An accel-
 erated workflow for untargeted metabolomics using the METLIN database. Nat. Biotechnol. 30, 826–828.
Tohge, T., and Fernie, A. R. (2009). Web-based resources for mass-
spectrometry-based metabolomics: a user’s guide. Phytochemistry 70, 450–456.
Tokimatsu, T., Sakurai, N., Suzuki, H., Ohta, H., Nishitani, K., Koyama, T., et al. (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.

Tsuchiya, H., Tsujimoto, Y., Arita, M., Bamba, T., and Fukusaki, E. (2011). GC/MS based metabolomics: development of a data mining system for metabolite identification by using soft independent modeling of class analogy (SIMCA). BMC Bioinformatics 12:131. doi:10.1186/1471-2105-12-131

Udagakumar, M., Chandar, D. P., Arun, N., Mathangi, J., Hema-vathi, K., and Seenivasagam, R. (2012). PMDR: plant f database – a metabolomic approach. Med. Chem. Res. 21, 47–52.

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

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

Usadel, B., Pozez, E., Nagel, A., Lohse, M., Czedik-Eysenberg, A., and Stitt, M. (2009). A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species. Maize. Plant Cell Environ. 32, 1211–1229.

Wang, Y., Xiao, J., Szeck, T. O., Zhang, J., Wang, L., and Bryant, S. H. (2009). PubChem: a public information system for analyzing bioactivities of small molecules. Nucleic Acids Res. 37, W623–W633.

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

Weckwerth, W. (2011). Unpredictability of metabolism – the key role of metabolomics science in combination with next-generation genome sequencing. Anal. Bioanal. Chem. 400, 1967–1978.

Weckwerth, W., Loureiro, M. E., Wenzel, K., and Fein, O. (2004). Differential metabolic networks unravel the effects of silent plant phenotypes. Proc. Natl. Acad. Sci. U.S.A. 101, 7809–7814.

Wishart, D. S. (2009). Computational strategies for metabolite identification in metabolomics. Bioanalysis 1, 1579–1596.

Wishart, D. S., Knox, C., Guo, A. C., Eisner, R., Young, N., Gautam, B., et al. (2009). HMDB: a knowledgebase for the human metabolome. Nucleic Acids Res. 37, D603–D610.

Wolfgang, M., Haldiya, P. K., Willighagen, E., Kind, T., and Fiehn, O. (2010). The chemical translation service – a web-based tool to improve standardization of metabolomic reports. Bioinformatics 26, 2647–2648.

Wurtele, E. S., Chappell, J., Jones, A. D., Celiz, M. D., Ransom, N., Hur, M., et al. (2012). Medicinal plants: a public resource for metabolomics and hypothesis development. Metabolites 2, 1031–1059.

Xia, J., Mandal, R., Sinelnikov, I. V., Broadhurst, D., and Wishart, D. S. (2012). metaboAnalyst 2.0 – a comprehensive server for metabolomic data analysis. Nucleic Acids Res. 40, W127–W133.

Xiao, J. F., Zhou, B., and Ressom, H. W. (2012). MetaboSearch: tool for mass-based metabolite identification using multiple databases. PLoS ONE 7:e40096. doi:10.1371/journal.pone.0040096

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 19 December 2012; accepted: 15 March 2013; published online: 04 April 2013.

Citation: Fukushima A and Kusano M (2013) Recent progress in the development of metabolome databases for plant systems biology. Front. Plant Sci. 4:73. doi: 10.3389/fpls.2013.00073 This article was submitted to Frontiers in Plant Systems Biology, a specialty of Frontiers in Plant Science. Copyright © 2013 Fukushima and Kusano. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.