

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  
研究開発課題「メタボローム・データベースの開発」

研究開発終了報告書

研究開発期間：平成23年4月～平成26年3月  
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)



## § 1 研究開発実施の概要

### (1) 実施概要

メタボローム研究においては、質量分析装置により出力されるピークデータをもとに代謝物の推定を行う。さらにこれらの代謝物についての機能を注釈づけることが必要とされる。そこで、本プロジェクトでは、質量分析装置から得られるピークデータ、ピークデータと代謝物の関係づけ、さらには、代謝物の機能注釈についてのデータベース構築ならびにツール群の研究開発および、低コストで継続維持可能なデータベース・システムの研究開発を以下の三つの項目に焦点をあてて 3 グループ(奈良先端大、かずさ DNA 研、理研)で密に連絡をとりながら進めた。

#### 1. 質量スペクトル DB(化合物 MSDB とメタボローム MSDB)の拡充

##### 1.1 化合物 MS DB の開発(奈良先端大、理研)

2011年4月以降、米国、スイス、ドイツ、日本の6研究グループが高精度に m/z を測定した約 9 千件のデータを公開した。本 MSDB へのアクセスは 1 千件/日 (unique IP address) を超えている。EU の 17 加盟国の環境科学研究機関が加盟する NORMAN Association は MassBank を公式 MS DB とした。これを契機として、ドキュメント管理システム(Subversion)を利用したミラーサーバ MassBank.eu をドイツに設置した。MassBank.jp と MassBank.eu は独立に化合物 MSDB を収集してデータを交換する対等な関係にあり、これまでの主従関係に基づくミラーサーバとは一線を画する。2 つのサーバが常に全 MassBank レコードの最新の同一コピーを所有している。さらに、分散型 DB の欠点とされた「どれか1つのデータサーバがサービスできないときには検索結果が異なる」を解決し、「常に全てのレコードを検索した結果をユーザに返す」安定なシステムを実現した。また全 MassBank レコードのコピーを誰もがダウンロードできることを実現した。このコピーを利用して、創造的な研究開発が生まれることを期待する。

##### 1.2 メタボローム MS DB の開発(奈良先端大、かずさ DNA 研、理研)

MassBank 開発キットをベースにした「メタボローム MSDB (public repository)」である Bio-MassBank を開発した。これまでに、かずさ DNA 研究所が植物 10 種(29 栽培種)、微生物 3 種、哺乳動物 1 種を LC-MS, MS<sup>2</sup> 分析したデータ 51 分析に由来する 53,236 件のマススペクトルを Bio-MassBank より公開した。これによって、データセット間で類似したマススペクトルを手がかりとして共通な未知代謝物を容易に検索することができるようになった。

#### 2. 代謝物情報 DB の構築

##### 2.1 代謝物と生物活性の関係データベース(奈良先端大)

メタボローム研究の効率化に寄与する目的で、学術論文を地道に調査し、データを蓄積し、生物種と代謝物の関係データベース KNApSAcK Core DB を開発した。現在まで、101,500 対の生物種-代謝物の関係、5 万種の代謝物、2 万種の生物が格納されており、世界のメタボローム研究者に認知され、現在までに、Nature 姉妹誌などの文献を含む 130 を越える論文において引用されるに至っている。二次代謝物における機能アノテーションを可能にすることを目標に、科学文献調査により集めたデータをもとに KNApSAcK Family 代謝物活性データベース(Metabolite Activity DB)の開発を進めた。Metabolite Activity DB は、天然物-対象生物-生物活性の三組の関係からなるデータベースである。現在までに、9,584 件のこれらの三組の関係の登録を完了した。

##### 2.2 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(奈良先端大、理研)

約 500 化合物の MassBank ESI-MS/MS データを解析して、ピークに chemical annotation を行い、m/z 値を分子式に置き換えた Chemically accurate MS データを MassBank から提供した。このデータを利用して、ピークの分子式と部分化学構造式との化学的関係を解析した。この関係知識を利用して、未同定代謝物の化学構造推定をおこなう Metabolite Prediction Tool を開発して、MassBank に実装した。

#### 3. メタボローム統合 DB(かずさ DNA 研、理研)

研究者間のデータ共有を目指してメタボロームデータの標準化フォーマット(The Togo Metabolome Data Format, TogoMD)の開発を行った。TogoMD フォーマットを用いて、

Bio-MassBank (<http://bio.massbank.jp/>) へのデータ提供を開始した。データ処理パイプラインの改良により、Bio-MassBank からのスペクトルデータ公開は加速度的に進み、初年度は 1,300 スペクトル、2 年目は 3,242 スペクトル、3 年目終了までに 50,644 スペクトルを公開した。

フォーマットの統一については、米国の MSI コンソーシアム(2005 年頃～)、欧州 COSMOS コンソーシアム(2012 年～)等で検討が進められてきており、TogoMD フォーマットは、現在でも統一化されていないメタボロミクスデータ記述方法の世界標準となりつつある。今後、世界標準として TogoMD を利用するための基盤を構築すると共に、データ利用を促進すると期待される。

データ公開における最大の律速要因である実験手法の詳細情報(メタデータ)の管理を効率化し、データ公開を加速する目的で、メタデータの管理と共有を専門としたデータベースシステム Metabolonote の開発を行った。Metabolonote は、2012 年より欧州バイオインフォマティクス研究所(EBI)で発足した既存の公共メタボロームデータリポジトリ MetaboLights と相補的に機能し、メタボロミクス研究の推進に貢献している。

また、RefSpec サーバーに掲載する統合スペクトルは、各機関が公開する実測するスペクトルから統合ソフトウェアによって半自動的に作成する。そのため実測スペクトルを各所から定期的にクロールリングする必要が生じる。スペクトル情報は頻繁に変更されるものではないので、データの差分を利用すれば効率よくクロールリングできる。その基本となるソフトウェアを作成した。

このようにメタボローム実験データの標準的記載、生物サンプルにおける質量分析データベースとそのケミカルアノテーションならびに機能アノテーションまでの一連の作業を容易にするためのデータベースならびにツール群を開発した。

## (2) 研究開発成果のデータベース等

別紙に記載

## § 2. 研究開発構想(および構想計画に対する達成状況)

### (1) 当初の研究開発構想

「メタボローム・データベースの開発」プロジェクトでは、図 1 に示す構想にもとづいてプロジェクト間で密に連携をとり、以下の3項目について研究開発を進めた。

#### 1. 質量スペクトル DB(化合物 MSDB とメタボローム MSDB)の拡充

##### 1.1 化合物 MS DB の開発

化合物 MS DB は既知の代謝物質や化合物について測定した MS データを共有するための公共リポジトリとして機能する。MS データを共有するために、各レコードにクリエイティブ・コモンズ・ライセンスの表示を義務付ける。

化合物ごとに分類された MS データは、「MS データー化学構造の関係知識」を得るためにも使用する。日本質量分析学会公式の MS DB として論文投稿時に MS データを MassBank に投稿するためのシステム開発をおこない、日本脂質生化学会など関連他学会にもこのシステムを普及することによって、LipidBank など二次代謝物の MS データの収集を加速する。

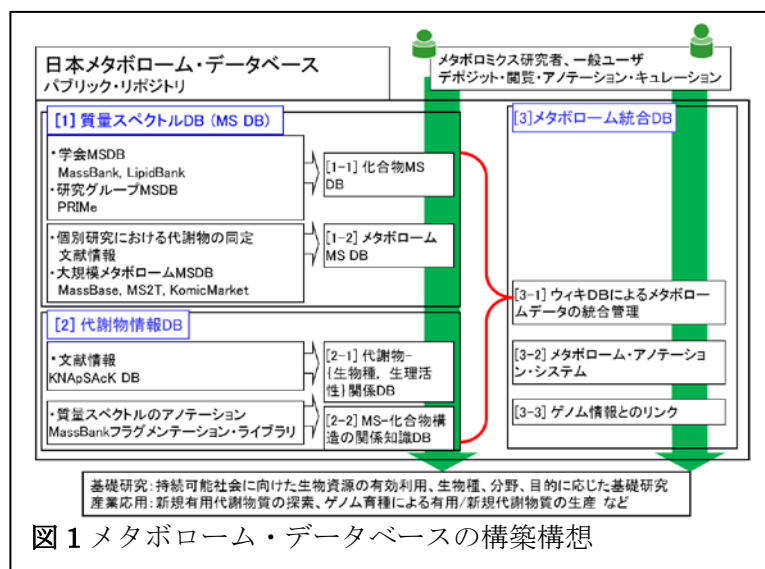


図 1 メタボローム・データベースの構築構想

## 1.2 メタボローム MS DB の開発

生物試料からの質量スペクトル・プロファイルの標準化法を開発し、メタボローム MS DB として公開する。MassBank、MassBase/KomicMarket などの既存データベースで開発された要素技術をもとに、化合物名が不明の場合でも質量スペクトル・ピークに対応したタグを付けるなどして、タグから生物種、部位、時期などの情報を取得することを目指す。これはゲノム解析が多数の機能未知遺伝子を発見したとき、それらの塩基配列をタグとして DB 化したことに通ずる発想である。すなわち、化学構造が未同定であっても、同じ MS データで関係づけられる代謝物が生合成されている生物種や組織、細胞などを比較・解析し、存在量の時間変動を遺伝子発現のそれと関係づけるインフォマティクス技術を開発・実装し、研究者に提供する。

生物試料を LC-MS や GC-MS で分析し、検出された代謝物の MS データを識別子として記録・収集するためには、これらの分析が試料中の代謝物質を網羅していること(網羅性)と、MS データが代謝物質の化学構造を表現していること(識別能が高い)の2つが必要条件となる。本開発課題では、各研究者がそれぞれの生物試料に最適化した分析法を用いて測定した各 MS データについて、網羅性と高識別能の実現法を測定者と協力して開発することによって、分析法に依存しない(ユニバーサルな)識別能の高い MS データを日本メタボローム・データベースとして公開することを目指す。そのために必要なデータベース設計やデータ形式の定義、検索ツールの開発も実施する。

### 2.代謝物質情報 DB の構築

#### 2.1 代謝物質と生物活性の関係データベース

メタボローム研究を医・薬・農学に関わる基礎と応用研究へ発展させる目的で、文献情報をもとに、現在 KNApSAcK に登録されている 1 割にあたる約 5 千の代謝物について生理活性情報を文献より抽出しデータベースに格納し、生理活性のオントロジーなども考慮した記述の標準化を図る。また、LipidBank データベースに格納される生理活性情報とも統合する。

#### 2.2 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

MS 機器の高精度化にもかかわらず、ESI-MS<sup>2</sup> データから二次代謝物質あるいはその化学構造式を推定する実用的なインフォマティクス技術は未だ開発されていない。

本開発研究では MassBank で公開されている ESI-MS<sup>2</sup> データを解析することによって得られる経験的な「MS データと化学構造の関係知識」に基づいて、代謝物質の化学構造式を推定する。すでに MassBank では、各イオンの分子式を同定し、解裂した化学結合を特定すること(「MS データの化学情報化」)によって、イオン(ピーク)と部分化学構造式との関係を解析する。これをもとに各ピークに関係づけられている各部分化学構造式を確率的に表現し「MS データと化学構造の関係知識 DB」を構築する。この DB と KNApSAcK データベースを用いて、MS データから化学構造式を推定し、部分化学構造の組合せを満たす代謝物質を未同定代謝物質の候補化合物として提示するツールを開発する。

### 3.メタボローム統合 DB

上記のデータベースコンテンツや知識は、それぞれ単一機関で入力・集約されるものではなく、研究者コミュニティ及びデータ生産者が協働して作り上げるものである。そのため、ウィキ形式でコンテンツを公開したまま作成できるソフトウェア基盤を構築する。具体的には、画像・テキスト・代謝物の物理化学情報についてウィキ化が進んでいる Metabolomics.JP データベースを発展させ、クロマトグラムや映像のようなファイルサイズの大きな情報、質量スペクトルのようにユーザーがインタラクティブに内容を閲覧・比較したい情報も扱えるシステムの構築を目指す。初年度に設計する統合システムの内容を順次ウィキ化することにより、本計画終了後に特定機関に多大な負担をかけずにメンテナンスのコストを下げることを目標とする。

また、最終的にはこれらの研究を進めることによりメタボロミクス研究における統合データベースの世界拠点の形成をめざす。

#### (2) 新たに追加・修正など変更した研究開発構想

##### (1) 東日本大震災(H23.3.11)による停電への対応

東日本大震災(H23.3.11)による停電を契機として、MassBank の特徴である分散型データベ

ースの欠点が顕在化した。すなわち、国内に分散していたデータサーバのどれかがサービスを停止する状態が頻発した。そこで、(当初国内に設置予定であった)ミラーサーバを国外(EU)に設置することにした。すなわち、H25.1 から MassBank.jp と MassBanmk.eu の2つのサーバで運営し、いずれのサーバにデータが update されても他方にただちに反映するように SVN サーバを利用して、双方で同一のデータ backup を所有することにした。この副産物として、ユーザーに提供していたデータのダウンロードサービスもいつも最新のコピーを提供できるようになった。

### (2) MassBank Developer's Kit の提供

MassBank が国際的に普及して実質的に世界標準と認められるとともに、MassBank システムを研究分野に応じて、カスタマイズしたいという希望が国内外の研究機関や企業から多く寄せられている。これに応えるために、全てのソースコードを Java だけで記述した MassBank Developer's Kit (MDK)を開発した。MDK を利用した Bio-MassBank と併せて SourceForge の Git から公開した(<https://sourceforge.net/p/massbank/MDK/ci/master/tree/>および <https://sourceforge.net/p/massbank/Bio-MassBank/ci/master/tree/>; H26.2)。

### (3) Java Code signing 証明書の付与

Java に深刻なセキュリティーホールがあることがわかった(H25 はじめ)。Java の開発元である Oracle は対策として、2つの認証システム、1.サーバのサイトを証明する認証、2.プログラムの Java ソースコードを証明する認証、を導入することにした(H25.9)。Java 7 Update 51 からは証明書がないと利用できなくなった(H26.1)。国内外の企業から証明書がない MassBank システムを利用することはできない、との指摘があり、MassBank はサーバに認証を必要とするだけでなく、分散型データベースであるので国内外の他の研究機関でも MassBank サーバを設置している。そこで code signing をつけた Java applet を作成して、提供した。

### (3) 達成状況

研究項目	H23 年度	H24 年度	H25 年度	変更点
1. 質量スペクトル DB(化合物 MSDB とメタボローム MSDB)の拡充				
1.1 化合物 MS DB の開発 (奈良先端大・理研・かずさDNA研グループ)				
・プロトタイプシステムの開発	←	←	←	・予定通り
・データの DB への蓄積ならびに公開		←	←	・予定通り
1.2 メタボローム MS DB の開発 (奈良先端大・理研・かずさ DNA 研グループ)				
・DB プロトタイプシステムの開発	←	←	←	・予定通り
・データの DB への蓄積ならびに公開		←	←	・予定通り
2.代謝物質情報 DB の構築				
2.1 代謝物質と生理活性の関係データベースの開発(奈良先端大グループ)				
・データベースの構築	←	←	←	・予定通り
・データベース蓄積ならびに公開		←	←	・予定通り
2.2 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				
・システム開発	←	←	←	・予定通り
・データの充実ならびに公開		←	←	・予定通り

3.メタボローム統合 DB(理研グループ) ・wikiを中心としたデータ統合技術開発 ・wiki を中心とした統合データの公開	←-----→			・予定通り
	←-----→			・予定通り

#### (4) 研究開発の今後の展開について

MassBankに収集されているデータそのものが科学技術イノベーションに貢献することはないので、MassBank がどのような研究分野で利用されているのかを調べた。

MassBank の論文(Horai et al., J. Mass Spectrom., 2010)を引用している論文(154 件)を掲載した学術国際誌の研究分野を Web of Science で調べると、薬学、医化学、植物科学、農学、環境学、表面化学など多岐にわたる。なかでも医療、医薬開発をはじめとする生命科学研究が最も多い。これは質量分析の主要な利用研究分野と重なり、MassBank は名実ともに質量分析を行うために必須の参照データベースの国際標準であり、(著者の所属機関から)基礎研究のみならず企業における応用研究においても広く利用されていることがわかる。

### § 3 研究開発実施体制

#### (1) 研究チームの体制について

① 「研究代表者:金谷重彦」奈良先端科学技術大学院大学グループ  
研究参加者

氏名	所属	役職	研究開発項目	参加時期
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	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4~H26.3
西岡 孝明	奈良先端科学技術 大学院大学情報科 学研究科	特任教授	MassBank, Bio-MassBank の開発	H23.4~H26.3
二瓶 義人	同上	本事業専任 研究員	システム開発	H23.4~H26.3
池田 奨	同上	本事業専任 研究員	システム開発	H23.4~H25.3
尾畷 雄也	同上	本事業専任 研究員	MS データと化 学構造の関係 知識	H23.4~H24.3
小笠原直毅	奈良先端科学技術 大学院大学バイオ サイエンス研究科	教授	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4~H26.3
橋本 隆	奈良先端科学技術 大学院大学バイオ サイエンス研究科	教授	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4~H26.3
中村 健介	奈良先端科学技術 大学院大学情報科 学研究科	研究員	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4~H26.3
平井 晶	奈良先端科学技術 大学院大学情報科 学研究科	研究員	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4~H24.3

和田 眞昌	奈良先端科学技術 大学院大学情報科 学研究科	D3	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4～H24.3
庄條 昌之	奈良先端科学技術 大学院大学情報科 学研究科	D3	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4～H24.3
中村由紀子	奈良先端科学技術 大学院大学情報科 学研究科	研究員	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4～H26.3
桂樹 哲雄	奈良先端科学技術 大学院大学情報科 学研究科	D2	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4～H26.3
秋葉 正毅	奈良先端科学技術 大学院大学情報科 学研究科	M1	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4～H24.3
池田 俊	奈良先端科学技術 大学院大学情報科 学研究科	D3	代謝物-活性、 代謝物-生物種 関係 DB の開発	H23.4～H26.3

② 「研究分担者:櫻井 望」かずさ DNA 研究所グループ

研究参加者

氏名	所属	役職	研究開発項目	参加時期
櫻井 望	かずさ DNA 研究所産 業基盤開発研究部	研究員	総括、システム 設計、開発	H23.4～H26.3
荒 武	同上	プロジェクト研究 員	システム設計、 データ構築	H23.4～H26.3
榎本 光男	同上	派遣技術員	プログラム開発	H23.11～H26.3
柴田 大輔	同上	産業基盤開発研 究部長	研究開発に関 する助言	H23.4～H26.3

③ 「研究分担者:有田正規」理研グループ

研究参加者

氏名	所属	役職	研究開発項目	参加時期
有田正規	理研横浜研究所 環境 資源科学研究センター	チームリーダー	システム設計、 プログラム開発	H23.4～H26.3
Ramon Mejia	理研横浜研究所 環境 資源科学研究センター	特別研究員	プログラム開発	H24.10～ H26.3
津川裕司	理研横浜研究所 環境 資源科学研究センター	特別研究員	プログラム開発	H24.10～ H26.3
福島敦史	理研横浜研究所 環境 資源科学研究センター	研究員	システム作成助 言	H23.4～H26.3
草野都	理研横浜研究所 環境 資源科学研究センター	研究員	システム作成助 言	H23.4～H26.3

(2) 国内外の研究者や産業界等との連携によるネットワーク形成の状況について

(2-1) 国内学会との連携

日本質量分析学会とは公式データベース(H20.3 から)として協力を得ており、MassBank の普及、向上を担うスペクトルデータ部会(H24.3 から)からのサポートを得ている。アメリカ質量分析学

会 (AMSM) でも Metabolomics interest group のワークショップを開催した (H24.6)。

## (2-2) 企業との連携

エーザイ(医薬企業)と田中耕一島津最先端研究所(質量分析機器製造業)とは H20 ごろから MassBank を利用したマススペクトルのデータベース化を推進している。とくに、「質量分析機器から MassBank レコード形式にしたマススペクトルを出力して MassBank に deposit して企業内でデータベースを構築したい」、「質量分析機器から直接 MassBank を検索して化合物同定ができるようにしたい」という強い要望が国内外の多くのユーザーから長年にわたってリクエストされている。それを実現すべく、ツール Mass++ の開発に連携をしてきた。Mass++ は MassBank を利用したメタボローム解析と、タンパク質データベースを利用したメタボローム解析のいずれにも利用できる。H26.1 に 最 新 版 Mass++ v2.6.4 が リ リ ー ス さ れ た (<http://www.first-ms3d.jp/english/achievement/software>)。

## (2-3) 国際的な連携

・NORMAN network (<http://www.norman-network.net/>) は EU19 ヶ国が参加する公立環境科学研究機関連合である。この NORMAN network から「環境物質の LC-MS<sup>2</sup> マススペクトルのデータベースを MassBank と連携して構築したい」との申し出があった (H24.5)。そこで、アムステルダムで開催された NORMAN 学会 (H24.11) の機会に、MassBank のレコード作成、データベース構築、検索、利用についての Workshop を共催した。NORMAN はその後、MassBank.jp と同等の機能をもつ MassBank.eu (相互にミラーとなる) サーバを立ち上げた (H25.1)。また、Mass++ (2-2) に相当するツールである RMassBank を独自開発して、会員や一般のユーザーに提供している (<http://bioconductor.org/packages/2.11/bioc/html/RMassBank.html>)。さらに MassBank Developer's Kit を利用して、環境科学 MassBank を立ち上げる計画がある。

・CASMI2013 の共催: マススペクトルから化学構造式を推定する国際コンテストを本 NBDC 統合化推進プログラム (金谷代表) と日本質量分析学会との共催として開催した (<http://www.casmi-contest.org/2013/index.shtml>)。これは CASMI の第2回コンテストであり、EU、米国、英国、日本など MassBank に関連あるメタボローム研究者による CASMI consortium として運営している。

・Thomson-Reuters 社との連携: Thomson-Reuters 社は Web of Science や Citation Index で知られる。Public repository の各レコードの学術論文での引用回数を調べる Data Citation Index (DCI) に協力している (H25 秋から)。論文中では「MassBank のレコード IDxxxxxxx を利用した」としか書かれない傾向がある。このために MassBank は参考文献セクションで引用されないことが多いので、正しく Citation Index に反映していない。これに対して、DCI ではこれらも正しく引用回数として記録され、レコードの提供者についても DCI に反映される。

## (2-4) NBDC プロジェクトならびに国内における連携

メタボローム実験におけるメタデータ一元管理システム Metabolonote (<http://metabolonote.kazusa.or.jp>) を通し、外部データベースとの連携を進めた。現在、28 生物種、375 分析を含む 764 の生物測定データに関するメタデータを公開しており、Bio-MassBank、MassBase、KomicMarket、RIKEN PRIME、PGDBj (統合化推進事業、田畑哲之代表) 等 7 つのデータベースと連携している。

生物試料全体における質量分析スペクトル・データ MassBase、Bio-MassBank の開発および、生物種・代謝物および生理活性・代謝物の関係データベース KNApSAcK Core データベースならびに Metabolite Activity データベースの開発を通して「ゲノム情報に基づく植物データベースの統合プロジェクト (田畑哲之代表)」と強く連携をとり、RDF 化に向けた基盤技術の開発を進めた。また、KNApSAcK Core データベースと「生命と環境のフェノーム統合データベース (豊田哲郎代表)」とも連携をとり研究を進めた。NBDC と連携して開発が進められている医薬基盤研究所における創薬・疾患のための横断検索サービス Sagace において、KNApSAcK において RDFa lite による html のマークアップを行い、Sagace からの KNApSAcK Core への検索を可能にし、連携



を強めた。

## § 4 研究実施内容及び成果

4.1 研究課題名:質量スペクトル DB(化合物 MSDB とメタボローム MSDB)の拡充(奈良先端大・理研・かずさ DNA 研究所グループ)

研究開発実施内容及び成果

4.1.1 化合物マスマススペクトルデータベース MassBank の開発

(1) Batch 検索機能の開発

大量マスマススペクトルをクエリとする batch 検索機能の開発: メタボローム研究では生物試料について LC-MS, MS<sup>2</sup> 分析などの測定を行う。生物試料1件あたり数百から数千の代謝物が検出される。これら多数のマスマススペクトルを手作業で参照マスマススペクトルと照合して代謝物を同定するスペクトル検索はとても時間と労力を要する作業であった。そこで MassBank を一括してスペクトル検索を行うための SOAP-API インターフェイスを開発提供した。ここで REST-API についても検討したが、http データには文字制限があるために、ピークデータが多く出現しているマスマススペクトルを入出力することは困難であった。そのために SOAP-API を採用した。前述した Mass++を利用して、あるいは Python などから MassBank API を利用して batch 検索が行うことが可能となった。

(2) MassBank データにクリエイティブ・コモンズ(CC)・ライセンスを表示する

・MassBank では Windows 版 MassBank システムの配布をおこなっていたので、企業内の LAN 環境で MassBank を構築、利用することができるために好評であった。それとともに、「Internet で公開している MassBank データのコピーを提供してほしい」という要望が国内外から寄せられていた。

・各レコードにクリエイティブ・コモンズ(CC)・ライセンスの表示を義務付けた。これまでデータの著作権は公開した研究グループが所有していたので、データのコピーを配布するために公開した研究グループの承諾を取得した。これは大変に手間であった。全研究グループに CC ライセンスの主旨を説明し、承諾を得る努力をした。しかし、一部の化学系や企業の研究グループは営利目的でのデータコピーの利用が認められず残念であった。

・CC ライセンス表示によって、MassBank データのコピーを自由に配布することが可能になっただけでなく、MassBank 側で品質管理のためにレコードの修正をおこなうための手続きが簡素になった。このことはデータの品質向上につながっている。

(3)分散型データベースの欠点の克服

分散型データベースの欠点を克服することは、MassBank データのコピーのダウンロードにつながるのので、この開発を優先して実施した。MassBank ではこれまで分散型 DB を採用してきた。このメリットは公開する研究者が MassBank レコード形式で作成し、これらのレコードをデータサーバ上で管理する(レコードの追加、修正、削除)ので、データサーバの維持とレコード管理にかかる経費を研究者が負担している。データを一カ所に集めて、管理する通常の DB と比べると、MassBank ではレコード作成をはじめ DB の維持、管理にかかる経費を研究者が分担していることになる。さらに分散並列処理を実現して検索速度を向上している。このような利点の一方で分散型 DB

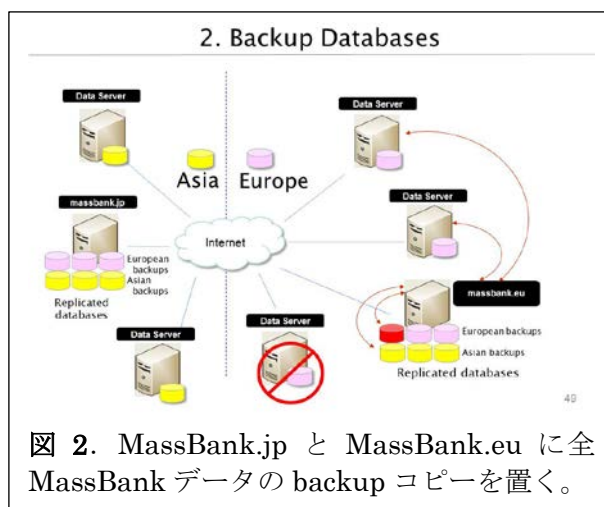


図 2. MassBank.jp と MassBank.eu に全 MassBank データの backup コピーを置く。

さらに分散並列処理を実現して検索速度を向上している。このような利点の一方で分散型 DB

には深刻な欠点もある。現在、MassBank には計 11 個の分散データサーバがあるが、これらのどれか1つ(あるいはそれ以上)のサーバにおいてサービスができない状態にある場合に、全てのサーバがサービスしている場合とは異なった検索結果が得られる。これがユーザーにとって深刻な問題となっていた。以上の欠点を解決する方法として MassBank.jp と対等な機能を有する MassBank.eu を EU に設置した。さらに分散データサーバ上にある全データの最新版コピーを backup として2つの MassBank に置いておくシステムを開発した(図2)。どれかのデータサーバが利用できない状態にあるときには、代わりに backup コピーを検索して結果を返す。backup コピーは文書管理システムである Subversion(svn)を利用して、svn サーバは奈良先端大に設置され、全ての MassBank データは svn サーバのレポジトリに蓄えられる。このデータは backup データを提供するとともに、ユーザーが download して自由に利用すること(データコピーの提供)を可能にした。

#### (4) MassBank の将来展望

分散データサーバ上にある全データの最新版コピーを backup として2つのサーバ、MassBank.jp と MassBank.eu に置いておくことは、MassBank の当初のコンセプトである「分散型データベース」に矛盾する。このコンセプトの背景となっているのは MassBank 維持経費を提供者が分散して負担しようとするものである。このうち最も大きな経費は MassBank レコードの作成にかかる費用である。Mass++ や RMassBank などのサードパーティによって開発、提供されるようになって、データ提供者の負担も軽くなりつつある。分散型データベースを廃止することによって将来のメンテナンスが格段に容易になると期待される。そこで、MassBank.jp と MassBank.eu を除いてデータサーバを廃止した(H25.12)。また、Java だけで記述され、冗長を取り除いたソースコード、MassBank Developer's kit を提供して、ユーザーそれぞれの環境や目的にカスタマイズした MassBank システムを構築することによって MassBank レコードはさらに増加し利用されるであろう。

#### (5) 学会や他の研究グループとの連携

##### 日本質量分析学会と連携した活動

MassBank は日本質量分析学会の公式データベースである(H20.3から)。MassBank の普及とデータベース化、データの品質向上を事業の目的とするスペクトルデータ部会が設置された(H24.3)。本部会の最初の事業として、H23.7 に学会員 8 名、MassBank4 名、奈良先端大グループ 3 名が集まり、データの化学構造式の誤り修正や分子関連イオンが正しく観測されているかどうかなどデータの品質をチェックした。その後、データを各自持ち帰り、H24.1 までに約 1,000 化合物のデータの品質監修を実施し、合計 568 レコードの修正をおこなった。

また国際学会誌に論文原稿が受理されると、論文中のマスマスペクトルを MassBank に登録するための投稿の手引きを作成した。しかし、「MassBank レコード形式への変換、MassBank への登録を著者自身がおこなうことは著者への負担が大きい」と学会誌編集委員会から反対意見が多かった。そのため MassBank のほうでレコード形式への変換、登録を H26 年度から著者に代わって本プロジェクトで実施することになった。

##### ユーザイと田中耕一最先端研究開発支援プログラムとの連携

LC-MS<sup>2</sup> データは各機器メーカーに特有のデータ管理システム上にバイナリ形式で保存されている。MassBank などの外部データベースを検索するためには text 形式でピークデータを手動で出力して、text ファイルを MassBank に読み込んで検索するしか方法は無かった。これは効率が悪く、誤りも多かった。Mass++ プロジェクトと連携して、バイナリ形式やプロテオミクス用の共通データ形式である mzXML で MS<sup>2</sup> データを入力し、SOAP API を利用して MassBank の batch 検索を実現した。約 1 千スペクトルをクエリとした場合に 30-40 分でスペクトル検索の結果がえられる。このような検索機能は MassBank 独自のもので、米国のスペクトルデータベースの開発を進める NIST データベースにも未だ無い。

## 国際連携

NORMAN 連合から MassBank と連携したい旨の申し出 (H24.5) があった。EU 加盟国の環境科学研究機関のネットワークであり、水や土壌中の環境汚染物質の標準マススペクトルを MassBank で公開し、共有することを希望してきた。環境汚染物質はヒトやその他の哺乳動物に生理活性を示すものの、それらの ESI-MS<sup>2</sup> データを収集したデータベースは小規模なものを除いてこれまでなかった。MassBank は生体内小分子のマススペクトルを対象としていたが、動物組織のメタボローム解析をおこなうと食物や飲料水に残留する環境汚染物質が検出されるものの、MassBank にはそれらのマススペクトルが収集されていないので、これまで「不純物」とされ、化合物同定はできなかった。このような状況の中、「医薬や環境汚染物質などの生理活性物質のマススペクトルを積極的に公開してほしい」という要望が国内外の研究者からあったので、NORMAN からの申し出を受け入れた。そこで EMBL-EBI MIRIAM に MassBank が登録された (<http://www.ebi.ac.uk/miriam/main/collections/MIR:00000273>, H24. 5)。

### 4.1.2 メタボロームマススペクトルデータベース Bio-MassBank の開発

#### (1) Bio-MassBank のコンセプト

Bio-MassBank では、生物試料をメタボローム解析したマススペクトルの public repository であり、代謝物が同定の有無にかかわらず全て公開することを目的とする。同一の試料から得られたデータは1つのデータセットとする。マススペクトルを代謝物の tag として利用して、マススペクトルが互いに同じ、類似しているものは同じ代謝物とみなすことによって、これまで未同定代謝物として一括されていた代謝物を、個々に比較することを可能にする。これに対して、EMBL-EBI で開発している public repository である MetaboLights ではマススペクトルの類似性を比較する検索機能がないので、代謝物を同定したり、未同定代謝物の比較をすることができない。これらの点を考慮して、生物試料におけるメタボローム・マススペクトル・データベース Bio-MassBank を設計し、データを蓄積した。

#### (2) レコード形式の決定

かずさ DNA 研究所グループとメタデータ項目のフォーマットや記述項目を決定した。マススペクトルは MassBank レコード形式で記述することにした。

#### (3) MassBank Developer's Kit の開発

従来の MassBank は Java と Perl を中心に C++ などの言語でプログラムを作成していた。開発の途中で新しい機能を追加したので、システム全体としては構成がツギハギであり、極めて複雑で、重複も多いソースコードであった。このままではシステムの改良や普及を妨げることから、全てを Java で書き換え、ソースコードの冗長部分を除いて、すっきりとした MassBank Developer's Kit を新たに開発し、SourceForge の Git から公開した (<https://sourceforge.net/p/massbank/MDK/ci/master/tree/>, H26.2)。MassBank Developer's Kit の特徴は、1.Linux や Windows をはじめどのような OS 上でも利用できる。2.実行速度が格段に向上した。これはソースコードの冗長性がなくなりシンプルになった

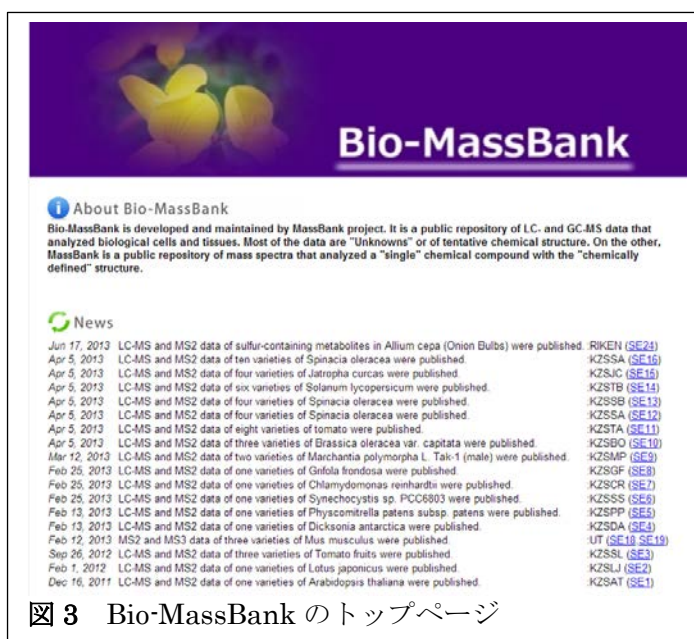


図3 Bio-MassBank のトップページ

ことや異なる言語の混在が無くなった、通信のオーバーヘッドが無くなったなどの理由による。3.レコードの登録処理の時間が格段に短くなった、などである。近い将来、この開発キットを利用して多くの MassBank システムの仲間が増えることを期待している。NORMAN 連合では、このキットを利用して、環境科学 MassBank を開発中である。

この開発キットを利用した最初の応用システムが Bio-MassBank であり、ソースコードも公開した (<https://sourceforge.net/p/massbank/Bio-MassBank/ci/master/tree/>, H26.2)。

#### (4) データの公開

これまでに、かずさ DNA 研究所が植物 10 種 (29 栽培種)、微生物 3 種、哺乳動物 1 種を LC-MS, MS<sup>2</sup> 分析したデータ 51 分析に由来する 53,236 件のマススペクトルを Bio-MassBank より公開した (<http://bio.massbank.jp/>) (H23.12.16)。これによって、データセット間で同じ、類似したマススペクトルを手がかりとして共通な未知代謝物を容易に検索することができるようになった (図3)。

#### 4.1.3 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

##### (1) ピークに分子式を付与 (chemical annotation) するアルゴリズムの開発

これまで 5 年間をかけて ESI-MS<sup>2</sup> データで観察されたピークの m/z 値から、手計算で分子式を推定して chemical annotation をおこなってきた。しかし、なかなか作業がはかどらなかったので、かずさ DNA 研グループと連携して自動に分子式を付与するアルゴリズムを開発した。これにより 1 つの化合物を分析した 2 つ以上の異なる ESI-MS<sup>2</sup> データに観察された共通のイオンについて、ピークの m/z 値に偶数個の電子を含む分子式だけを選んで付与することが可能となった。

化合物 (重複のないユニークな 2,121 化合物) を高精度に測定した MassBank ESI-MS<sup>2</sup> データ (9,602 レコード) を対象として、正イオンピークに 7,469 分子式を、負イオンピークに 3,437 分子式を付与した。これらの分子式はこれまでに手作業でおこなってきた chemical annotation の拡張に寄与した。

##### (2) 数値 (m/z) データから分子式データへ

これまでマススペクトルはピーク (m/z) を数値で表現してきた。この数値の精度を測定精度 (accuracy) と呼ぶ。質量分析計はピーク (m/z) を高精度に測定する高分解能化 (測定精度 1ppm 以下; 有効数字 6 桁) へと向かっており、メタボローム解析研究への導入がはじまっている。MassBank で現在公開されている ESI-MS<sup>2</sup> データは測定精度が数 ppm から 30ppm のものが多い。すなわち MassBank のデータは数年後には参照データとして役立たなくなることが予想される。かといって、全ての化合物を高分解能な分析計で測定しなおすことは MassBank の resource と時間を無駄にすることになる。そこで、m/z を数値ではなく、chemical annotation (1) で得られた分子式で表現した MassBank レコード (Chemically Accurate Data) を提供することにした。分子式は有効数字の制約を受けないので、将来にわたって参照スペクトルとして利用することができる。

また、分子式データを利用してスペクトル検索をおこなうツールの開発も終了し、現在、論文を準備しているところである。論文が受理され次第 Chemically Accurate Data と検索ツールを公開する予定である。

##### (3) MS データと化学構造の関係知識 DB の構築

ESI-MS<sup>2</sup> データ (マススペクトル) で観察されたピークについて、そのイオンが由来する部分化学構造との関係 (「ピークと部分化学構造との関係」) を解析、収集した。ESI-MS<sup>2</sup> データ (正イオン 788 データ、負イオン 676 データ) について、ピークと部分化学構造式との関係を図示した wiki ベースの「MassBank Fragmentation Library」への入力を継続した (<http://metabolomics.jp/wiki/Index:MassBank>)。この入力を容易におこなうツールである Ojimatix editor も作成した。

並行して「MS データと化学構造の関係知識 DB」の研究開発を進めた。一般に与えられた分



子式に対応する部分化学構造は複数ありうるが、ピークと部分化学構造との関係のうち1:1の対応関係にあるものを精選したところ、269 分子式と 94 部分化学構造との関係(630 ペア)が得られた(1つの部分化学構造から複数のイオンに開裂するので、分子式の数の部分化学構造の数よりも多くなっている)。この経験的關係は Excel 上に整理した。

#### (4) マススペクトルから代謝物の推定ツールの実用化

開発した化学構造式ツールは、先ず ESI-MS で観測された分子イオンピーク(精密質量)に該当する代謝物を KNApSAcK データベースから選び、次いで、ESI-MS<sup>2</sup>で観察されたピークについて「ピークと部分化学構造との関係」から推定される部分化学構造を有する代謝物を絞り込むことにより、代謝物を推定しようとするものである。このツールは MassBank に「Metabolite Identification」として実装した。「ピークと部分化学構造との関係」の数はいまのところ限られるが、このツールによる質量スペクトルからの代謝物推定に関する実証的サービスを提供することができた(図4)。

#### 4.2 研究課題名:メタボローム・データベースの構築(かずさDNA研究所 産業基盤開発研究部) 研究開発実施内容及び成果

本課題では、メタボロームデータの公開・流通・利用を促進することを目的として、(1)データ標準フォーマットの開発、(2)実験手法の情報(メタデータ)を軸としたデータ共有システムの開発、および(3)データの大量公開を進めた。メタボロミクスでは、多様な分析装置が色々な組み合わせで使用されるため、出力される様々なデータを統一的に記述できるフォーマットの整備が遅れている。結果として、独自のフォーマットで実装された多様なデータベースが散在する状況となった。そこで本課題では、国内の研究者が活用しやすい標準化フォーマットを整備し、そのフォーマットに従ってメタデータを介してデータベース間あるいは研究者間でデータを共有できるシステム構築を目指した。一方で、これまでかずさDNA研究所で取得した膨大な解析データについてデータの公開を進めるとともに、データ公開を加速する技術開発を行った。具体的な公開目標としては、課題開始時に公開可能であった36,000件の質量分析データの一般公開を目指した。

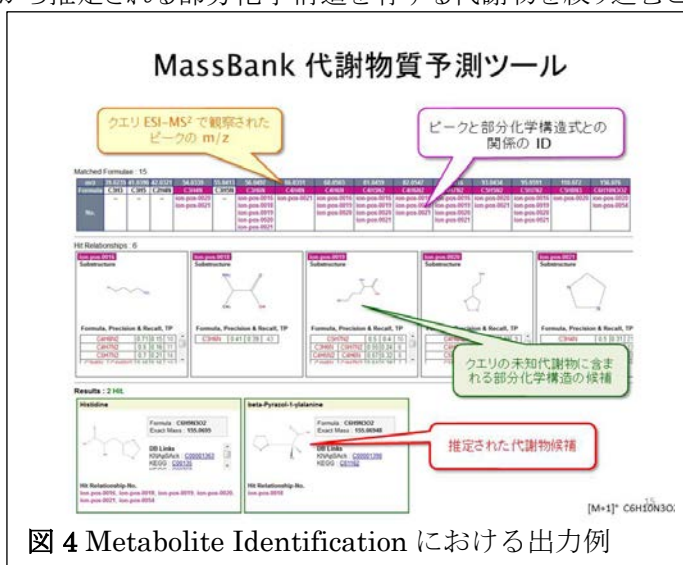


図4 Metabolite Identification における出力例

#### (1)メタボロームデータ標準フォーマット(TogoMD)の開発

国内の研究者が活用しやすいメタボロームの標準化フォーマット(The Togo Metabolome Data Format, TogoMD, <http://metabolonote.kazusa.or.jp/TogoMD>)の開発を行った。メタボロミクスでは、同一の生体サンプルが異なる分析装置で複数回測定される場合などがあり、実験手法の詳細情報(メタデータ)は階層構造で記述できる。また、それぞれの実験データからメタデータを参照するためには、一意の ID で任意の階層のメタデータにアクセスすることが望ましい。メタデータの階層構造を管理でき、任意の階層のメタデータを一意の ID で参照できるようなフォーマットの開発が世界においてもなされておらず、独自に開発することとした。かずさDNA研究所と奈良先端大チームが中心となり、化合物 MS DB の開発において MassBank で多く扱われている液体クロマトグラフィー(LC)質量分析(MS)のデータを主な対象として、ガスクロマトグラフィー、キャピラリー電気泳動(CE)-MS 等にも対応できるものを開発した。初年度後半から、TogoMD フォーマットを用いて、Bio-MassBank (<http://bio.massbank.jp/>)へのデータ提供を開始した。データ処理パイプラインの改良により、Bio-MassBank からのスペクトルデータ公開は加速度的に進み、

初年度は 1,300 スペクトル、2 年目は 3,242 スペクトル、3 年目終了までに 50,644 スペクトルを公開した。

TogoMD フォーマットは、現在でも統一化されていないメタボロミクスデータ記述方法の世界標準となり得る。フォーマットの統一については、米国の MSI コンソーシアム(2005 年頃～)、欧州 COSMOS コンソーシアム(2012 年～)等で検討が進められてきたが、現在まで世界標準となりうるものは提案されていない。また、米国保健省 NIH は独自のフォーマットでヒトメタボロームデータを公開し始めている(2013 年～)。このような状況下では、いかに利用者のニーズに合ったデータを大量に提供するかが、標準化の鍵となる。TogoMD で大量のデータを公開し、まずは国内で流通を図れたことは、世界標準として TogoMD を利用するための基盤を構築すると共に、データ利用の促進への寄与できると期待される。

## (2)メタデータを軸としたデータ共有システム(Metabolonote)の開発

データ公開における最大の律速要因である実験手法の詳細情報(メタデータ)の管理を効率化し、データ公開を加速する目的で、2 年目から、メタデータの管理と共有を専門としたデータベースシステムの開発を行った。メタデータの管理がデータ公開の律速となることは当初より想定されていたが、標準化フォーマットが完成しデータ処理パイプラインが効率化されたことで、これがより顕在化した。メタボロミクスのデータは、同一のサンプルが様々な機器の組み合わせで分析されること、分析実験が次々に追加されること、解析データの中から化合物を同定・推定(アノテーション)する作業が時間をかけて徐々に行われることなど、研究の進捗に沿って徐々に変化してゆくことが特徴である。このような変化に対応するためには、キュレーターが存在するレポジトリではなく、ユーザー自身が即時編集できるシステムが必要である。また、同一の分析から生じたデータが、生データのレポジトリ、マスマグメントのライブラリなど、用途が異なる複数のデータベースで使用されることがある。複雑なメタデータをユーザーに理解しやすく提供するためには、メタデータを一元管理し、それを外部から参照できるシステムが望ましい。そこで、このような特徴を兼ね備えるシステム Metabolonote(<http://metabolonote.kazusa.or.jp>)を開発した。Metabolonote は SemanticMediaWiki をベースに開発されているため、メタデータを即時編集し、公開することができる。また、実験データを登録したデータベースや論文サイトへのリンク、サンプルの画像等の補足情報などを任意に追加することができる。このような自由度により、登録者のメタデータがハブとなって、散在しているデータリソースが統合化されるため、閲覧者によるメタボロームデータの利用が促進され、また、メタボローム実験の体系的整理においても期待できる。3 年目は、API によるデータ取得機能の強化などを行い、外部データベースとの連携を進めた。現在、28 生物種、375 分析を含む 764 の解析データに関するメタデータを公開しており、Bio-MassBank、MassBase、KomicMarket、RIKEN PRIME、PGDBj(統合化推進事業、田畑グループ)等 7 つのデータベースと連携した。

Metabolonote は既存の公共メタボロームデータリポジトリ MetaboLights と相補的に機能し、メタボロミクス研究の推進に貢献するだろう。2012 年より欧州バイオインフォマティクス研究所(EBI)で発足した MetaboLights は、ISA-Tab と呼ばれる形式で、メタデータおよび実験データの URI を記述することができる。ISA-Tab は、Metabolonote のような研究目的/サンプル/機器分析/データ解析という粒度で階層構造を表現するには適しておらず、また、各メタデータ毎の ID でアクセスする仕組みも持たない。また、メタデータを配布して利用することを前提としており、一元管理をすることは想定されていない。一方で、キュレーターを介した登録により、しっかりしたメタデータを記載してゆくことには注力されている。恒久的なデータ保管場所としての MetaboLights と、データ生産に即した Metabolonote は、互いに相補的な役割を果たして、メタボロミクス研究に貢献するにちがいない。

## (3)データの公開

本課題では、これまでの公開総数(約 2.4 万件)を上回る 36,000 件の質量分析データの公開を当初目標とし、データ処理パイプラインの改良などを行った結果、これを十分に達成することが出来た。かずさ DNA 研究所が取得し非公開であったデータ約 6 万件のうち、権利がフリーであり、

即時、公開が可能な 36,000 件について、研究者が利用可能な形で公開することを目指した。公開の形態としては、分析機器に依存するバイナリ生データと、それをテキスト形式に変換し情報処理研究者が利用できるようにしたものを、これまで我々が構築してきた MassBase (<http://webs2.kazusa.or.jp/massbase>)に登録することとした。データ処理パイプラインの改良により、公開は 2 年目で劇的に進み、1 年目に 5,040 件、2 年目に 33,294 件を新規に公開して、当初目標を達成した。3 年目には、その後解析されたデータを少数追加した。最終的に、132 生物種を含む合計 38,344 件のデータを、本課題を通じて公開した。生データに近いデータの公開だけでなく、そこからピークを検出し、データベース検索などの一次アノテーションを行った代謝物プロフィールデータは、KomicMarket (<http://webs2.kazusa.or.jp/komicmarket>)から公開した。本課題を通じ、17 生物種、200 分析のプロファイルデータを公開することができた。Bio-MassBankへ登録したデータについては、すべて TogoMD フォーマットでプロフィールデータを入手できるよう、KomicMarket 一時サイト ([http://webs2.kazusa.or.jp/new\\_km\\_tmp](http://webs2.kazusa.or.jp/new_km_tmp))から公開している。このように大量のプロファイルデータを公開できたのは、高速な組成式演算や化合物データベース検索を可能とする MFSearcher の開発 (Sakurai et al., 2013, *Bioinformatics* 誌)と、LC-MS 解析ソフトウェアの PowerGet において TogoMD に従ったデータ出力機能の開発 (Sakurai et al., *BioMed Research International* 誌 in press)、また、アノテーションオントロジーの内部的な整備が大きな要因となっている。

本課題を通して行われた大量のデータ公開は、メタボロミクス研究にとって大きな貢献となっている。2012 年 2 月から稼働している MetaboLights には、NMR のデータを含め、現在 31 件の研究課題 (Study) が登録されている。一方 2012 年 12 月から公開された Metabolonote には 34 の研究課題が登録されており、MetaboLights を上回る速度で情報公開が進んでいる。世界で公開されている主要な分析データ (MassBase, PRIME MS2T, MetaboLights, KomicMarket)のうち、国内から公開されているデータは約 92%にも及んでおり、データ公開における我が国の貢献は大きい。

#### 4.3 研究課題名:生物種-代謝物関係データベースならびに代謝物-生物活性データベースの開発(奈良先端大グループ)

##### 研究開発実施内容及び成果

ゲノム決定において新型シーケンサなどにより出力されるショートリードなどの大量情報からゲノム全体の配列へとアセンブルし、それぞれの遺伝子の機能を注釈する。生物はこのゲノム情報にもとづいて、種々の酵素遺伝子を発現し最終的に代謝物を生合成し、生物の生存戦略における役割を担っている。メタボローム情報処理についてみると、質量分析装置により出力されるピークデータをもとに代謝物の推定を行う。ゲノム情報処理ならびに遺伝子注釈づけと同様に、生物が生合成する代謝物情報を得て、さらにこれらの代謝物についての生物学における機能を注釈づけることが必要とされる。そこで、本プロジェクトでは質量分析器からえられるピーク情報を候補代謝物と関連づけることを目的に、生物種と代謝物の関係データベース KNApSACk Core を、また、代謝物の生物学的機能に関する注釈づけを行うことを目的に代謝物と生理活性の関係データベース Metabolite Activity の開発を進めた。

##### (1)生物種と代謝物関係データベース(KNApSACk Core)ならびに代謝物と生理活性の関係データベース(Metabolite Activity)

近年メタボローム研究において使用される質量分析装置では、代謝物の精密分子量を測定できるため、精密分子量をもとに分子式さらには代謝物の候補を選択することができれば、ピークデータの解釈が非常に容易になる。生物種と代謝物の関係を文献情報から網羅したデータベースがあれば、生物サンプルにおける質量分析をもとに精密分子量から分子式を推定し、生体内の二次代謝物の候補を列挙することが出来る。大きな問題点は、天然物(代謝物)と生産する生物については科学論文に記載されているのみの情報が多く、それらをデジタル化し格納したデータベースは皆無であった。そこでメタボローム研究の効率化に寄与する目的で、学術論文を地道に調査し、データを蓄積し、生物種と代謝物の関係データベース **KNApSAcK Core DB** を開発した。また、質量分析データからの検索エンジンの開発も進め公開した。現在まで、**KNApSAcK Core DB** には 101,500 対の生物種-代謝物の関係、5 万種の代謝物、2 万種の生物が格納されており、世界のメタボローム研究者に認知され、現在、130 を越える論文において引用されるに至っている。

地球上のほぼ唯一の生産者は植物である。また、森林のような生態系では植物自身が放出する様々な物質を介した生物間相互作用によりバランスが成り立っていることがわかってきている。このような二次代謝物を通じた生物間の関係を理解することは、地球の生態系を分子レベルで理解することへとつながる。ゲノム科学を通して、このような二次代謝物を介した生物間の関係を理解し、代謝物の機能性としての注釈づけを可能にすることを目的に、科学文献調査により集めたデータをもとに **KNApSAcK Family 代謝物活性データベース(Metabolite Activity DB)** の開発を進めた。**Metabolite Activity DB** は、天然物-対象生物-生物活性の三組の関係からなるデータベースである。当初の目標は **KNApSAcK Core** の 1 割程度、すなわち 5,000 件を目標に天然物-対象

生物-生物活性の三組の関係を整理することを目標とした。結果としてではあるがその倍の 9,584 件のこれらの三組の関係を登録することができた。このうち約半分はヒトを除く生物活性であった。生物活性として摂食誘導物質“**feeding attractant**”によるデータ検索結果(図 5)において、天然物 **Dihydroquercetin** は生物種 **Scolytus mediterraneus** に対して摂食誘導物質としての活性があることを示している。さらにここで天然物の ID(C00000677)をクリックすることにより、**Dihydroquercetin** を生合成する生物種リストが得られる。このようにして、どのような生物が、どのような天然物を生合成し、どのような対象生物にいかなる生物活性を示すかという四組の関係を把握することができる。

The screenshot shows the 'Metabolite Activity' database interface. At the top, there is a search bar and a table with columns: C\_ID, Metabolite Name, Activity Category, Biological Activity (Function), Target Species, and Reference. A red box highlights the entry for C00000677, Dihydroquercetin (C15H16O10), which is categorized as a 'Feeding attractant' and is active against the target species *Scolytus mediterraneus*. Below the table, a detailed view for Dihydroquercetin is shown, including its chemical structure and a list of organisms that produce it, such as *Aspergillus nidulans*, *Aspergillus niger*, and *Aspergillus terreus*.

図 5 KNApSAcK Core ならびに Metabolite Activity DB の出力例、両方のデータベースによりある生物が生産する天然物(代謝物)が他の生物にどのように影響をおよぼすかを把握することができる。これにより代謝物の注釈づけが可能となる。

今後さらに、より詳細な生物学機能を代謝物に注釈づけすること、ならびに地球の生態系を分子レベルで把握することを目指し、植物二次代謝物において同様なデータを蓄積することにより、地球上全体における生物間の関係の DB 構築を進めたい。



## (2)データの公開

本プロジェクトで研究開発を進めた生物種-代謝物関係データベースならびに代謝物-生物活性データベースについては **KNApSAcK Family** ([http://kanaya.naist.jp/KNApSAcK\\_Family/](http://kanaya.naist.jp/KNApSAcK_Family/))により公開した。研究者のボランティアにより本プロジェクト以外で構築されたゲノム・メタボローム関係データベースについても同様のサイトから公開を進めている、**KNApSAcK Family** に整理されているデータベースの属性間の関係を図 6 に示す。**KNApSAcK family** における月ごとのアクセス数は、約 20 万件であり、年間 240 万程度のアクセスがある。現在までに約 110 の世界のドメインからアクセスがあり、その内訳は、おおよそ日本が 50%、中国 20%、米国 20%、その他の国 10%となっている。これらのデータは全て **National Bioscience Database Center (NBDC)**からも公開されている。なお、2012 年に **Plant Cell Physiology** に受理された **KNApSAcK Family** データベースに関する論文(Farit et al., 2012)は、2013 年 **PCP awards** を受賞するに至った。このことは、日本植物生理学会においてもこのようなデータベース構築を研究として評価していただいたことを意味しており、今後、このようなバイオインフォマティクス・データベース構築の発展に貢献すべく努力を続けたい。

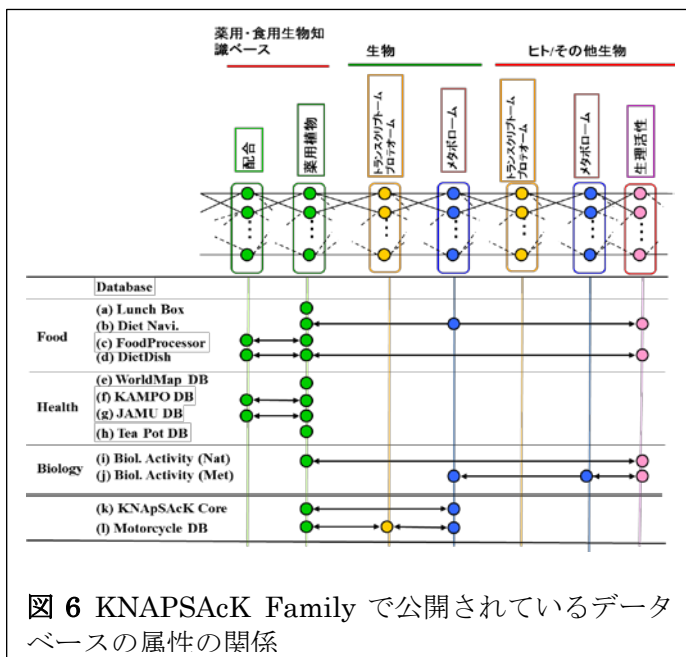


図 6 KNApSAcK Family で公開されているデータベースの属性の関係

## 4.4 研究課題名:メタボローム統合情報処理技術の研究開発(理研グループ)

### 研究開発実施内容及び成果

#### (1)統合スペクトル作成ソフトウェアの作成と、統合ライブラリ

複数機関で計測したスペクトルを簡単に比較検討することを目的に、複数スペクトルを並行表示して平均化や共通部分を拾い出すソフトウェアを構築した。これはアルゴリズム開発のために内部で利用するソフトウェアである。その結果、GCMS で同一標準物質を計測した場合でも全く異なるスペクトル(同一機関なら **MassBank ID OUF00116** と **OUF00117**、異なる機関では **PR010234** と **KZ000267** など)が観測されることがわかった(相違はおそらく誘導体化の相違に基づく)。こうした結果を体系的に整理し、標準物質毎に信頼のおけるスペクトルを掲載することは重要である。統計的に信頼できるスペクトルを「統合スペクトル」と呼ぶ。

統合スペクトルは実測値(またはその平均等)である必要はない。むしろ数理モデルや統計結果に基づいてノイズを除去し、理想とする仮想スペクトルであるべきと思われる。こうした思想に基づく(仮想)スペクトルライブラリは他には存在しない。

#### (2)RefSpec データベース(仮称)の開発

統合スペクトルの掲載先として **wiki** 形式の **RefSpec** サーバーを開発した。(名称は **Reference Spectra** の略称で、今後変更の可能性もある。)このデータベースでは特定のネームスペース上に統合スペクトルが掲載され、**Java Script** によりスペクトルを描画する。**(MassBank** は **Java Applet** を用いた実装であった。)関連機関にはこのサーバーを公開しているが、一般公開は今後内容を充実させてからの予定である。**MassBank** が各機関より投稿された標準スペクトルレポジトリとして機能するのに対し、**RefSpec(仮称)**はそれらを集約して作成した統合スペクトルのライブラリである点が特徴である。

### (3) データネットワーキング&クローリングソフトウェアの作成

RefSpec サーバーに掲載する統合スペクトルは、各機関が公開する実測スペクトルから統合ソフトウェアによって半自動的に作成する。そのため実測スペクトルを各所から定期的にクローリングする必要が生じる。スペクトル情報は頻繁に変更されるものではないので、データの差分を利用すれば効率よくクローリングできる。その基本となるソフトウェアを作成した。

### (4) MassBank light スペクトル検索ソフトウェアの開発

データベースにおけるデータ部分とそれを検索するソフトウェアは独立したコンポーネントであるべきである。そのためスペクトル検索ソフトウェアを単独のアプリケーションとして開発した。スペクトル検索のアルゴリズムは MassBank に実装されるものと同じで、部分構造検索を除いて基本機能は同じである。上記 RefSpec サーバーは一括ダウンロード可能なスペクトルを提供するサイトとし、検索その他は各ユーザーがローカルにおこなえる環境を整えた。

## § 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 13 件)

1. Nakamura Y, Afendi F M, Kawsar P A, Ono N, Tanaka K, Hirai A M, Sato T, Sugiura T, Altaf-Ul-Amin M, Kanaya S, KNApSAcK Metabolite Activity Database for Retrieving the Relationships Between Metabolites and Biological Activities, *Plant Cell Physiol.*, 55:e7.1-9 (2014) doi: 10.1093/pcp/pct176
2. Sakurai N, Ara T, Enomoto M, Motegi T, Morishita Y, Kurabayashi A, Iijima Y, Ogata Y, Nakajima D, Suzuki H, Shibata D, Tools and databases of the KOMICS web portal for preprocessing, mining, and dissemination of metabolomics data, *BioMed Research International*, in press (2014)
3. Tsugawa H, Arita M, Kanazawa M, Ogiwara A, Bamba T, Fukusaki E "MRMPROBS: Data Assessment and Metabolite Identification Tool for Large-scale MRM-based Widely Targeted Metabolomics" *Analytical Chemistry*, 85(10), 5191–5199, 2013 doi: 10.1021/ac400515s
4. Ogawa T, Furuhashi T, Okazawa A, Nakai R, Nakazawa M, Kind T, Fiehn O, Kanaya S, Arita M, Ohta D, Exploration of Polar Lipid Accumulation Profiles in *Euglena gracilis* Using LipidBlast, a MS/MS Spectral Library Constructed in Silico, *Bioscience Biotechnology and Biochemistry*, 78(1), 2014
5. Ikeda S, Abe T, Nakamura Y, Kibinge N, Hirai M A, Nakatani A, Ono N, Ikemura T, Nakamura K, Altaf-Ul-Amin M, Kanaya S, 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.*, 54, 711-27, 2013, doi: 10.1093/pcp/pct041
6. Afendi F M, Darusman L K, Morita A H, Altaf-Ul-Amin M, Takahashi H, Nakamura K, Tanaka K, Kanaya S, Efficacy prediction of jamu formulations by PLS modeling, *Curr Comput Aided Drug Des.*, 9, 46-59, 2013, doi: 10.2174/1573409911309010005
7. Katsuragi T, Ono N, Yasumoto K, Altaf-Ul-Amin M, Hirai M Y, Sriyudthsak K, Sawada Y, Yamashita Y, Chiba Y, Onouchi H, Fujiwara T, Naito S, Shiraishi F, Kanaya S, SS-mPMG and SS-GA: tools for finding pathways and dynamic simulation of metabolic networks, *Plant Cell Physiol.*, 54, 728-39, 2013, doi: 10.1093/pcp/pct052
8. Nakamura K, Shimura N, Otabe Y, Hirai M A, Nakamura Y, Ono N, Md Altaf Ul-Amin and Kanaya S, KNApSAcK-3D: A Three-Dimensional Structure Database of Plant Metabolites, *Plant Cell Physiol.*, 54,2,e4(1-8) ,2012, doi: 10.1093/pcp/pcs186
9. Sakurai N., Ara T., Kanaya, S, Nakamura, Y., Iijima, Y., Enomoto, M., Motegi, T., Aoki, K., Suzuki, H., and Shibata D., An application of a relational database system

- for high-throughput prediction of elemental compositions from accurate mass values., *Bioinformatics*, 29 (2): 290-291, 2013, doi:10.1093/bioinformatics/bts660
10. Afendi F M, Okada T, Yamazaki M, Hirai M A, Nakamura N, Nakamura K, Ikeda S, Takahashi H, Md. Altaf-Ul-Amin, Darusman L K, Saito K, Kanaya S, KNApSAcK Family Databases: Integrated Metabolite Plant Species Databases for Multifaceted Plant Research, *Plant Cell Physiol.*, 53,2,1-12,2012, doi: 10.1093/pcp/pcr165
  11. Takahashi H, Morimoto T, Ogasawara N, Kanaya S, AMDORAP: Non-targeted metabolic profiling based on high-resolution LC-MS, *BMC Bioinformatics*, 12, 259-260,2011, doi:10.1186/1471-2105-12-259
  12. Tsugawa H, Tsujimoto Y, Arita M, Bamba T, Fukusaki E, 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, 2011 doi:10.1186/1471-2105-12-131
  13. Redestig H, Kusano M, Ebana K, Kobayashi M, Oikawa A, Okazaki Y, Matsuda F, Arita M, Fujita N, Saito K "Exploring molecular backgrounds of quality traits in rice by predictive models based on high-coverage metabolomics" *BMC Systems Biology* 5(1), 176, 2011 doi:10.1186/1752-0509-5-176

(2)その他の著作物(総説、書籍など)

1. 西岡孝明、データベース MassBank、現代質量分析学 - 基礎原理から応用研究まで(高山光男、早川滋雄、瀧浪欣彦、和田芳直編)、第 32 章、465-480、発行 化学同人、京都市、2013.
2. 櫻井望、有田正規、西岡孝明、金谷重彦、メタボローム・データベースの開発、日本バイオインフォマティクス学会ニュースレター、No 27, 3-4 (October 2013).
3. 西岡孝明、スペクトルデータ部会の歴史、日本質量分析学会創立 60 周年記念号、質量分析、61(6), 59-61, 2013.
4. 中村由紀子、森田(平井)晶、西岡孝明、金谷重彦、KNApSAcK Family 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”、*バイオサイエンスとインダストリー*、70 (4), 267-272, 2012.
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. 金谷重彦、平井 晶、旭 弘子、高橋弘喜、中村建介、Md. Altaf-Ul-Amin、二瓶義人、池田奨、尾島雄也、西岡孝明、メタボロームデータベース:多様な研究への対応とデータの共有化、「使えるデータベース・ウェブツール」*実験医学*、29,(15), 2460 (2011).

(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 8 件、国際会議 8 件)

1. Afendi F M (ボゴール農大)、岡田岳人(徳島文理大)、山崎真巳(千葉大)、平井(森田)晶 (NAIST)、中村由紀子(NAIST)、中村建介(NAIST)、池田俊、高橋弘喜(千葉大)、Amin Md. Altaf-Ul(NAIST)、Darusuman Latifah K.(ボゴール農大)、斉藤和季(千葉大)、金谷重彦(NAIST)、KNApSAcK Family Databases: 植物研究における多目的活用に向けた生物種・代謝物関係データベース、*Plant Cell Physiology award*2013, 3/11, 富山, 2014
2. 有田 正規 「マススペクトルのデータベースを使いこなす」第 31 回日本植物細胞分子生物学会 バイオインフォマティクス講習会 9/12 (10-12) 札幌 2013
3. 西岡孝明(NAIST)、櫻井望(かずさ DNA 研)、二瓶義人(NAIST)、”AIST)10-データの解析から得られた ESI-MS/MS フラグメンテーション情報”、ワークショップ「フラグメンテーション: 反応機構と構造」、第 61 回質量分析総合討論会(1B-W1-1830)、日本質量分析学会主催、エポカル、つくば市、2013 年 9 月 10-12 日
4. Takaaki Nishioka(NAIST), MassBank and Bio-MassBank: Databases of Mass Spectra of Known and Unknown Metabolites toward Reconstruction of Metabolic

- Pathways by Metabolomics, Complex Biodynamics and Networks (cBio2013), Tsuruoka, Yamagata, 12, 13 November 2013.
5. Takaaki Nishioka, "MassBank, Database of Mass Spectra, and Its Application to Metabolomics, Autumn School of Chemoinformatics in Nara, 日本化学会情報化学部会、情報計算法学生物学会、東京大学工学研究科船津研究室共催、奈良県立新公会堂、奈良市、2013年11月27-28日。
  6. 櫻井望(かずさDNA研究所)、メタボロームデータの公共化と再利用に向けて何が必要か? 第8回メタボロームシンポジウム、2013年10月4日、九州大学医学部 百年講堂
  7. \*Nishioka, T., Introduction to MassBank, NORMAN MassBank workshop (NORMAN Association 主催), 11月27日, 2012, IVM VU University Amsterdam, Amsterdam, Netherland
  8. 西岡孝明(NAIST)、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、2012年10月10-12日、山形県鶴岡市
  9. 荒武(かずさDNA研究所)、MassBase活用術〜ドライからウエットまで、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
  10. Arita M Wiki-based Databases: Integration of knowledge through the web, C-NAIR Seminar, 2011年12月5日、Nasional, Gadjah Mada University (Indonesia)
  11. 有田 正規「Knowledge Management using a Wiki-based System」細胞を作る研究会、2011年10月27、大阪
  12. Arita M Update on Lipid Databank and other lipidomics initiatives, EMBL-EBI Industry Programme & MetaboLights Project Workshop, 2012年5月22-25日, Hinxton
  13. Arita M What can metabolomics learn from genomics and proteomics?, International Conference of Research and Application on Traditional Complementary and Alternative Medicine (TCAM), 2012年6月21-23日, Muhammadiyah University of Surakarta (Indonesia)
  14. Arita M Cyberinfrastructure for metabolomics and synthetic biology, International Symposium on Biotechnology for Green Growth, 2012, 10月24-26日, 神戸
  15. \*Nishioka T, Data processing, software, databases & standardization, Invited panelist of ASMS Metabolomics Interest Group, 59th ASMS Conference on Mass Spectrometry and Allied Topics, Colorado Convention Center, Denver, CO, USA. 6月5-9日, 2011. (Invited Panelistとして講演、討論)
  16. 有田 正規「コミュニティとして機能するためのインフラストラクチャ」第84回日本生化学会大会シンポジウム「ゲノムデザインの最前線」2011年9月22日、京都

②口頭発表(国内会議4件、国際会議3件)

1. CASMI2013 出題グループ、西岡孝明(奈良先端大)、””代謝物推定コンテスト CASMI2013 第8回メタボロームシンポジウム、2013年10月3、4日、九州大学医学部百年講堂、福岡市。
2. 西岡孝明(奈良先端大)、”メタボロミクス研究における MassBank の高度な利用法の開発”、FIRST 田中 ms3d プロジェクト成果発表会、2013年9月11日、エポカル、つくば市、2013.
3. \*Nishioka T(奈良先端大), Nihei Y(奈良先端大), Ojima Y(奈良先端大), & Ikeda T(奈良先端大). MassBank: Public Mass Spectral Database, 19<sup>th</sup> International Mass Spectrometry Conference, Kyoto International Conference Center, Kyoto, Japan, September 20, 2012.
4. \*Nishioka T(奈良先端大). “Constructing MassBank Database on PCs”, 19<sup>th</sup> International Mass Spectrometry Conference, Workshop Mass++ and MassBank: Tools for Data Processing and Database on PC, Kyoto International Conference

- Center, Kyoto, Japan, September 18, 2012.
5. \*Nishioka T(奈良先端大), Arita M(東大). Ojima Y(奈良先端大). Ikeda T(奈良先端大). & Nihei Y(奈良先端大). Chemical Annotation of MassBank ESI-MS2 Data, 59th ASMS Conference on Mass Spectrometry and Allied Topics, Colorado Convention Center, Denver, CO, USA. June 5-9, 2011.
  6. 荒武(かずさDNA研究所)、櫻井望(かずさDNA研究所)、鈴木秀幸(かずさDNA研究所)、柴田大輔(かずさDNA研究所)、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、2012年8月4日
  7. 西岡孝明(奈良先端大)、池田 奨(奈良先端大)、““先端大 hemic の検索から構造推定まで”、第4回バイオアナリシス研究会、大阪大学大学院工学研究科生命先端工学専攻サントリーメモリアルホール、吹田市、2011年7月8日。
- ② ポスター発表(国内会議14件、国際会議3件)
1. \*T.Schulze(UFZ), E.Schymanski(EAWAG), S.Neumann(IPB), J. Slobodnik(NORMAN), M. Krauss(NORMAN), Y. Nihei(奈良先端大), G. Schramm (UFZ), T. Nishioka(奈良先端大), J. Hollender(EAWAG), W. Brack(UFZ), NORMAN MassBank – bridging communities from metabolomics to environmental chemists, SETAC, Glasgow, England, June, 2013.
  2. 二瓶義人(奈良先端大)、西岡孝明(奈良先端大)、実用的・先進的なマススペクトルデータベース MassBank、第61回質量分析総合討論会(3P-01)、日本質量分析学会主催、2013年9月10-12日、エポカル、つくば市。
  3. 荒武(かずさDNA研)、榎本光男(かずさDNA研)、二瓶義人(奈良先端大)、有田正規(理研)、金谷重彦(奈良先端大)、西岡孝明(奈良先端大)、櫻井望(かずさDNA研)、メタボロミクス研究におけるメタデータに特化した Wiki による管理システム、第61回質量分析総合討論会(3P-01)、日本質量分析学会主催、2013年9月10-12日、エポカル、つくば市。
  4. 青島健(エーザイ)、福田充(エーザイ)、井川幹之(エーザイ)、高橋健太郎(エーザイ)、木村剛之(エーザイ)、梶原茂樹(島津田中最先端研)、宇都宮眞一(島津田中最先端研)、田中聡(島津田中最先端研)、吉沢明康(島津田中最先端研)、田中耕一(島津田中最先端研)、池田奨(奈良先端大)、二瓶義人(奈良先端大)、西岡孝明(奈良先端大)、小田吉哉(エーザイ)、小質量分析用データ解析ソフトウェア Mass++ とマススペクトルデータベース MassBank 連携プラグインの開発、第61回質量分析総合討論会(3P-08)、日本質量分析学会主催、2013年9月10-12日、エポカル、つくば市。
  5. 田中聡(島津田中最先端研)、吉沢明康(島津田中最先端研)、草野麻衣子(島津田中最先端研)、藤田雄一郎(島津田中最先端研)、福田充、宇都宮眞一(島津田中最先端研)、梶原茂樹(島津田中最先端研)、青島健(エーザイ)、小田吉哉(エーザイ)、二瓶義人(奈良先端大)、西岡孝明(奈良先端大)、田中耕一(島津田中最先端研)、田中耕一(島津田中最先端研)、MassBank を用いた脂質同定システムの構築、第38回日本医用マススペクトル学会年会、2013年9月26-30日、神戸市産業振興センター、神戸市。
  6. 荒武(かずさDNA研究所)、榎本光男(かずさDNA研究所)、二瓶義人(奈良先端大)、有田正規(理研)、金谷重彦(奈良先端大)、西岡孝明(奈良先端大)、櫻井望(かずさDNA研究所)、Metablonote: メタボロミクス研究におけるメタデータ専用の Wiki による管理システム、第8回メタボロームシンポジウム、九州大学百年講堂、2013年10月3、4日
  7. 荒武(かずさDNA研究所)、榎本光男(かずさDNA研究所)、二瓶義人(奈良先端大)、有田正規(理研)、金谷重彦(奈良先端大)、西岡孝明(奈良先端大)、櫻井望(かずさDNA研究所)、Metablonote: メタボロミクス研究におけるメタデータに特化した Wiki による管理システム、第61回質量分析総合討論会、つくば国際会議場、2013年9月12日
  8. \*Nishioka T(奈良先端大), Nihei Y(奈良先端大), Ojima Y(奈良先端大), & Ikeda T(奈良先端大). " MassBank: Public Repository for Sharing Mass Spectral Data of

Metabolomics", 8<sup>th</sup> International Conference of the Metabolomics Society, Marriot Wardman Park Hotel, Washington, DC, USA. June 25-28, 2012.

9. 池田奨(奈良先端大)、二瓶義人(奈良先端大)、尾鷲雄也(奈良先端大)、西岡孝明(奈良先端大)、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 年 10 月 10- 12 日
10. 荒武(かずさ DNA 研究所)、榎本光男(かずさ DNA 研究所)、西岡孝明(奈良先端大)、有田正規(理研)、金谷重彦(奈良先端大)、櫻井望(かずさ DNA 研究所)、Metabolonote: メタボロミクスのデータリソースを統合するメタデータ専用管理システム、トーゴの日シンポジウム 2013、時事通信ホール、2012 年 10 月 5 日
11. 荒武(かずさ DNA 研究所)、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、2012 年 10 月 5 日
12. \*Tanaka S(島津田中最先端研)、Kajihara S(島津田中最先端研)、Utsunomiya S(島津田中最先端研)、Tabata T(エーザイ)、Aoshima K(エーザイ)、Oda Y(エーザイ)、Nihei Y(奈良先端大)、Nishioka T(奈良先端大)、Tanaka K(島津田中最先端研). Development of a Client Tool for a Mass Spectral Database, The 59th Conference on Mass Spectrometry and Allied Topics, American Society of Mass Spectrometry, Denver, CO. USA, Jun 5-9, 2011.
13. 二瓶義人(奈良先端大)、池田奨(奈良先端大)、尾鷲雄也(奈良先端大)、西岡孝明(奈良先端大)、マススペクトルデータベース MassBank の機能拡張、第 59 回質量分析学会総合討論会(日本質量分析学会主催)、ホテル阪急エキスポパーク、吹田市、2011 年 9 月 13 日~9 月 15 日。
14. 池田奨(奈良先端大)、二瓶義人(奈良先端大)、西岡孝明(奈良先端大)、先端学会総合ホスティングサービスを活用して WAN 上でマススペクトルを共有する、第 59 回質量分析学会総合討論会(日本質量分析学会主催)、ホテル阪急エキスポパーク、吹田市、2011 年 9 月 13 日~9 月 15 日。
15. 尾鷲雄也(奈良先端大)、二瓶義人(奈良先端大)、有田正規(東大)、西岡孝明(奈良先端大)、代謝物の ESI-MS<sup>2</sup> データにおけるピークと部分化学構造式の関係、第 59 回質量分析学会総合討論会(日本質量分析学会主催)、ホテル阪急エキスポパーク、吹田市、2011 年 9 月 13 日~9 月 15 日。
16. Howell Parry(島津田中最先端研)、田中聡(島津田中最先端研)、田畑剛(エーザイ)、青島健(エーザイ)、小田吉哉(エーザイ)、二瓶義人(奈良先端大)、西岡孝明(奈良先端大)、宇都宮真一(島津田中最先端研)、梶原茂樹(島津田中最先端研)、田中耕一(島津田中最先端研)、宇都宮真一(島津田中最先端研)質量分析解析フリーウェアの新バージョン、第 59 回質量分析学会総合討論会(日本質量分析学会主催)、ホテル阪急エキスポパーク、吹田市、2011 年 9 月 13 日~9 月 15 日。
17. 櫻井望(かずさ DNA 研究所)、荒武(かずさ DNA 研究所)、鈴木秀幸(かずさ DNA 研究所)、柴田大輔(かずさ DNA 研究所)、オミクスデータ統合解析のための代謝経路地図ツール KaPPA-View4、トーゴの日シンポジウム 2011、日本科学未来館、2011 年 10 月 5 日

#### (4)知財出願

##### ① 内出願 (0 件)

本プロジェクトではデータベースならびにツール群は、開発研究を進めると同時に直ちに公開したため、出願はない。

##### ②海外出願 (0 件)

本プロジェクトではデータベースならびにツール群は、開発研究を進めると同時に直ちに公開したため、出願はない。

③その他の知的財産権

本プロジェクトではデータベースならびにツール群は、開発研究を進めると同時に直ちに公開したため、出願はない。

(5)受賞・報道等

① 受賞(顕著な受賞の前に\*を付記してください)

\*1. Plant Cell Physiology 2013 awards

Afendi FM. (ボゴール農大), 岡田岳人(徳島文理大)、山崎真巳(千葉大)、平井(森田)晶(NAIST)、中村由紀子(NAIST)、中村建介(NAIST)、池田俊、高橋弘喜(千葉大)、Amin Md. Altaf-UI(NAIST)、Darusuman Latifah K.(ボゴール農大)、斉藤和季(千葉大)、金谷重彦(NAIST)、日本、KNAPSAcK Family Databases: 植物研究における多目的活用に向けた生物種・代謝物関係データベース

「日本植物生理学会における国際誌 Plant Cell Physiology に投稿した多目的活用に向けた生物種・代謝物関係データベースに関する論文が、2013 年度論文賞として評価された。本プロジェクトで開発した成果物についての国際的評価と位置付けられる。」

\*2. 日本植物細胞分子生物学会 2012 年度技術賞

櫻井望、「トランスクリプトーム解析手法の開発と受託サービスの実用化」というテーマで、大場利治博士(タカラバイオ株式会社ドラゴンジェノミクスセンター)ら他3名と共同受賞した。

3. 日本植物細胞分子生物学会 2012 年度論文賞

Kosuke Kai(大阪府大), Hiroki Takahashi(NAIST), Hirohisa Saga(大阪府大), Takumi Ogawa(大阪府大), Shigehiko Kanaya(NAIST) Daisaku Ohta(大阪府大) (2011) Metabolomic characterization of the possible involvement of a Cytochrome P450, CYP81F4, in the biosynthesis of indolic glucosinolate in Arabidopsis, Plant Biotechnology, 28, 379-385.

② マスコミ(新聞・TV等)報道

1. 「未来の講義 奈良先端大の研究から、植物が含む 5 万の物質 DB 化」(毎日新聞、奈良、2012 年 5 月 29 日)
2. 「進むメタボロミクス研究:奈良先端科技大、5 万種をデータベース化」(健康食品新聞、2012 年 6 月 27 日)

③ その他

なし

## § 6 研究開発期間中に主催した会議等

主なワークショップ、シンポジウム、アウトリーチ等の活動

年月日	名称	場所	参加人数	目的と概要
H23.7.30	MassBank データ監修会	奈良先端科学技術大学院大学	15名	日本質量分析学会スペクトルデータ部会と共催。MassBankのデータを監修して、品質が劣るデータの削除、修正をおこなった。
H24.9.17	質量分析の基礎と最先端(特別勉強会)～プロダクトイオンスペクトルの解釈から薬物動態質量分析へ	国立京都国際会議場	80名	第19回国際質量分析学会の日本語プログラムとして、MassBankを紹介、解説。
H24.9.18	Workshop “Mass++ and MassBank: Tools for Data Processing and Database on PC”	国立京都国際会議場	150名	第19回国際質量分析学会でMassBankを紹介するワークショップを田中最先端研と共催。PCでデータベースを構築、利用する方法を紹介。
H24.11.27	NORMAN MassBank workshop (NORMAN 連合主催)	IVM VU University Amsterdam, Amsterdam, Netherland	44名 (EU19ヶ国、日本)	EUの公的環境科学研究機関が所属するNORMAN連合が主催して、MassBankの利用講習会を開催。
H25.3.22	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	約150名	植物関連のデータベースの統合化について、植物研究者に紹介することを目的に、金谷グループと田畑グループからそれぞれ1演題ずつ発表を行った。
H25.9.5～ H25.9.6	Mass Spectrometry Informatics in Natural Product Research	公立学校共済組合「立山高原ホテル」	20名	天然物化学とマスマススペクトル情報を融合させるための国際ワークショップ
H25.9.3～ H26.2.3	Critical Assessment of Small Molecule Identification (CASMI2013)国際コンテスト・本NDBCプロジェクトと日本質量分析学会との共催	Internet上 <a href="http://www.casmi-contest.org/2013/index.shtml">http://www.casmi-contest.org/2013/index.shtml</a>	出題:11名 解答:7名 研究グループ(国内1、国外6)	マスマススペクトルから化学構造式を推定する国際コンテストを開催した。16マスマススペクトルを出題した。解答者は分子式、化合物を3ヶ月をかけて解答した。
H25.7.19	第9回質量分析技術者近畿ブロック研究会	奈良先端科学技術大学院大学	30名	大学で質量分析に携わる研究者を対象として、MassBankを紹介、解説。
H25.12.12	ビッグデータが世界を変える、あなたに迫る超大規模データ	奈良先端大	305名	一般市民を対象に、メタボロミクスの生活へのつながりを公開講座として講演した。



H26.1.26 ～ H26.1.31	BioHackason H13.13 + 実務者会議	万国津梁館	50名	ハッカソンと合同で統合プロジェクトに参加する6グループが連携するための合同会議
---------------------------	-------------------------------	-------	-----	-----------------------------------------

## § 7 ユーザー評価結果への対応

≪平成25年7月に実施した「NBDCにおける事業活動のユーザー評価」(<http://biosciencedbc.jp/user-hyouka-2013/result-summary>)で得られたユーザーの意見、提案等(詳細は別紙2を参照)に対し、実施済み若しくは実施予定の対応策等を具体的に記載してください。)

### ① 実施済み

MassBank、Bio-MassBankなどのスペクトルデータベース、ならびに代謝物情報を中心としたデータベース KNApSAcK Familyにおいて、メタボロミクスの分野からは世界的に高く評価されるに至っている。一般利用者がいかにこの分野に関心をもっていたか、今後の課題であり、随時マニュアルを完備する努力を進めている。また、メタボロミクスから広がる化学生態学、医薬学の分野への関連づけについてさらにデータベースを拡張すべく研究開発を進めているところである。

### ② 実施予定

メタボロミクスに関わる研究分は、ゲノム生物学にはじまり医薬学、食品学、さらには生態学と拡大しつつあるなかで、多様なユーザーのニーズに応えられるようにデータベースの拡張を図りたい。

## § 8 その他

(1) 研究代表者として、研究開発、プロジェクト運営等について、上記以外に報告したいことがあれば、自由に記載してください。

学会附属展示会において成果物の広報活動ならびにユーザーからの要望収集を積極的に行った。

櫻井望(かずさDNA研)、メタボロミクスのポータルサイト KOMICS、日本食品科学工学会第60回記念大会附属展示会、実践女子大、2013年8月29日～8月31日

櫻井望(かずさDNA研)、メタボロミクスのポータルサイト KOMICS、第61回質量分析総合討論会附属展示会、つくば国際会議場エポカルつくば、2013年9月10日～9月12日

櫻井望(かずさDNA研)、メタボロミクスのポータルサイト KOMICS、第65回日本生物工学会附属展示会、広島国際会議場、2013年9月17日～9月20日

金谷重彦(奈良先端大)、櫻井望(かずさDNA研)、有田正規(理研)、メタボローム・データベースの開発、第36回日本分子生物学会年会附属展示会、神戸コンベンションセンター、2013年12月3日～12月5日

櫻井望(かずさDNA研)、金谷重彦(奈良先端大)、メタボローム・データベースの開発、日本分子生物学会2014年度大会附属展示会、明治大学生田キャンパス、2014年3月28日～3月30日(予定)

以上