

ライフサイエンスデータベース統合推進事業
統合データ解析トライアル
研究開発課題「植物代謝物プロファイリングデータ
ベース AtMetExpress の開発と
オミックスデータ統合化の推進」

研究開発終了報告書

研究開発期間：平成25年9月～平成26年1月
研究代表者：福島 敦史
(独立行政法人理化学研究所 環境資
源科学研究センター 統合メタボロミク
ス研究グループ メタボローム情報研究
チーム、研究員)

§ 1 研究開発のねらい

【研究開発の目的】

ある生物種が持つ代謝物の総体（メタボローム）を測定し、有限個の代謝物情報へと絞り込むことはチャレンジングな問題である。本研究は、質量分析計を使って測定したモデル植物シロイヌナズナのメタボローム研究論文およびその代謝物プロファイルデータを整理し、シロイヌナズナのメタボロームデータベース（AtMetExpress と呼ぶ）を構築する。特に、計算機の取り扱いやデータマイニングを専門としないウェットの研究者（分子生物学者や植物研究者など）に対し、広く植物の代謝物プロファイルデータを閲覧・加工・共有するフレームワークの開発・情報基盤づくりを行う。想定するユーザは、(1) 自身のメタボロームデータについて過去データを参考にして解釈したい、(2) 遺伝子機能解析を進める段階で、検証可能な仮説の構築や実験計画にヒントを得たい人々である。メタボロームデータ統合手法の開発および NBDC の植物関連データベース、メタボローム関連データベースと統合することで、新たな知識発見につなげる。

【研究背景】

本研究が取り扱う代謝物プロファイルデータとは、ある生体由来サンプルにおける代謝物の蓄積量を指す（マススペクトルとは異なる）。このデータは、質量分析計から出力される生データを分析化学、化合物の構造情報およびバイオインフォマティクス技術によるデータ処理がなされたあと、代謝物蓄積量データ行列へと変換され利用可能となる。このデータ行列は、さらに下流の統計解析によってデータ解釈され、生物現象に関する仮説の構築へと貢献する。この代謝物プロファイルデータの共有は、マススペクトルあるいは化合物情報に比べて遅れをとっている。したがって、シロイヌナズナの代謝物プロファイルデータを例として、これらのデータ収集と多変量解析等の探索的データ解析による再解析によって、生物現象との結びつけを伴うデータ解釈と実験検証可能な仮説の構築に資するデータの共有とそのデータベースの開発が急務である。

これまでのいわゆるメタボロームデータベースの開発とは、主に MassBank (<http://www.massbank.jp/>) のような質量分析計のマススペクトルや KEGG (<http://www.genome.jp/kegg/>) および KNApSACk (<http://kanaya.naist.jp/KNApSACk/>) のような化合物情報に関するものであった。代謝物プロファイルデータベースの例として、カナダの HMDB (<http://www.hmdb.ca/>) および PMR (<http://metnetdb.org/PMR/>) が先行している。前者には、代謝物の構造情報やデータ解析ツールが備わっている一方、代謝物プロファイルデータへのアクセスには制限がある。また後者の PMR は、モデル植物シロイヌナズナ変異体と薬用植物等の代謝物プロファイルを公開しており、代謝物プロファイルデータへのアクセスがよく、またデータ解析ツールも備えている。研究代表者自身が開発中の代謝物プロファイルデータベース MeKO (<http://prime.psc.riken.jp/meko/>) も PMR と同コンセプトでかつ PMR に相補的な 50 種のシロイヌナズナ変異体に関する代謝物プロファイルデータを格納しており、データ解析ツールをも備えている。

【対象データとデータベース】

本研究は、以下を研究開発対象とする。

データ： シロイヌナズナ代謝物関係データおよび質量分析計由来のメタボロームデータ

データベース名： AtMetExpress

データ量： 14 データセット（代謝物プロファイルデータ）

約 1,200 代謝物（現時点で照合した結果であり、若干の重複が存在する）

代謝物プロファイルデータの生データサイズは、約 100MB/sample

代謝物蓄積量データ行列のサイズは、約 300KB/データ行列（データセット）

本研究課題に直接関連する NBDC データベース：

KNApSACk Core DB, MassBank, Bio-MassBank, metabolomics.jp, 植物ゲノム統合データベース (PGDBj)

§ 2 研究成果

【本研究で開発したデータベース】

データベース名: AtMetExpress

URL: <http://prime.psc.riken.jp/AtMetExpress/>

(別紙 1 にて、ツール機能、使用法、入出力例などを示した)

【研究開発の意義】

文献に基づいた代謝物—生物種に関するデータベース KNApSAcK を出発点として、文献に報告のあるシロイヌナズナ代謝物と、PMR と理研において質量分析計で検出された代謝物とを照合した。その結果、本研究は KNApSAcK に存在するシロイヌナズナで検出された報告のある代謝物(既知化合物数=約 600) と、現在の技術(質量分析計のみ)で実測できる代謝物との関係を明確にできた(上記ウェブサイトを参照のこと)。研究背景でも述べたように、代謝物プロファイルデータは、MassBank のような化合物の実測マススペクトルデータほどのデータ共有はなされていない。収集したデータの再解析の強力さを実証するには、さらなる時間と研究が必要であるが、本研究で開発した AtMetExpress はその第一歩となる。

【得られた効果】

- (1) 専門家以外の方々が気軽に代謝物プロファイルデータを閲覧できるサイトができた。
- (2) NBDC のメタボロームデータベース KNApSAcK や metabolomics.jp と部分的に統合することで、遺伝子—代謝表現型—表現型を貫くデータベースが構築可能となった。そのメタボロームデータについては、一貫した情報が共有可能となった。
- (3) 代謝物の蓄積パターンを中心としたデータマイニング法・データ再解析法の開発が部分的に試みられた。これによって、これまで未探索であった代謝レベルの有用知識発見につながるかもしれない。

§ 3 研究開発計画および計画に対する達成状況

(1) 達成状況

【研究計画概要と実施した研究項目】

1) 既存の化合物情報の収集・統合(平成 25 年 10 月)

シロイヌナズナに関して既存の化合物データベース KNApSAcK および AraCyc のデータベースにある化合物情報を統合した。同時に、PMR と理研で測定できたシロイヌナズナ・メタボロームデータとを統合した。具体的には、我々が開発した化合物 ID 統合ツール MetMask (Redestig et al., 2010) によって、CAS や KEGG ID などの化合物 ID が利用可能なものについて自動的に ID を統合した。現状で、いくつかの重複した代謝物についてマニュアルチェックが必要であるが、シロイヌナズナのメタボローム中、質量分析計で測定可能な代謝物数はおよそ 1,200 個であることがわかった。

2) 代謝物プロファイルデータの収集・再解析(平成 25 年 10 月・平成 26 年 1 月)

手始めに、理研でこれまで行われたシロイヌナズナの代謝物プロファイリングでカバーできる・できた化合物を整理した。現在、理研データのみ 14 個のデータセットを収集・整理した(未発表データ含)。先行して、50 種のシロイヌナズナ代謝物データベース MeKO(投稿中)を発表する予定である。

3) Wiki 化・データ視覚化・データ共有(平成 25 年 11 月・平成 26 年 1 月)

3-1) 代謝物情報は Wiki ページで管理している(別紙 1、Fig. 1 参照)。

3-2) 代謝物がシロイヌナズナのどの部位で検出されるか、蓄積するのかといった情報は、多変量

解析やヒートマップで表示できる（別紙 1、Fig. 2-3 参照）。

3-3) すべての生データ行列は NBDC および他の公共メタボロームデータベースへも登録すべく、データ共有を現在進めている。将来的にさらなる格納データの充実を図ることで植物システム生物データベースへと展開していく予定である。

【達成できなかった項目】

- 1) 代謝物名のユニーク化
- 2) 解析ツールの充実
- 3) 公共メタボロームデータベースとの連携

(2) ツールの将来性への展望

【研究開発のさらなる展開および成果の今後の展開見込】

AtMetExpress には、ユーザデータの解析を受け付けるツールを一部実装した（別紙 1、Fig. 4-5 参照）。このようなインタラクティブなデータ解析を可能としたことは、PMR など他データベースと比べても新規な点である。現時点で、さらなる発展プロジェクトとして以下の 3 つが考えられる：

- 1) メタボロームデータ解析・解釈をさらに自由自在にするプロジェクト
過去研究で利用可能な代謝物プロファイルデータを一か所に集める（データ離散を回避するため）。これらデータの二次加工を可能とする解析ツール、ユーザへのメリットが感じられるデータ解析ツールを提供し、データ解釈のためのノウハウ集めおよびデータ共有を促進する。
- 2) データ共有を推進するプロジェクト
データ共有する際の実験項目の詳細（メタデータ）の記述を簡便にするツール開発を行う。具体的には質量分析計ごと・メタボローム研究グループごとにメタデータのテンプレートを自動生成し、その雛形の中で実験ごとに変更する箇所を最小にする。これには、HUPO, Plant Ontology, Processing and separation ontology 等を採用したオントロジーで統一する必要があるとなる。
さらにメタボロームデータ共有のために次の問題を解決する必要がある：(A) データを共有するメリットをデータ生産者らに与える (B) 公共のレポジトリに登録する際の実験項目の詳細（メタデータ）の記述の煩雑さから解放する。
- 3) データ統合の強力を示すプロジェクト
他のオミックスレベルデータとの統合を行う。モデル植物シロイヌナズナを例にして、遺伝子機能解析を目的としたユーザのために、新型シーケンサによる RNA-seq トランスクリプトームおよびマイクロアレイデータとの統合を行う。具体的には、(1) ストレスへの発現応答や発達ステージに伴う遺伝子発現変化を表示する（発現差異解析）、(2) 発現相関（共発現）ネットワーク解析（クラスター解析）を想定している。NCBI の GEO など公共利用可能なデータの二次加工を経て、ユーザの対象遺伝子の機能についてヒントとなる候補遺伝子の提示する機能をもたせる。

さらなる代謝物プロファイルデータの「見える化」を行うことで、植物生理学、分析化学、代謝エンジニアリング、バイオインフォマティクスといった多様な背景を持つ学生・研究者のメタボローム分野への新規参入への貢献が期待できる。

【想定される統合化推進プログラムへの評価や発展への寄与】

本研究では、モデル植物シロイヌナズナが持つ代謝物を例にして、既存の化合物データ情報＋代謝物プロファイルデータの再解析＋柔軟な Wiki スタイルの WEB 公開を行った。生データ共有を目的とした既存のデータベースとして MetaboLights (<https://www.ebi.ac.uk/metabolights/>) と MetabolomeExpress.org (<https://www.metabolome-express.org/>) が機能している。また NBDC 統合化推進プログラム第一期においても、代謝物プロファイルの生データを格納・公開して

いる MassBase (<http://webs2.kazusa.or.jp/massbase/>) 等も整備されつつある。しかしながら、これらのデータが爆発的に蓄積している傾向は見られない。また分析化学の専門家ですら、これら生データを代謝物蓄積量データ行列に変換し、さらに二次加工するのは容易ではないのが現状である。その問題点は以下の通りである。

- (A) 上記レポジトリに生データを登録し、共有したとしても生物学者は活用できない
- (B) レポジトリに登録するには、細かいメタ情報を整理する必要があり、未だ煩雑である
- (C) そもそもデータ共有するメリットが不明（登録者に何の得もない！）

これらの問題点のうち (C) については少なくとも部分的に本研究で解決を試みた（ユーザデータのインタラクティブなデータ解析ツールの実装）。さらなる解決手段の開発がメタボロームデータの共有および他オミックスデータとの統合を促すであろう。本研究課題の研究期間内にその生合成経路に関する新たな知識発見を促すには至らなかったが、今後、さらに統合解析ツールの開発を促して遺伝子アノテーションの改善と NBDC データベース間の仲介をすることで、総合能力を引き出せるであろう。このような研究開発は将来の NBDC 統合化推進プログラムでサポートすべき重要な研究課題の一つであると期待している。

§ 4 研究参加者

氏名	所属	役職	研究開発項目	参加時期
○福島 敦史	独立行政法人 理化学研究所	研究員	研究開発立案、 データ収集、サイ ト構築、論文執筆	H25.10-H26.1

§ 5 成果発表等

- (1)原著論文発表（国内(和文)誌 0 件、国際（欧文）誌 0 件）

なし

- (2)その他の著作物(総説、書籍など)

なし

- (3)国際学会発表及び主要な国内学会発表

- ① 招待講演（国内会議 0 件、国際会議 0 件）

なし

- ② 口頭発表（国内会議 0 件、国際会議 0 件）

なし

- ③ ポスター発表（国内会議 0 件、国際会議 2 件）

1. Atsushi Fukushima, Makoto Suzuki, Makoto Kobayashi, Yozo Okazaki, Ryo Nakabayashi, Kenji Akiyama, Tetsuya Sakurai, Miyako Kusano, Masanori Arita, Kazuki Saito, “Development of metabolite-profiling database in Arabidopsis: AtMetExpress – Meta-analysis of metabolome data”, Asia Pacific Bioinformatics Conference (APBC) 2014, Shanghai, China, 17-19 January, 2014.
2. Atsushi Fukushima, Makoto Suzuki, Makoto Kobayashi, Yozo Okazaki, Ryo Nakabayashi, Kenji Akiyama, Tetsuya Sakurai, Miyako Kusano, Masanori Arita, Kazuki Saito, “Development of metabolite-profiling database in Arabidopsis: AtMetExpress – Meta-analysis of metabolome data”, The 2014 Winter q-bio Conference, The Island of Hawaii, USA,

17-20 February, 2014.

(4)知財出願

《本事業の成果に係わるものを、出願人(研究機関、JST、その他)に係わらず漏れなく記載してください。守秘契約や研究の進捗の関係などで公開を希望されないものについて、件数のみ記載しますので、下の＜非公開希望案件＞欄にハッチをかけて、”発明の名称、発明者、出願人、出願日、出願番号”を記載してください。》

①内出願 (0 件)

なし

②海外出願 (0 件)

なし

③その他の知的財産権

なし

(5)受賞・報道等

なし

§ 6 自己評価

本研究ではシロイヌナズナが持つ代謝物を例にして、柔軟な Wiki スタイルの化合物データ情報＋代謝物プロファイルデータの再解析ツールを備える AtMetExpress の開発を行った。達成したこととして、§3 で述べたように 1) 既存の化合物情報の収集・統合、2) 代謝物プロファイルデータの収集・再解析ツールの開発、3) Wiki 化・データ視覚化・データ共有を含んでいる。全体の研究計画に照らし合わせても当初の計画をはるかに超えた、高い地点まで到達したと自負している。

以上

別紙 1

Supplementary Document 1

Development of the metabolite profile database in Arabidopsis:

AtMetExpress

Atsushi Fukushima^{1,2*}, Ramon Francisco Mejia¹, Makoto Suzuki³, Makoto Kobayashi¹, Yozo Okazaki¹, Ryo Nakabayashi¹, Yutaka Yamada¹, Kenji Akiyama¹, Tetsuya Sakurai¹, Miyako Kusano^{1,4}, Masanori Arita^{1,5}, Kazuki Saito^{1,6}

¹ RIKEN Center for Sustainable Resource Science, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa, 230-0045, Japan

² JST, National Bioscience Database Center (NBDC), 5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081, Japan

³ RIKEN Plant Science Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa, 230-0045, Japan

⁴ Kihara Institute for Biological Research, Yokohama City University, Yokohama, Kanagawa 244-0813, Japan

⁵ National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan

⁶ Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, Chiba 260-8675, Japan

* Corresponding author: Atsushi Fukushima (Fax: +81-45-503-9489, e-mail: a-fukush@psc.riken.jp)

Version: 0.2

Date: Mar 12, 2014

Abstract

Background

Since early 2000s, a number of metabolome analyses have been demonstrated using combined hyphenated chromatographic and mass spectrometry as well as nuclear-magnetic-resonance [1-3]. Although these profiling experiments are becoming routine in many research groups, making the studies remains laborious and expensive (For example, see [4] and [5]). Therefore the number of samples and conditions analyzed in individual projects is still relatively small and mostly ranges from only two to few hundreds of samples (For example, see [6]). While the metabolomics community is working towards the setup of sharing of metabolome data [7-12], mining the public information and demonstrating the

richness of integration of multiple metabolite profile datasets remain largely unexploited.

Description

The AtMetExpress is a freely available database including detailed information about small molecule metabolites detected in the model plant *Arabidopsis thaliana* (Arabidopsis). This is intended to be used for applications in metabolomics, functional genomics, biomarker discovery, and scientific education. The aims of the project are (i) to establish a new web-based platform to conduct meta-analysis of metabolite profile datasets and (ii) to explore the diversity of complex metabolic networks using the multiple profiling experiments in Arabidopsis. We integrated multiple metabolome dataset in Arabidopsis and constructed AtMetExpress to store the information. By analyzing the integrated data we found that Arabidopsis has ~1,200 metabolites which we can detect using mass spectrometry-based metabolite profiling. The current AtMetExpress also contains 'development LC-MS dataset' by Matsuda and colleagues [13], which is opened to the public via the PRIME web site (<http://prime.psc.riken.jp/lcms/AtMetExpress/>).

Conclusions

A small and simple GUI tool in AtMetExpress has built for configuring and performing these analyses, allowing easy metabolome meta-analysis for plant biologists. We demonstrate how meta-analysis of metabolite profile datasets can be used for new hypothetical generations of metabolic behavior in Arabidopsis across a wide range of experiments.

URL: <http://prime.psc.riken.jp/AtMetExpress/>

Keywords: Metabolome, Database, Meta-analysis, Metabolite profiling, Arabidopsis thaliana

Construction and implementation of AtMetExpress

We collected 11 articles related to mass spectrometry-based metabolomics in Arabidopsis from our previous publication and 3 unpublished dataset. For metabolite information we integrated compound identifiers from AraCyc [14] and KNApSack [15] using MetMask tool [16]. Then we manually curated these datasets, and have built AtMetExpress tools by using R/Bioconductor [17] packages including shiny package (<http://www.rstudio.com/shiny/>).

Overview, utility, and general characteristics of AtMetExpress

There are two major options available in the AtMetExpress: (i) Browse and (ii) Analysis.

Browse

The browse option provides a Wiki interface for a number of metabolites to the database, allowing us to search the database (Fig. 1). Each entry is linked back to other metabolome databases such as PubChem [18], KEGG [19], and KNApSack [15] for more information. The resultant data from statistical data analyses (tab-delimited (tsv) format) are also available for download.

The screenshot shows the 'AtMet:Malate::1' page. It includes a sidebar with navigation links like 'Main Page', 'Recent changes', and 'Random page'. The main content area has two tables: 'Malate' and 'IDs and Links'. The 'Malate' table lists properties such as 'Suzukicheck', 'Pmdetect', 'PmdetectAraCycID', 'Pmdetectname', 'Pmdetectkegg', 'KNApSack-name', 'Molecular_formula', 'MW_calculate_by_MWOC', 'Organism', 'datasetname', and 'each_dataset_Uniq_ID'. The 'IDs and Links' table lists identifiers from 'PUBCHEM', 'AraCyc_ID', 'CAS_ID', 'KEGG', and 'KNApSack-ID'. The page also has a 'Log in' button, a search bar, and a footer indicating the page was last modified on 7 February 2014.

Malate	
Suzukicheck	Pmdetect
PmdetectAraCycID	malic acid
Pmdetectname	Malate
Pmdetectkegg	malic acid
KNApSack-name	L-malic acid,636-61-3
Molecular_formula	C4H6O5
MW_calculate_by_MWOC	134.0215226
Organism	Arabidopsis thaliana
datasetname	17128_Nikolau_DataSheet2
each_dataset_Uniq_ID	PM714

IDs and Links	
PUBCHEM	222656
AraCyc_ID	MAL
CAS_ID	97-67-6
KEGG	C00149
KNApSack-ID	C00001192

Fig. 1. The molecule view of malate. Each entry is linked back to other metabolome databases such as PubChem [18], KEGG [19], and KNApSack [15] for more information. The page is accessible at <http://metabolomics.jp/wiki/AtMet:Malate::1>. Other characteristics of Wiki-style pages are according to Arita and Suwa [20].

Analysis

The Analysis option provides an interface for visualizing AtMetExpress datasets,

for analyzing user's data on our website, and for getting their results. The following information can be obtained: i) quality check of sample replicates by principal component analysis (PCA) (Fig. 2); ii) Visualization of their own data by using our AtMetExpress tools [e.g., hierarchical cluster analysis (HCA) and a heatmap] (Fig. 3) and iii) meta-analysis tools including metabolite name integration (Fig. 4 and Fig. 5): a tool for integrating metabolite identifiers from local reference libraries based on SQLite and public databases that do not depend on a single common primary identifier. Users can use this option to integrate two data matrices in GENERIC format.

PRIME Visualization Tools

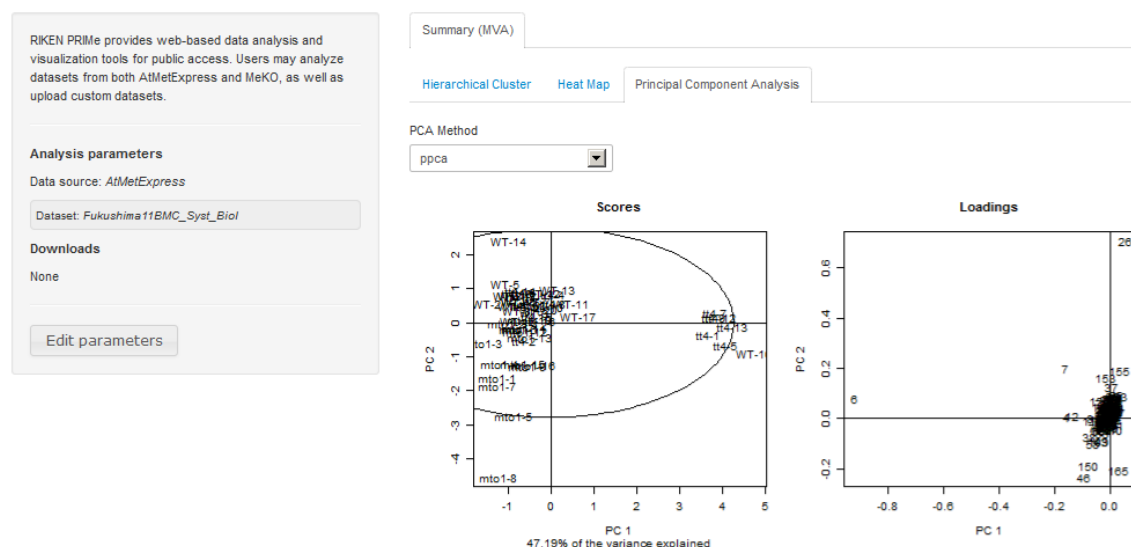


Fig. 2. Principal component analysis of the roots of 3 *Arabidopsis* genotypes [Col-0, *methionine-over accumulation 1* (*mto1*), and *transparent testa 4* (*tt4*)]. Score scatter and loading plots for roots using all detected peaks are shown. These data include 53 root samples (WT, $n = 17$; *mto1*, $n = 16$; and *tt4*, $n = 20$). The data were from the work of Fukushima et al. [21].

PRIME Visualization Tools

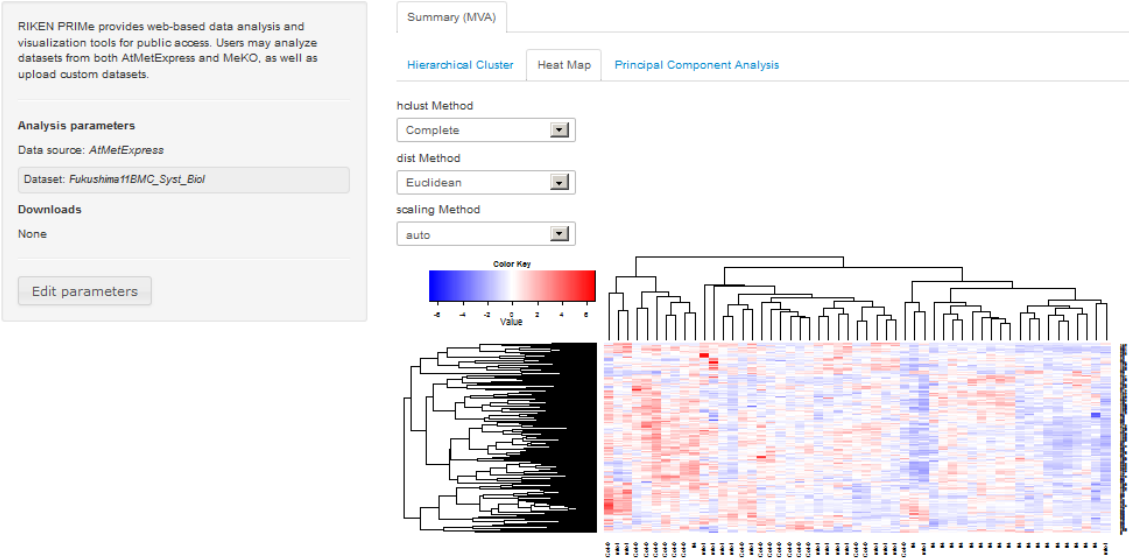


Fig. 3. Visualization of the depositing data by using our AtMetExpress tools [e.g., a heatmap (bottom panel)]. Red and blue colors represent relative increased or decreased metabolite levels compared to median values, respectively. The data were from the work of Fukushima et al. [21].

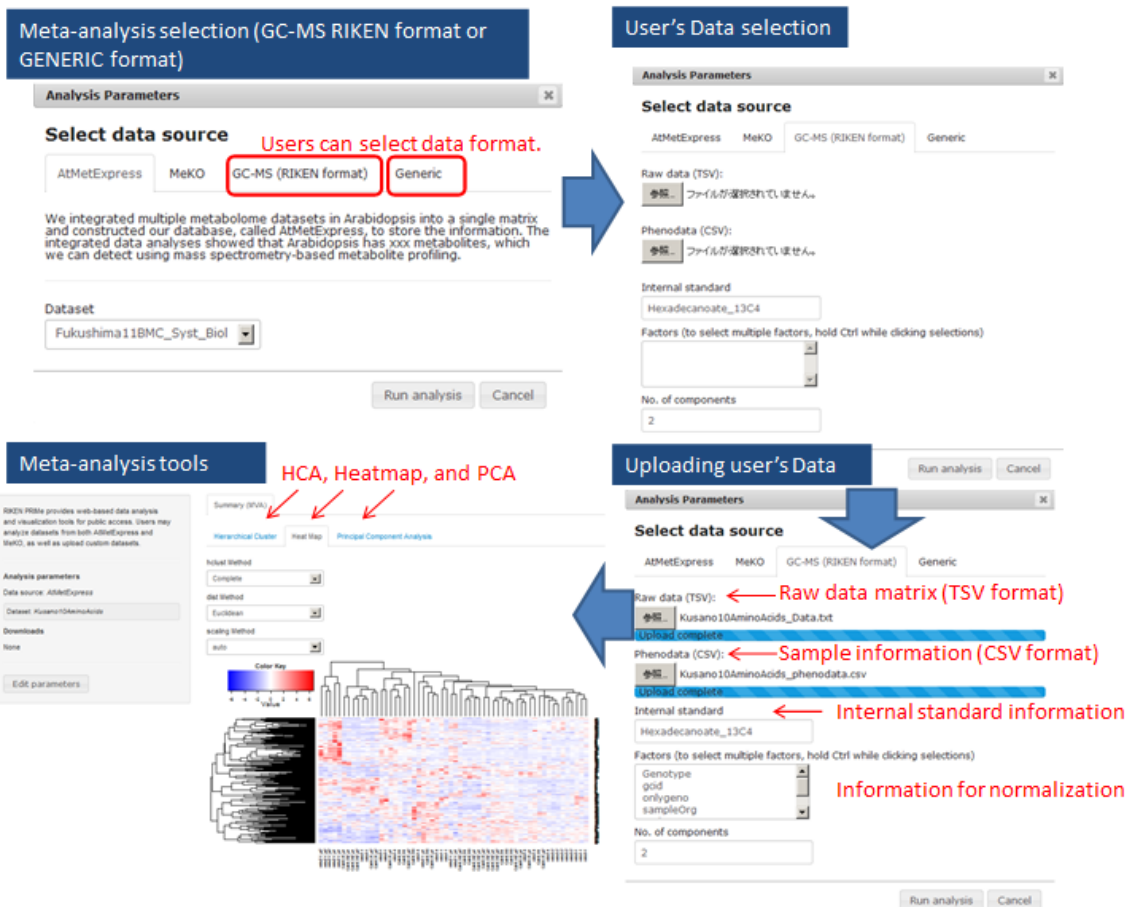


Fig. 4. Meta-analysis workflow of user's own data by using our AtMetExpress tools in GC-MS RIKEN format. an example format is given in AtMetExpress site (<http://prime.psc.riken.jp/AtMetExpress/exdata/exdata.zip>). With raw data matrix and their sample information, users can do meta-analysis of their own data including integration of different datasets in GENERIC format (see Fig. 5).

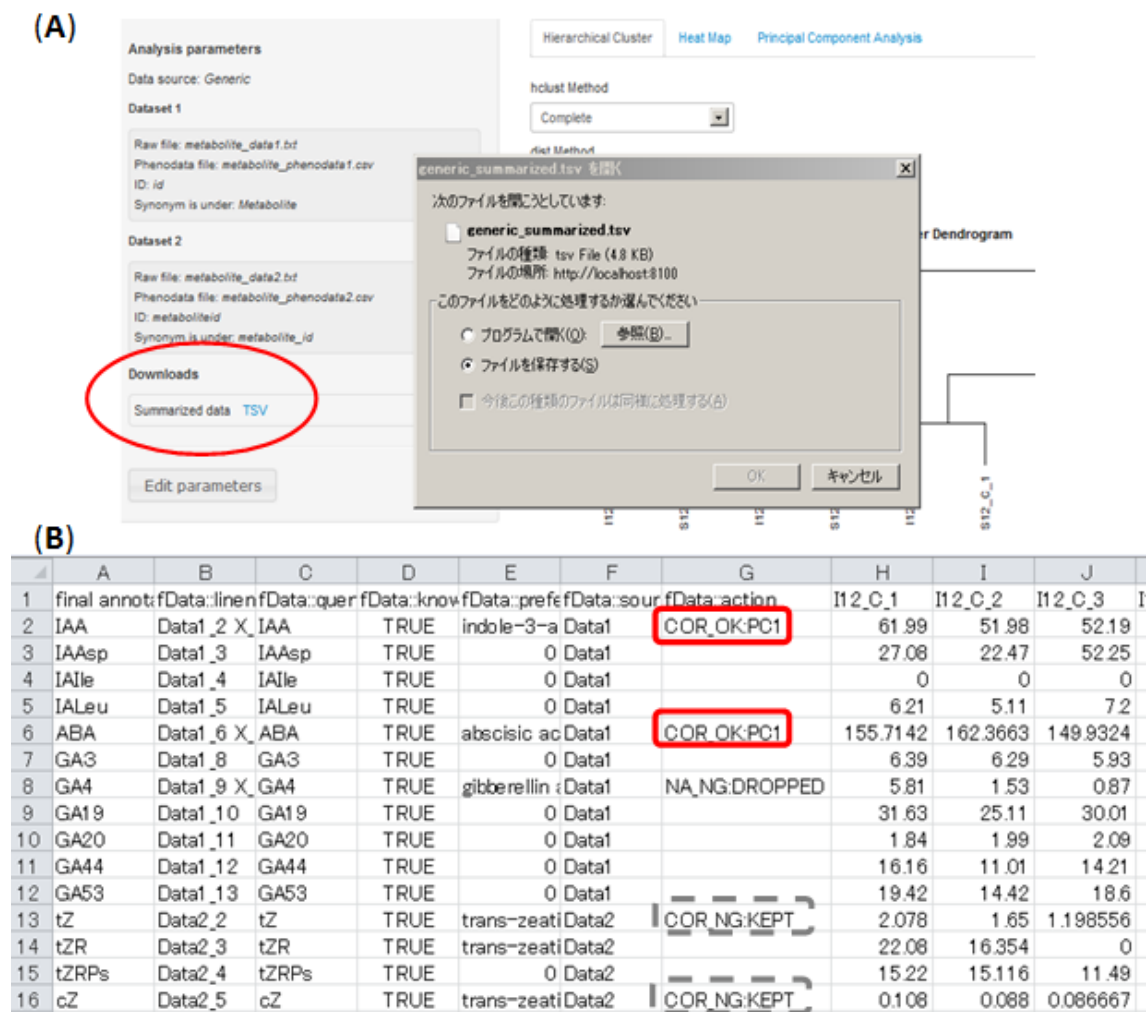


Fig. 5. An example of GENERIC analysis. Users can upload their own data to AtMetExpress in GENERIC format (Typically the rows and columns represent metabolite names and samples, respectively). Automatically integrated/summarized data are downloadable using GENERIC tab (A). You can see the summarized data matrix using principal component analysis (B). Highly correlated metabolites between datasets are replaced by principal component 1 (e.g., IAA and ABA with red rectangles), while weakly correlated metabolites are not replaced by the same manner (e.g., two cytokinins, tZ and cZ with dashed rectangles).

Acknowledgments

This work was partly supported by Japan Science and Technology Agency (JST)–Strategic International Collaborative Research Program (SICORP), Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and by Japan Advanced

Plant Science Network. Research activity is also partly supported by a Grant-in-Aid for Young Scientists (B; grant no. 23700355 to A.F.) from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We would like to thank Dr. Henning Redestig (RIKEN) for R programing and computational assistance.

References

1. Scalbert A, Andres-Lacueva C, Arita M, Kroon P, Manach C, Urpi-Sarda M, Wishart D: **Databases on food phytochemicals and their health-promoting effects.** *J Agric Food Chem* 2011, **59**(9):4331-4348.
2. Fiehn O, Barupal DK, Kind T: **Extending biochemical databases by metabolomic surveys.** *J Biol Chem* 2011, **286**(27):23637-23643.
3. Fukushima A, Kusano M: **Recent progress in the development of metabolome databases for plant systems biology.** *Front Plant Sci* 2013, **4**:73.
4. Lei Z, Huhman DV, Sumner LW: **Mass spectrometry strategies in metabolomics.** *J Biol Chem* 2011, **286**(29):25435-25442.
5. Bais P, Moon-Quanbeck SM, Nikolau BJ, Dickerson JA: **Plantmetabolomics.org: mass spectrometry-based Arabidopsis metabolomics--database and tools update.** *Nucleic Acids Res* 2012, **40**(Database issue):D1216-1220.
6. Fernie AR, Aharoni A, Willmitzer L, Stitt M, Tohge T, Kopka J, Carroll AJ, Saito K, Fraser PD, DeLuca V: **Recommendations for reporting metabolite data.** *Plant Cell* 2011, **23**(7):2477-2482.
7. Sansone SA, Fan T, Goodacre R, Griffin JL, Hardy NW, Kaddurah-Daouk R, Kristal BS, Lindon J, Mendes P, Morrison N et al: **The metabolomics standards initiative.** *Nat Biotechnol* 2007, **25**(8):846-848.
8. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TW, Fiehn O, Goodacre R, Griffin JL et al: **Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI).** *Metabolomics* 2007, **3**(3):211-221.
9. Carroll AJ, Badger MR, Harvey Millar A: **The MetabolomeExpress Project: enabling web-based processing, analysis and transparent dissemination of GC/MS metabolomics datasets.** *BMC Bioinformatics* 2010, **11**:376.

10. Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, Mahendrakar T, Williams M, Neumann S, Rocca-Serra P *et al*: **MetaboLights--an open-access general-purpose repository for metabolomics studies and associated meta-data.** *Nucleic Acids Res* 2013, **41**(Database issue):D781-786.
11. Salek RM, Haug K, Conesa P, Hastings J, Williams M, Mahendrakar T, Maguire E, Gonzalez-Beltran AN, Rocca-Serra P, Sansone SA *et al*: **The MetaboLights repository: curation challenges in metabolomics.** *Database (Oxford)* 2013, **2013**:bat029.
12. Sakurai T, Yamada Y, Sawada Y, Matsuda F, Akiyama K, Shinozaki K, Hirai MY, Saito K: **PRIME Update: innovative content for plant metabolomics and integration of gene expression and metabolite accumulation.** *Plant Cell Physiol* 2013, **54**(2):e5.
13. Matsuda F, Hirai MY, Sasaki E, Akiyama K, Yonekura-Sakakibara K, Provart NJ, Sakurai T, Shimada Y, Saito K: **AtMetExpress development: a phytochemical atlas of Arabidopsis development.** *Plant Physiol* 2010, **152**(2):566-578.
14. Zhang P, Foerster H, Tissier CP, Mueller L, Paley S, Karp PD, Rhee SY: **MetaCyc and AraCyc. Metabolic pathway databases for plant research.** *Plant Physiol* 2005, **138**(1):27-37.
15. Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Altaf-Ul-Amin M, Darusman LK *et al*: **KNAPSAcK family databases: integrated metabolite-plant species databases for multifaceted plant research.** *Plant Cell Physiol* 2012, **53**(2):e1.
16. Redestig H, Kusano M, Fukushima A, Matsuda F, Saito K, Arita M: **Consolidating metabolite identifiers to enable contextual and multi-platform metabolomics data analysis.** *BMC Bioinformatics* 2010, **11**:214.
17. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J *et al*: **Bioconductor: open software development for computational biology and bioinformatics.** *Genome Biol* 2004, **5**(10):R80.
18. Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH: **PubChem: a public information system for analyzing bioactivities of small molecules.** *Nucleic Acids Res* 2009, **37**(Web Server issue):W623-633.

19. Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M: **Data, information, knowledge and principle: back to metabolism in KEGG.** *Nucleic Acids Res* 2014, **42**(1):D199-205.
20. Arita M, Suwa K: **Search extension transforms Wiki into a relational system: a case for flavonoid metabolite database.** *BioData Min* 2008, **1**(1):7.
21. Fukushima A, Kusano M, Redestig H, Arita M, Saito K: **Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach.** *BMC Syst Biol* 2011, **5**:1.