

平成 24 年度 研究開発実施報告書

ライフサイエンスデータベース統合推進事業「統合化推進プログラム」

研究代表者

金谷 重彦

奈良先端科学技術大学院大学・情報科学研究科・教授

メタボローム・データベースの開発

§1. 研究実施体制

(1)「奈良先端大グループ」グループ(研究機関別)

① 研究代表者: 金谷 重彦 (奈良先端大学院大学情報研究科、教授)

② 研究項目

・質量スペクトル DB(化合物 MSDB とメタボローム MSDB)の拡充

化合物 MS DB の開発

・代謝物質情報 DB の構築

代謝物質と生物活性の関係データベース

MS データと化学構造の関係知識 DB と化学構造式推定ツールの開発・実装

(2)「かずさ DNA 研究所」グループ(研究機関別)

① 主たる共同研究者: 櫻井望 ((公財)かずさ DNA 研究所 産業基盤開発研究部 研究員)

② 研究項目

・メタボローム・データベースの構築

(3)「かずさ DNA 研究所」グループ(研究機関別)

① 主たる共同研究者: 有田 正規 (理化学研究所・植物科学研究センター・ユニットリーダー)

② 研究項目

・メタボローム・データベースの構築



§2. 研究実施内容

目的

メタボロミクスが他のオミクスと比較して大きく立ち後れている点は、一般研究者が利用できる公開データ(実験から得られた定量値を伴う網羅的な代謝物検出データ(メタボロームデータ))が不足している点である。初年度は、メタボロームデータの公開・利用を促進する目的でデータの標準フォーマットを策定した。平成 24 年度は、この標準フォーマットを用い、メタボロームデータを公開するに当たり最大の律速の一つとなっているメタデータ(実験手法の詳細情報)の収集を効率化させるシステムを構築して、データ公開を加速させ、さらに、代謝物構造ならびに活性情報との統合を図ることを目的にデータベース(DB)開発を進めた。

概要

化合物に関する知識ベースの情報は近年急速に拡充されている(KNApSAcK、KEGG 等)。このような状況の中、本プロジェクトでは、生物種・代謝物を KNApSAcK Core DB^{3,4)}に、また、個々の代謝物情報を KNApSAcK Metabolite Activity DB^{7,8)}に格納した。また、特定生物で検出された化合物の定性的な情報も整備されつつある(ヒト:HMDB、酵母:YMD 等)。しかし、遺伝子発現における GEO(NCBI)や ArrayExpress(EBI)に相当するような、定量値を伴う代謝化合物のプロファイルデータの蓄積は遅れており、一般からデータを収集する集約的なデータベースとしては MetaboLights(EBI)が 2012 年に初めて公開されたに過ぎない。この主な要因は、複雑なデータ形式を整形し、膨大な実験手法の情報を収集・整理して、メタボロームデータを公開すること自体に、大きな労力がかかっていることが挙げられる。初年度は、データの作成と利用の両者を円滑に行うことを目的に、データの標準フォーマットを策定して、この形式に従ったデータの提供を開始した。本年度は、もう一つの大きな律速である、膨大な実験手法の情報(メタデータ)の収集・整理に対して、これを効率化するシステムの構築を第一の目標とした。1) 可能な限り簡便にメタデータを記載でき、2) メタデータの階層構造を管理することができ、3) 異なる目的の複数の実データデータベース(生データ:MassBase、メタボロームデータ:KomicMarket、MS スペクトルデータ:MassBank、Bio-MassBank 等)から共通して参照でき、さらに4) RDF 化による他データベースとの連携もできるシステムとして、セマンティック MediaWiki をプラットフォームとして用いた、メタデータだけを専用で取り扱うデータベースの構築を行った。一方で、データ処理の効率化とデータの公開も進めた。また、質量分析データにおける精密質量から分子式を推定する方法を考案し¹⁾、データベース構築の過程の効率化を図った。

進捗状況

メタデータを専用で取り扱うデータベースとして、Metabolonote を完成させ、2012 年 12 月 28 日に一般公開した。また、データ処理パイプラインの効率化を図った。特に、PowerGet ソフトウェアに標準フォーマットでの出力機能を追加したことで、ピークデータの整形が劇的に改善された。精密質量値から迅速に組成式推定を行う MFSearcher システムは、さらなる高速化を達成し、従



来法の 100～2000 倍を達成した。Metabolonote によりメタデータの整理が効率化されたことで、12 月以降加速的にデータを搭載し、16 件の実験プロジェクトのメタデータを公開した。2012 年 2 月から公開されている MetaboLights が現在 16 件であることを考慮すると、効率の良いデータ公開が実現できていると言える。このメタデータを参照する実データとして、KomicMarket より 111 分析のメタボロームデータ(202,939 ピーク)を新規に公開し、51 分析に由来する 53,236 件の MS スペクトルを Bio-MassBank より公開した。さらに、質量分析の生データを蓄積する MassBase からは、本年度 33,294 件の分析データを公開し、昨年度との総計で 43,658 件の公開を 2 年間で達成した。これは、本プロジェクト 3 年間での目標値 36,000 件を大きく上回る公開数となった。

研究成果

代謝物-生物種関係データベース KNApSAcK Core DB では、現在 106,417 対の生物種-代謝物の関係が格納され、また、9450 レコードの代謝物の活性情報が KNApSAcK Metabolite Activity DB に格納されるに至っている。これらのデータベースはオミックスプラットフォームとしてのメタボローム・データベースとして国際誌ならびに国内誌に公表するに至っている²⁻⁸⁾。MFSearcher システムについて Bioinformatics 誌に論文公開を行った¹⁾。

今後の見通し

効率化されたメタデータ整理システム、データ処理パイプラインを用いて、引き続きデータ拡充を行う。また、セマンティック検索を利用し、Metabolonote を介して MassBase、KomicMarket と Bio-MassBank、KNApSAcK との連携を強化して、一般研究者がメタボローム実験データを効果的に利用するための基盤整備を進める。さらに、植物データベースのゲノム情報による統合チーム(代表:田畑哲之博士)の PGDBj との連携も進める。論文や学会発表、出展等により、広報・利用促進活動を積極的に行う。本プロジェクトで開発を進めている MassBank、Bio-MassBank、KNApSAcK Core、KNApSAcK Metabolite Activity のデータの充実を図るとともに、質量分析データにもとづいた構造推定の効率化を図り、メタボロームプラットフォームとしての世界標準を目指す。



§3. 成果発表等

(3-1) 原著論文発表

- ① 発行済論文数(国内(和文) 1件、国際(欧文) 4件):
- ② 未発行論文数(“accepted”、“in press”等)(国内(和文) 0件、国際(欧文) 3件)
- ③ 論文詳細情報
 1. Nozomu Sakurai, Takeshi Ara, Shigehiko Kanaya, Yukiko Nakamura, Yoko Iijima, Mitsuo Enomoto, Takeshi Motegi, Koh Aoki, Hideyuki Suzuki and Daisuke Shibata "Application of a relational database system for high-throughput prediction of elemental compositions from accurate mass values", *Bioinformatics*, vol. 29, No. 2, pp.290-291, 2013 (DOI: 10.1093/bioinformatics/bts660)
 2. Farit Mochamad Afendi, Tetsuo Katsuragi, Akira Kato, Noritaka Nishihara, Kensuke Nakamura, Kensuke, Yukiko Nakamura; Ken Tanaka, Morita Morita, Aki, Altaf-Ul-Amin, Hiroki Takahashi, Shigehiko Kanaya, Systems biology approaches and metabolomics for understanding Japanese traditional Kampo medicine, *Current Pharmacogenomics and Personalized Medicine*, 10, 2, 111-124, 2012 (DOI: なし)
 3. Farit Mochamad Afendi, Taketo Okada, Mami Yamazaki, Aki Hirai-Morita, Yukiko Nakamura, Kensuke Nakamura, Shun Ikeda, Hiroki Takahashi, Md. Altaf-Ul-Amin, Latifah K. Darusman, Kazuki Saito and Shigehiko Kanaya, KNApSAcK family databases: Integrated metabolite plant species databases for multifaceted plant research, *Plant and Cell Physiology*, 53, 2, 1-12, 2012 (DOI: 10.1093/pcp/pcr165)
 4. Kensuke Nakamura, Naoki Shimura, Yuuki Otabe, Aki Hirai-Morita, Yukiko Nakamura, Naoaki Ono, Md Altaf Ul-Amin and Shigehiko Kanaya, KNApSAcK-3D: A Three-dimensional structure database of plant metabolites, *Plant Cell Physiol.*, 54, 2, e4(1-8), 2012 (DOI: 10.1093/pcp/pcs186)
 5. 池田俊, 桂樹哲雄, 小野直亮, 中谷淳至, 中村由紀子, 森田晶, 金谷重彦, オミックス・プラットフォーム: バイオ・ビッグ・データに挑む, *生物工学会誌*, 90, 12, 777-781, 2012
 6. Farit M. Afendi, Latifah K. Darusman, Aki Hirai Morita, Md. Altaf-Ul-Amin, Hiroki Takahashi, Kensuke Nakamura, Ken Tanaka, Shigehiko Kanaya, Efficacy Prediction of Jamu Formulations by PLS Modeling, *Current Computer-Aided Drug Design*, 9, 4, 46-59, 2013 (DOI: in press)
 7. Shun Ikeda, Takashi Abe, Yukiko Nakamura¹, Kibinge Nelson, Aki Hirai Morita, Atsushi Nakatani, Naoaki Ono¹, Toshimichi Ikemura³, Kensuke Nakamura⁴, Md. Altaf-Ul-Amin¹, Shigehiko Kanaya, Systematization of the protein sequence diversity in enzymes related to secondary metabolic pathways in plants, in the context of big data biology inspired by the KNApSAcK Motorcycle database, *Plant Cell Physiol.*, (in press) 2013 (DOI: in press)



8. Farit M. Afendi, Naoaki Ono, Yukiko Nakamura, Kensuke Nakamura, Latifah K. Darusman, Nelson Kibinge, Aki Hirai Morita, KenTanaka, Hisayuki Horai, Md. Altaf-Ul-Amin, Shigehiko Kanaya, Data Mining Methods for Omics and Knowledge of Crude Medicinal Plants toward Big Data Biology, Comput. Struct. Biotech. J., 4, e20131010.1-14, 2013 (DOI: なし)

(3-2) データベースおよびウェブツール等の構築と公開

公開中のデータベース・ウェブツール等
別紙を参照

(3-3) その他の著作物(総説、書籍など)

なし

(3-4) 国際学会および国内学会発表

① 招待講演 (国内 2 件、国際 3 件)

〈国内〉

1. 荒武、MassBase 活用術～ドライからウェットまで、第 7 回メタボロームシンポジウム、慶應義塾大学鶴岡メタボロームキャンパス、2012 年 10 月 12 日
2. 柴田大輔、食品科学へのメタボロームデータの応用、平成 24 年度 H-InvDB データベース講習会、産業技術総合研究所 臨海副都心センター、2013 年 1 月 24 日

〈国際〉

- *1 Nishioka, T., "Introduction to MassBank", NORMAN MassBank workshop (NORMAN Association 主催), IVM VU University Amsterdam, Amsterdam, Netherland, November 27, 2012.
2. 金谷重彦 (NAIST), Jamu Database and Bioinformatics, Jamu-Informatic Training, ジャカルタ, 2012, 10
3. 金谷重彦 (NAIST), Jamu Database and Bioinformatics, National Seminar on Cultural Aspects, Policy and Science Philosophy of Jamu, ジャカルタ, 2012, 10

② 口頭講演 (国内 8 件、国際 3 件)

〈国内〉

1. 荒武、櫻井望、鈴木秀幸、柴田大輔、大規模メタボロームデータベース MassBase を用いた植物メタボローム解析、第 30 回植物細胞分子生物学会、奈良先端科学技術大学院大学(奈良)、2012 年 8 月 4 日
2. 櫻井望、鈴木秀幸、柴田大輔、植物オミクス解析のためのウェブツール KaPPA-View4 ファミリー、第 30 回植物細胞分子生物学会、奈良先端科学技術大学院大学(奈良)、2012 年 8 月 4 日



3. 金谷重彦 (NAIST), カレーライスのメタボロミクス? 食品化学にも活用できるオミックス手法とデータベース~, 日本食品化学学会 第18回総会・学術大会, 北海道, 2012, 6
4. 金谷重彦 (NAIST), KNApSAcK Family Database Extended to Comprehensive Understanding of Bio-machinery, International Conference of Natural Products Biosynthesis, 淡路市, 2012, 6
5. 金谷重彦 (NAIST), KNApSAcK Family DB: おミックス研究における医食同源の体系化, 2012年度第1回CACフォーラムセミナー, 大阪府, 2012, 7
6. 金谷重彦 (NAIST), KNApSAcK Family DB: おミックス研究における医食同源の体系化, BioJapan2012, 神奈川, 2012, 10
7. 金谷重彦 (NAIST), KNApSAcK ファミリーデータベース: メタボロミクスから展開する植物の多目的活用, トーゴの日シンポジウム 2012, 東京, 2012, 10
8. 西岡孝明, "高精度なマススペクトルを MassBank で活用する", 第7回メタボロームシンポジウム, 慶應義塾大学鶴岡メタボロームキャンパス, 山形県鶴岡市, 2012.10

〈国際〉

1. 金谷重彦 (NAIST), KNApSAcK Family Database Extended to Comprehensive Understanding of Bio-machinery, International Conference of Natural Products Biosynthesis, 淡路市, 2012, 6
2. Nishioka T, Nihei Y, Ojima Y, & Ikeda T. "MassBank: Public Mass Spectral Database", 19th International Mass Spectrometry Conference, Kyoto International Conference Center, Kyoto, Japan, September 15-21, 2012.
3. Nishioka T. "Constructing MassBank Database on PCs", 19th International Mass Spectrometry Conference, Workshop "Mass++ and MassBank: Tools for Data Processing and Database on PC", Kyoto International Conference Center, Kyoto, Japan, September 15-21, 2012.

③ ポスター発表 (国内 2件、国際 1件)

1. 荒武, 榎本光男, 有田正規, 金谷重彦, 櫻井望, Metabolonote: メタボロームデータの連携と高度利用を目指した Wiki ベースのメタデータ管理システムの構築, トーゴの日シンポジウム 2012, 時事通信ホール(東京), 2012年10月5日
2. 池田奨, 二瓶義人, 尾寫雄也, 西岡孝明, "MassBank のツール拡充と信頼性向上への取り組み", 第7回メタボロームシンポジウム, 慶應義塾大学鶴岡メタボロームキャンパス, 山形県鶴岡市, 2012.10

〈国際〉

1. Nishioka T, Nihei Y, Ojima Y, & Ikeda T. " MassBank: Public Repository for Sharing Mass Spectral Data of Metabolomics", 8th International Conference of the



Metabolomics Society, Marriot Wardman Park Hotel, Washington, DC, USA. June 25-28, 2012.

(3-5) 知財出願

①平成 24 年度の累積件数(国内 0 件、海外 0 件)

なし

②平成 23、24 年度の累積件数(国内 0 件、海外 0 件)

なし

③その他の知的財産権

なし

(3-6) 受賞・報道等

① 受賞

*1. 2012 年度 日本植物細胞分子生物学会 技術賞、大場利治、佐藤将一、辻本善政、北川正成、櫻井望、2012 年 8 月 4 日

*2. 2012 年度 日本植物細胞分子生物学会 論文賞、Kosuke Kai, Hiroki Takahashi, Hirohisa Saga, Takumi Ogawa, Shigehiko Kanaya, Daisaku Ohta, 8 月 4 日

② 新聞報道

1. 毎日新聞、植物が含む5万の物質、DB化、5月29日

2. HJ 健康食品新聞、進むメタボロミクス研究、6月27日

3. IPB (ボゴール農大ニュース・インドネシア)、IPB kembangkan “Database” Jamu、10月23日

③ その他の成果発表

なし



§4. 研究開発期間中に主催した活動(主催したワークショップ等)

年月日	名称	場所	参加人数	概要
2013年3月22日	第54回日本植物生理学会年会データベース講習会	岡山大学津島キャンパス	80人	オーガナイザーの一人として櫻井が参加した。 植物関連のデータベースに関する話題として、PGDBjの市原寿子博士、KNApSAcKファミリーの話題提供をした。
2012年9月15-21日	19th International Mass Spectrometry Conference	Kyoto International Conference Center, Kyoto	44人	「国際学会におけるシンポジウムのオーガナイザー(共催)」, Japan,において Workshop “Mass++ and MassBank: Tools for Data Processing and Database on PC”, を企画した(オーガナイザー)。

