

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）
「生物種メタボロームモデル・
データベースの構築」
研究開発期間：平成26年4月～平成29年3月

研究開発終了報告書

研究代表者：有田正規
(理化学研究所環境資源科学研究センター
チームリーダー)

§1. 研究開発実施の概要

研究開始時に研究開発目的として主要三項目を掲げた。それぞれの項目について、全体計画書に記した内容に沿ってその実現状況を記述する。

1. メタボローム情報のアノテーション

理化学研究所およびかずさDNA研究所で計測した大量の代謝物測定データより、代謝物ピークを抽出し、通常よくある測定生データではなく、代謝物アノテーションを施した情報として公開した。具体的には141生物種の1114サンプル(解析データ3079件)、量にして10テラバイト以上のデータを解析し、とりわけ野菜に関して20種以上のメタボロームを明らかにした。

フォーマットはTogoMD形式とし、同定結果はかずさDNA研究所のKomicMarket2と理化学研究所のPRIME、測定条件はMetabolonoteより公開している。代謝物データはPRIME、KomicMarket2だけでなく、新規に開発したMassBank-wikiサーバからも公開している。測定条件は、63405トリプルのRDFとしてNBDC RDFPortalに提供した。

本作業に必要な代謝アノテーションソフトウェアとして、かずさDNA研究所でFT-ICR用ソフトウェアPowerGetと、フラボノイドに特化したFlavonoidSearch、理化学研究所でスペクトルから構造を予測するソフトウェアMS-FINDER等を開発した。

2. 文献情報の蓄積と検証

奈良先端科学技術大学院大学ではケミカルエコロジーや生薬学における代謝物の生理活性(生物活性)情報を、学術文献や一般書籍から幅広く収集した(それぞれKNApSAcKデータベースのBiological Activity, Metabolite Ecologyセクションで公開)。生物活性を記載する際には、代謝物—生物活性、代謝物—生物種の関係情報の統計が取れるように活性についてのオントロジーも構築した。具体的な件数は以下のとおりである。

- ・代謝物+生物活性情報 10758 エントリー
- ・代謝物+生態活性+生物種情報 8285 エントリー

前期から構築を続けるKNApSAcKデータベースは50899エントリーに達し、薬用植物情報も生物-利用国の組み合わせで56275件(9924件、地域-食用植物、46351件、地域-薬用植物)収集した。これらの文献情報はBioCycプロジェクトやKEGG等のゲノムデータベースから収集した代謝物アノテーション結果の検証に利用した。

3. データ統合とモデルDB作成

代謝物アノテーションの結果から、生物種ごとの代謝物リストを作成した。とりわけBioCycプロジェクトにおいて化合物情報が利用可能なシロイヌナズナおよびモデル植物について、二次代謝物および脂質を既知の代謝パスウェイと照合し、生合成および修飾遺伝子を介してゲノム情報と代謝物をリンクした。この情報をBio-MassBankにおける情報と統合し、MassBank-wiki上にモデル生物種のメタボローム情報を掲載した。

MassBank-wikiにおいてマスマスペクトルを閲覧する仕組みとして、JavaScriptによる新しいスペクトルブラウザを作成した。このブラウザはJavaScriptのみで構築され、各種ブラウザに対応している。

各種情報を共有する汎用システムとして、JSONフォーマットのデータをユーザ単位でアップロード、共有、ダウンロードできるMetabolome Node Network (MNN)システムを構築した。このサイトにマスマスペクトルや代謝物情報を登録し、変更トリガーを元に情報を共有できる仕組みを作成した。

§2. 研究開発実施体制

1. 研究グループ

(1) 「理研」グループ(研究代表者グループ)および遺伝研グループ

(平成28年度のみ、研究代表者の本務先である国立遺伝学研究所にグループを設置し人員を雇用した。)

人員構成

氏名	所属機関	役職	研究開発項目	参加時期
有田 正規	理化学研究所 環境資源科学研究センター	チームリーダー	総括、メタボモデルの構築、	H26.4～ H29.3
福島 敦史	理化学研究所 環境資源科学研究センター	研究員	メタボモデルの構築	H26.4～ H29.3
津川 裕司	理化学研究所 環境資源科学研究センター	特別研究員	プログラム開発	H26.4～ H29.3
Ramon Francisco Mejia	理化学研究所 環境資源科学研究センター	特別研究員	プログラム開発	H26.4～ H28.12
川島 武士	国立遺伝学研究所	特定有期雇用職員	生物系統検証	H28.10～ H29.3
Nadeeka Nilmini Hettiarachchi	国立遺伝学研究所	技術補佐員	生物系統データ作成	H28.7～ H29.3
多田 一風太	総合研究大学院大学 (国立遺伝学研究所)	学生(修士相当)	生物系統データ作成	H28.11～ H29.3

担当項目

有田：メタボモデルを搭載するプラットフォームとして、MassBank-Wikiを構築。旧 MassBank サーバにおけるスペクトル検索、閲覧ツールを全て Wiki 向けに JavaScript にて作成。

津川：MS-DIAL, MS-FINDER などのアノテーション用ソフトウェアツールを作成。

Mejia：海外サーバとデータを共有するための基盤ソフトウェア MNN を構築。

福島、川島：ゲノムデータベースより、代謝ネットワークを介してモデル生物種毎の代謝物リストを作成。

Hettiarachchi、多田：代謝物リストの確認と入力。

(2) 「奈良」グループ(主たる共同研究者グループ(1))

人員構成

氏名	所属機関	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	総括、生物活性の関係データベース構築	H26.4～ H29.3

氏名	所属機関	役職	研究開発項目	参加時期
森田 晶	同上	研究員	代謝物と生物活性の関係データベース構築	H26.4～ H29.3
中村 由紀子	同上	研究員	同上	H26.4～ H27.3
Sony Hartono Wijaya	同上	D1-D3	同上	H26.12～ H29.3
大橋 美名子	同上	研究技術員	同上	H26.12～ H29.3
Azian Azamini binti Abdullah	同上	D2-D3	同上	H27.4～ H29.3
白 禹	同上	D2-D3	同上	H27.4～ H29.3
黄 銘	同上	特任助教	同上	H28.4～ H29.3
中本 雅俊	同上	特任助教	同上	H28.4～ H29.3
白石 磨貴男	同上	D1-D2	同上	H27.4～ H29.3
渡部 聡之	同上	D1-D2	同上	H27.4～ H29.3
和田 佳子	同上	D1-D2	同上	H27.4～ H29.3
劉 康	同上	D1-D2	同上	H27.4～ H29.3
津曲 優子	同上	D1-D2	同上	H27.4～ H29.3
渡邊 琢朗	同上	D1	同上	H27.4～ H28.3
Nelson K. Kibinge	同上	D3	同上	H27.1～ H28.3
Li Dong Hang	同上	D3	同上	H27.1～ H28.3
江口 遼平	同上	M2-D1	同上	H28.3～ H29.3
山森 明弘	同上	D1	同上	H28.4～ H29.3

担当項目

金谷： 関係データベースの設計と構築、データ入力。

森田： 関係データベースのレイアウト等、インターフェースの開発。

その他： 関係データベースの入力。

(3)「かずさ」グループ(主たる共同研究者グループ(2))
人員構成

氏名	所属機関	役職	研究開発項目	参加時期
櫻井 望	かずさ DNA 研究所メタボロミクスチーム	チームリーダー	総括、メタボローム DB の構築	H26.4～ H29.3
山田 学	同上	プロジェクト技術員	プログラム開発	H26.4～ H27.3
秋元 奈弓	同上	特任研究員	メタボローム DB の構築	H26.7～ H29.3
池田 千晶	同上	プロジェクト技術員	プログラム開発	H26.7～ H29.3
印出 美幸	同上	補助員	データ整理・入力	H28.4～ H28.9

担当項目

櫻井: PowerGet, KomicMarket2 等、ソフトウェアツールの作成と統括。

山田: Metabolonote サーバの構築。

秋元: フラボノイドデータベースの構築、アノテーション。

池田: KomicMarket2 サーバの作成。

印出: KomicMarket2 におけるデータ整理。

2. 有識者会議等

DBCLS が開催するスパークルソンに隔月で参加し、統合化推進プログラムのテーマを実施した。その中で、外部研究者との意見交換等を実施したために、有識者会議等は実施していない。また多くの学協会と連携しながら進めてきたため、それら学協会における会合やワークショップが、そのまま有識者会議として機能している。

スパークルソンの開催履歴は以下のウェブサイト に詳述してある。

<http://wiki.lifesciencedb.jp/mw/index.php/SPARQLthon>

基本的に、隔月で統合化推進プログラムの各研究開発課題のメンバーも参加し、スパークルソン内で協議および発表をおこなった。

§3. 研究開発の目的、実施内容及び成果

1. 研究開発の背景

本研究開発を実施するに至った経緯

1. 世界におけるメタボロームプロジェクト

平成25年度末時点で、世界には化合物標準マススペクトルを集めたDBや、生物サンプルの測定結果を集めるDBが点在し、少なくとも以下の5サイトで標準マススペクトルを利用できた。

1. NIST/EPA/NIH ライブラリ (Natl. Instit. Standards Technol., 米国)
GC/MSでは標準的な有償のスペクトルライブラリで、実質の世界標準であった。MS/MSスペクトルも収載を開始しはじめた。
2. SDBS (産総研、日本)
マススペクトルを含む様々なスペクトルのライブラリだが、一括ダウンロードは不可。またESI マススペクトルは含まない。
3. METLIN (Scripps 研究所、米国)
XCMS 解析ソフトウェアと連携したオンラインスペクトル検索を提供。しかしライブラリは非公開でダウンロードも不可能。
4. GMD (Max-Planck Instit. Mol. Plant physiol., ドイツ)
GC/MSの実測等を収載したライブラリ。植物の代謝物が中心だが、開発者が不在になり、近年はアップデートされていない。
5. HMDB (アルバータ大学、カナダ)
ヒト試料で検出される代謝物について GC/MS 等のスペクトルを公開し、その他、代謝物構造や文献情報を充実させている。

いずれも、主要機関が統一された実験条件で測定したスペクトルのDBであり、それ以外の研究グループが得たスペクトルを第三者が再利用するための仕組みではなかった。またデータを無償公開していても、ライセンス形態が不明であった。

平成25年、欧州バイオインフォマティクス研究所 (EBI) は、MetaboLights プロジェクトを開始した。MetaboLights は、測定データ (機器依存フォーマットを許す) と実験条件を集積する国際共同DBである。投稿データには恒久的なID番号を付与し、EBI が責任を持って情報を維持するとしている。また同年、米国保健省 (NIH) はヒト由来試料を対象を限定したメタボローム DB 構築を開始し、公式リポジトリである Metabolomics Workbench (Univ. California San Diego) を開始した。NIH グラントによる研究成果は当該リポジトリへの登録が義務付けられ、その後、ダウンロード可能な生データの規模は急速に拡大した。

欧米が相次いで公式の恒久的なリポジトリを開設したのに対し、国内では生データを格納するリポジトリを構築できる組織は資金や規模の面で存在しなかった。ただし MassBank という標準スペクトルのデータベースは世界的に知られていたため、所有するリソースを存在感を持って世界に示しつつ、欧米の動きに遅れを取らない体制づくりが早急に望まれていた。

2. 第一期統合化推進プログラム後に抱えていた課題

第一期統合化推進プログラムにおいて申請者らは、化合物標準マススペクトルDBおよび生物サンプルからのマススペクトルDB (MassBank, Bio-MassBank)、生物種および生理活性と代謝物の関係DB (KNAPSAcK Family Databases)、生物由来サンプルの分析データおよび実験条件の情報DB (MassBase, Metabolonote, KomicMarket) など幅広いDBを構築・無償公開していた。とりわけマススペクトルDBであるMassBankは、化合物ごと、実験条件ごとに28研究組織の情報を集約したオープンアクセスDBとして (平成25年度末時点で) 世界最大規模であり、上述のマススペクトルDBと並んで国際的に利用されていた (標準スペクトル数は4万件)。またヨーロッパにNorman MassBankと呼ばれる支部を設置し、名実ともに国際公共DBとして機能していた。しかし検索のインターフェースが各機関における分析条件や、スペクトルの類似度に限られており、

例えば、MassBank 内に登録される化合物数すら把握できていなかった。また、数十以上の類似スペクトルが登録される代謝物がある一方で、精度の低いスペクトルもあり、精度の観点からは必ずしも標準スペクトルとは言えなかった。端的に言えば、情報集約型のサイトが抱える問題を例外なく抱えていた。

技術面で致命的であったのは MassBank と Bio-MassBank のソースコードが部分的に失われていた点である。基本となる Apache サーバ上に C++、Perl、Java を駆使した複雑な構造が組み上がっており、システムのバージョン更新やデータベースの機能改変ですら、極めて難しい状態にあった。また利用可能な Apache や Tomcat のバージョンが限定され、バージョンを殆んど変更できなかった点は、セキュリティ面からも致命的であった。スペクトルを閲覧するインターフェースも Java Applet という古い技術に基づいており、これも一新する必要に迫られていた。

3. 本研究開発課題における対象データベースとの相違点や連携内容

本研究課題はまず、世界に点在するメタボロームデータベース間で連携する枠組みを作ること前提とした。平成25年度末時点で、欧米と同レベルのリポジトリを整備することは明らかに不可能であった。当時の国際的な地位を保つことすら困難に思われた。また同時に、セキュリティの観点から、ソースコードが失われた MassBank と Bio-MassBank のサーバ全体を作り直す作業も必要であった。そのため、統合化推進プログラムに同期に採択される研究チームや DBCLS と定期的に研究会を開催し、オントロジーの整備と標準フォーマットを通して連携することを第一目標とした。

国際レベルの連携方針として、日米欧の三極構造を目指すことにした。欧州には既に Norman MassBank があり、環境メタボロミクス情報を登録していた。研究開始時点では米国において対応するサーバが不在であった。当時より共同研究を実施していた UC Davis のオリバー・フィーン研究室と新サーバ MassBank Of North America (MONA) の設置を検討し、2015 年より MONA データベースが稼働しはじめた。スペクトルや測定サンプルの数では、コホート研究等を長年実施してきた Fiehn 研究室の MONA が圧倒的に多くのリソースを提供できる。そこで本課題では、日本の独自性を出す一つの方向として、標準スペクトルに注力すること、そして生データのリポジトリではなく、アノテーション結果から得られる知識を集約するサイトの構築に注力した。(つまり生データや個々の計測結果は扱わない。)この方針は初期 MassBank の設立主旨にも沿うものであり、Metabolights および Metabolomics Workbench との役割分担を目指したものであった。

4. 研究開発開始以降の背景変化と、課題への影響

この後、セクション3「達成目標及び実施計画」の(2)「期間中に追加・削除・変更した実施計画・達成目標」において、背景を含めた変化を詳述する。

2. 研究開発対象のデータベース・ツール

(1) データベース

主要なもの

正式名称	略称	概要
MassBank および MassBank-wiki	MassBank-wiki	代謝物標準マススペクトルのデータベース
KNApSAcK Database	KNApSAcK	代謝物—生物種関係のデータベース
KomicMarket2	KomicMarket2	代謝物アノテーション結果のデータベース

上記以外のもの

正式名称	略称	概要
------	----	----

MassBase	MassBase	生物種メタボローム生データのデータベース
Metabolonote	Metabolonote	メタボローム解析メタ情報のデータベース
KNApSAcK Biological Activity	Biological Activity	代謝物の生物活性のデータベース
KNApSAcK Metabolite Ecology	Metabolite Ecology	代謝物、代謝物質の生態学的活性および対象生物という三つ組のデータベース

(2) ツール等

正式名称	略称	概要
Metabolome Node Network	MNN	JSON フォーマットのデータをクラウド形式で共有するためのプラットフォーム
PowerGet	PowerGet	代謝物アノテーション用ソフトウェア
MS-FINDER	MS-FINDER	代謝物構造予測ソフトウェア
FlavonoidSearch	FlavonoidSearch	フラボノイド構造予測ソフトウェア

※データベース、ツールの詳細は別紙参照。

3. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

研究開始当初に記載したガントチャート

	26年度	27年度	28年度
理研	ナズナ、イネ、酵母モデル データの共有ソフト	の作成 ウェアの開発	有用種モデルの作成 データの共有の実施
奈良		生物種—代謝物情報蓄積 代謝物—生物活性蓄積	
かずさ		データ大規模整備 外部DBとの連携	
	ソフトウェア技術	開発	

研究開発当初に記載した「マイルストーン」および「開発するデータベースのまとめ」

マイルストーン

・2014年度末

連携する統合化推進プログラム4チーム間に共通するアクセス用ガジェット
の作成
代謝プロファイルデータ500件の登録

・2015年中旬 モデル生物種(ナズナ、イネ、酵母)のメタボモデル作成

・2015年度末 モデルDBの開始、MassBankサーバの廃止

- ・2016 年度 メタボローム国際連携の中で代謝リファレンス情報を共有

開発するデータベースのまとめ

本計画では第一期統合化推進プログラムの方針に従い、グループ間でゲノムおよび代謝物オントロジーを共通化する。共通化はそれぞれのデータベース毎におこない、各サーバを物理的に統合・廃止することはしない。そのため KNApSAcK, MassBase, Metabolonote, MassBank など個々のデータベースは個別にアップデートをおこなう。その間に、Bio-MassBank の情報を再整理してモデル DB として再構築する。その後、MassBank サーバを停止してモデル DB に統合する。

データ処理とアノテーションの方針を下表に示す。成果は全て無償公開する。

データ型及び現在の DB 名	統合の内容と公開時期・方法	連携先 (T:研究チーム)
代謝物マススペクトル (MassBank, Bio-MassBank, MassBase, PRIME, KomicMarket)	スペクトルを測定条件、生物種毎に整理して統合。スペクトル標準として各機関に頒布。Bio-MassBank に含まれる情報をモデル DB に改組。	日本質量分析学会、MassBank コンソーシアム、欧州の標準化コンソーシアム
マススペクトル測定条件 (Metabolonote, PRIME)	測定法のオントロジー作成。測定情報は Metabolonote に統合し RDF 化。	JST -NSF メタボロミクス
代謝化合物 (KNApSAcK, LipidBank, Metabolomics.JP)	生物種との関係データを拡充しつつ、InChI, CAS 番号を付与。糖鎖や母核構造など特徴を付与して構造化。	日本脂質生化学会、成松 T、JST 日化辞
生物活性、生薬効能等 (KNApSAcK, Metabolomics.JP)	生物種との関係データを拡充しつつ、オントロジーを作成して RDF 化。	和漢医薬学会、植物細胞分子生物学会、榎屋 T
生物種(系統分類) (KNApSAcK, Metabolomics.JP)	Taxonomy ID を付与。GBIF 等を介して NBRP および INSDC と生物種情報を共通化。	遺伝研 NBRP, 田畑 T, 黒川 T, 中村 T

4. 実施内容

(1) 実施内容

理研グループ

理研1. リファレンススペクトルサーバ MassBank wiki の構築(有田、津川、Mejia)

MassBank wiki は旧 MassBank のスペクトルのみならず、スペクトル精度を精査し、修正 (curate) したスペクトルセットも掲載する。研究開始時点では旧 MassBank における化合物数すら不明だったが、InChIKey を用いて整理することで 3127 化合物の ESI-MS/MS スペクトルを確認できた。そのうち 2098 化合物のスペクトルが信頼できる精度であった²¹⁾。MassBank wiki ではこのコアセットを検索、ダウンロード可能にしている。

マススペクトルを拡大縮小しながら複数比較できる JavaScript インターフェースを完成させ、MediaWiki サーバ上で動作するようにした(図1)。また質量やピーク値に基づく検索インターフェースを作成し、MassBank wiki サーバ上で稼働させている。

ブラウザ機能の最大の特徴は、(旧 MassBank におけるブラウザや他データベースと異なり)

- ・ 100% JavaScript で OS やブラウザ非依存であること
- ・ http アドレスのコピー&ペーストで同一画面を復元できること(いわゆるステートレス)

- レイアウトの順序やサイズをインタラクティブに変更できることである。とりわけスペクトルの削除や追加を自由にウィンドウ間でやり取りできる機能を備えている。またスペクトル右上のアイコンをクリックすると画像のダウンロードもできる。

理研 2. モデル生物のメタボローム情報作成 (有田、福島、川島、Nilmini)

かずさグループが整理した 136 の生物種から代謝物リストを抽出し、系統関係に合わせた代謝物リストを作成した。また、理化学研究所 CSRS で計測したメタボローム情報のメタデータは本プロジェクトで作成した Metabolonote に集約している²³⁾。

ただ、メタボローム測定による代謝物のリストとゲノムに基づく代謝物の推定リスト、そして文献でレポートされる代謝物は殆ど一致しないこともわかった。(例えばイネの場合、ゲノムから推定される 1550 物質のうち、メタボロームで同定されるのは 100 程度、文献で報告されるのは 20 程度である。)今後この不一致を解消するべくデータのキュレーションやアノテーション率の向上を目指す必要がある。

理研 3. 代謝物予測ソフトウェアの構築 (有田、津川)

母核構造を一切指定せずに直接化合物を予測、候補をランキングできるソフトウェアシステム MS-FINDER を構築した²⁷⁾。このシステム作成に伴い、化合物の組成式リストおよび InChIKey 一覧表も作成した。

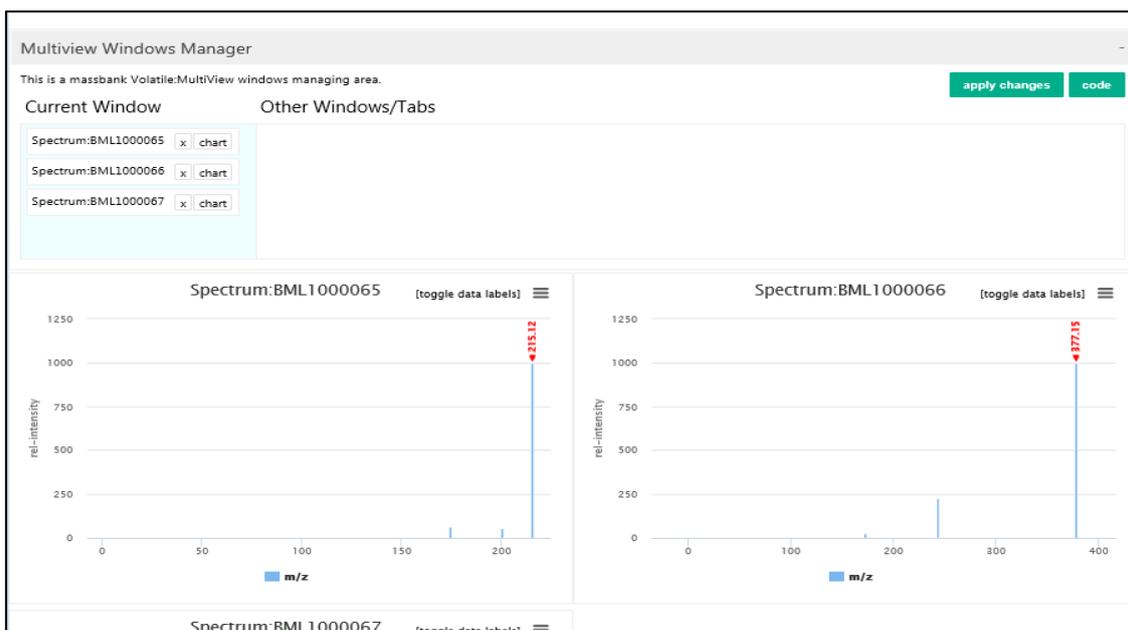


図1. MassBank wiki のスペクトル表示および Multiple Windows Manager

奈良グループ

生物種と代謝物の関係データベース KNApSAcK の充実 (金谷、他)

標準的かつ人間が読める方法で分子情報を提供する目的で、生物種と代謝物の関係データベースに含まれる全代謝物 50899 件について InChI(International Chemical Identifier) キーを付加した。これにより、理研グループとかずさグループとの連携を強化し、欧米のデータベースとの共通部分も計算できるようになった。

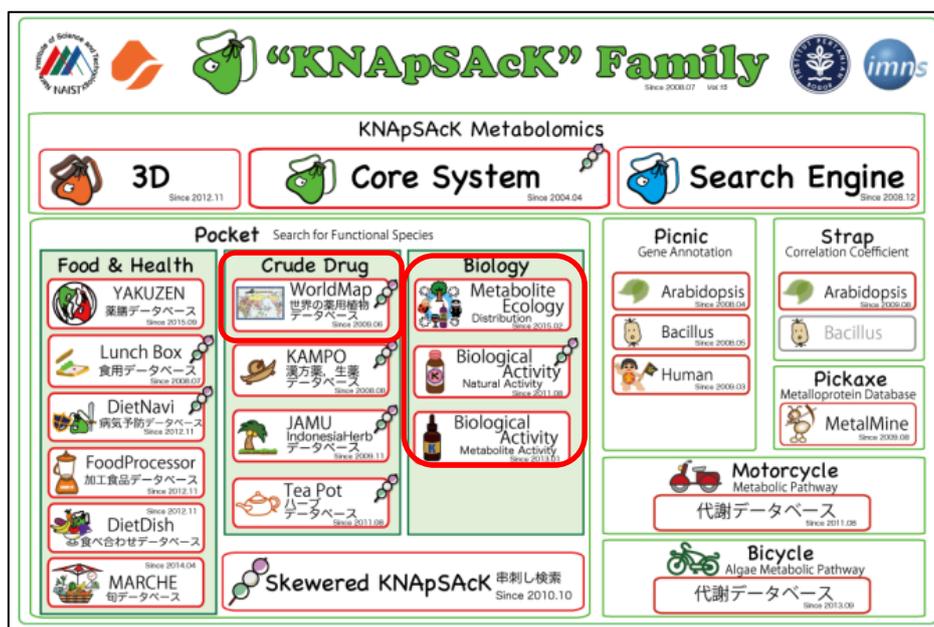


図2. KNApSAcK ファミリーDB のトップページ。赤枠内が本プロジェクトによるもの。

利用国情報の追加

食用と薬用生物に関して、世界の 222 地域(国と島)ごとの利用情報を生物-利用国の組み合わせで 56275 件(9924 件、地域-食用植物、46351 件、地域-薬用植物)整理し、生物種-代謝物関係データベースへのリンク付けをおこなった¹⁵⁾。

代謝物の生物活性情報の追加

KNApSAcK Metabolite Activity DBとして、各代謝物の生物活性を 10758 件登録した。人に対する生理活性のみならず、生体間の相互作用の表現を 140 カテゴリーに階層分けして整理した。登録された 3 千の代謝物(約 800 生物種)とその生物活性の関係から、個々の物質ではなく代謝経路の生物活性としての役割を説明することができた^{1,9)}。この成果は、2014 年、Molecular Informatics において 2015 年に論文賞を受賞した(69 報のうち最優秀)。

エコロジー情報の追加

環境・生態に及ぶ広範なメタボロミクス研究に対応すべく二次代謝物による生物間の関係性を体系化する目的で、データベース Metabolite Ecology DB の設計を行い、データの蓄積を進めた⁶⁾。二次代謝物を、大気中(VOC)、その他、植物体の葉、根、根茎という4種の分布様式に分類した。また、それぞれの代謝物によるエコロジーとして抗カビ、抗菌、誘引、防御、成長促進、成長阻害、忌避などの生物活性に注目し、8285 組の代謝物、代謝物質の生態学的活性とその対象生物という3組のデータを実際に Metabolite Ecology DB に格納し、公開を進めた。また生物種ならびに二次代謝物をキーとした遺伝子ならびに代謝経路に関する情報の蓄積の充実を図っている。

かずさグループ

第一期では、測定生データを大量に公開する基盤として、レポジトリである MassBase、測定条件を管理するデータベース Metabolonote、公開のためのフォーマット TogoMD を整備した²³⁾。第二期は、生データから高品質な代謝物ピークを抽出し、大規模に公開し、活用するため、1) 高精度ピーク抽出ソフトと品質確認ツールの開発、2) アノテーションの向上、3) 公開をおこなった。

かずさ1. 大規模ピーク抽出ソフト・ツールの開発(櫻井)

LC-精密質量MSの生データ(汎用のmzXMLフォーマット)からピークを抽出してアノテーションを行い、TogoMDファイルで出力を行うための、バッチ処理ソフトウェア PowerGetBatchを開発した。質量精度とアダクト判定精度を向上したことで、化合物データベースへのヒット率を大幅に向上させた(例:ヒトの尿で約70%)。また、生データやピーク情報の品質確認ツールを開発し、公開した(MassChroViewer, Sakurai et al., submitted)。

かずさ2. アノテーションの向上(櫻井、秋元)

機能性成分としても関心の高い植物由来の二次代謝物であるフラボノイド類(約7000種類)について、MSスペクトルから構造を推定するシステム FlavonoidSearchを開発した(Akimoto et al., submitted)。化学構造から理論的に予測されるMSスペクトルを、膨大な手作業でデータベース化したことで、植物由来のフラボノイド同定数を大幅に向上できた。

質量値から化合物データベースを検索すると、荷電分子や複合構造(塩など)の登録により、偽陽性が最大20%も生じる。単一構造を取りだし中性化した化合物データベース N2Dを開発・公開したことで、偽陽性の排除に成功した(Sakurai et al., submitted)。

以上の開発と、データベース検索結果を集計するモジュールを PowerGetBatch に組み込むことで、質の高いアノテーションを付与できるようになった。

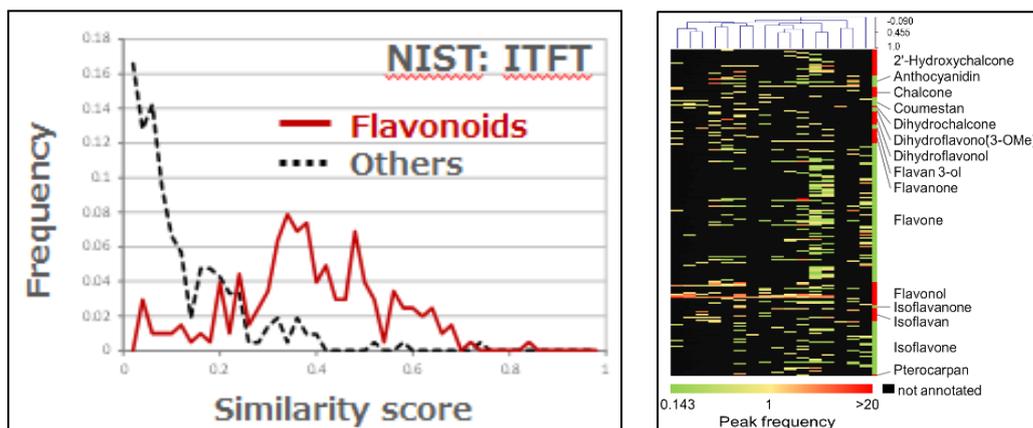


図3. フラボノイド同定スコアの適正度を示すグラフと、植物各種サンプルにおけるフラボノイド。

かずさ3. データの公開(櫻井、山田、池田、印出)

第二期開始以降、86の陸上植物、32の真菌を含む141生物種について、アノテーションを付与したデータを PowerGetBatch で構築し、TogoMDフォーマットで KomicMarket2 より公開した(2308件の分析から抽出された1,849,135ピークを含む)。Metabolonoteに追加したメタデータは、総サンプル数1114件、分析データ2709件、解析データ3079件にのぼる。

Metabolonote が管理する測定条件の情報を SPARQL で検索できるよう、オントロジーの整備とRDF化を行い、63405トリプルのRDFを構築して、NBDC RDFPortalに提供した。現在RDFPortalからの公開準備を進めている。

(2) データベースの利便性に関する利用者ニーズと具体的な対応

- 海外のデータベース作成者と情報交換するため、各種データベース(MassBank, KNApSAcK, MS-FINDER)に分子構造のハッシュキーであるInChIKeyを付与する作業をおこなった。この作業を通して構造情報の整理を簡便に実施できるようになった。
- 同様に、スペクトルのハッシュキーである SPLASH を海外と共同開発し、MassBank 等に実装した。この作業を通して、複数サーバ間でスペクトルをシェアしても、簡単に重複を判定できるようになった。
- MassBank やアノテーションソフトウェアの利用講習会を、バイオインフォマティクス学会や

DBCLS と連携して実施した。

- 統合データベース講習会 AJACS で 4 回の講習会(第二期)を行い、メタボローム解析や化合物データベースのユーザの要望を収集し、データベースやツール開発に取り入れた。

(3) 持続的なデータベース運用体制の構築に向けた取り組み

- 各研究グループそれぞれが類似したサーバを設置するのは、維持管理の手間とコストを増やす。しかし成果をどこかに集約する作業は各研究機関の方針と相容れない。そのため、本研究ではデータ管理の役割分担を行なった。大まかには、理研グループが海外とのデータ共有、奈良先グループが代謝物とその効能情報、かずさグループがメタボロームデータの収集と管理である。そのため、理研で計測したデータのメタデータをかずさの Metabolonote に登録、かずさで計測したメタボロームデータの結果を MassBank wiki に登録し、相互の情報流通を図った。

(4) 統合化推進プログラムの他のチームや DBCLS との連携

DBCLS が開催するスパークルソンを隔月で共催し、統合化推進プログラムの複数チームとともにスパークルソンを通じた情報交換、共有をおこなった。

(5) データ産出を行う研究組織や研究室、プロジェクトとの連携

以下の研究組織、学協会と連携するかたちでプロジェクトを実行してきた。

日本質量分析学会 … 旧 MassBank を公式サーバとする学術団体。その機関誌に掲載されるマススペクトルは MassBank に登録することが推奨されている。

学会ウェブサイト: http://www.mssj.jp/about/society/society_massbank.html

日本脂質生化学会 … LipidBank を公式サーバとする学術団体で、LipidBank に収載される脂質構造の内、スペクトルが利用可能なものには MassBank へのリンクを張っている。

学会ウェブサイト: <http://jcbl.jp/wiki/>

質量分析インフォマティクス研究会 … 日本バイオインフォマティクス学会の公募研究会にあたり、2016 年 4 月に発足した。2016 年度は 2 回開催し、その中で MassBank の紹介や本プロジェクトで開発したソフトウェアの紹介を行なっている。また 2016 年度は日本バイオインフォマティクスの年会でも質量インフォマティクスのワークショップを主催した。

研究会ウェブサイト: <http://ms-bio.info/aim.html>

国際メタボロミクス学会 … メタボロミクス研究者に向けてデータの標準化や国際連携を進めている。MassBank を含むデータベースを紹介するワークショップを毎年開催している。

(6) 人材の育成

今後も質量分析インフォマティクス研究会等を通じて、マススペクトルに携わる研究者数を増やす予定である。

(7) その他

- 生物種-代謝物関係データベース KNApSAcK の論文が、2015 年にトムソン・ロイターにより、本分野の上位 5%論文として選定された。

§4. 主要なデータベースの利活用状況

1. アクセス数

(1) 実績

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
MassBank (wiki版は一般公開せず)	訪問者数	160,365	154,636	158,805	153,855
	訪問数	295,392	261,693	360,951	321,198
	ページ数	4,124,087	4,605,583	5,977,294	4,632,951
KNApSAcK	訪問者数	50,645	51,024	52,942	53,731
	訪問数	125,143	102,921	113,847	103,073
	ページ数	1,277,674	1,045,392	991,412	986,174
KomicMarket2	訪問者数	724	688	801	874
	訪問数	1,566	1,450	1,522	2,307
	ページ数	14,832	10,402	9,373	51,159

表 1-1 研究開発対象の主要なデータベースの利用状況 (年度別)

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
MassBank (wiki版は一般公開せず)	訪問者数	13,364	12,886	13,234	17,095
	訪問数	24,616	21,808	30,079	35,689
	ページ数	343,674	383,799	498,108	514,772
KNApSAcK	訪問者数	4,220	4,252	4,412	5,970
	訪問数	10,429	8,577	9,487	11,453
	ページ数	106,473	87,116	82,618	109,575
KomicMarket2	訪問者数	60	57	67	97
	訪問数	131	121	127	256
	ページ数	1,236	867	781	5,684

表 1-2 研究開発対象の主要なデータベースの利用状況 (月間平均)

(2) 分析

- MassBank データベースの訪問者数は 2013 年から毎月 10,000~15,000 で安定しており、年間訪問者数 15 万(毎月の訪問者数の 12 ヶ月分合計であり、重複を含む)である。2015 年のみページ数が多くなっているが、これはチリからのバルクダウンロードがおこなわれたため、1 回限りの大量アクセスになっている。
- 2014 年度から MassBank の主要サービスであったバッチ検索機能や部分構造検索機能を廃止しているが、これらのサービスを停止した影響は殆ど無いことがわかる。訪問時間が 5 分を超えるアクセスは全体の 15% 程度ある。
- 本プロジェクトで作成する MassBank wiki はまだ一般公開していないため、本統計は旧 MassBank のものである。
- KNApSAcK データベースの訪問者数も安定している。訪問者数と訪問数、閲覧ページ数の関係は MassBank とほぼ同じであり、ユーザあたりの閲覧スタイルが似ていることを示す。本プロジェクトで作成する Biological Activity や Metabolite Ecology は同じサーバ上で稼働しているため、この統計にすべて含まれている。
- KomicMarket2 データベースは、TogoMD フォーマットに移行する前の旧 KomicMarket と、移行途中の一時サイト(New KomicMarket Tmp および KomicMarket2 Tmp)を合計した

ものである。2016年に訪問数、訪問者数、ページ数ともに増加しており、公開データへの関心が高まっていることがうかがえる。国内・海外共にアクセスが増加していたが、特に国内からのアクセスが伸びており、AJACS講習会の効果などが影響しているかもしれない。

参考：KomicMarket, MassBase, Metabolonote を足したものは以下のとおりになる。MassBaseからは4テラバイトを超えるダウンロードもあったため、ページ数や訪問数が多くなっている。

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
KomicMarket, MassBase, Metabolonote	訪問者数	16,825	17,942	14,499	11,554
	訪問数	42,602	30,759	29,809	31,914
	ページ数	270,479	152,865	135,165	236,860

表 2-1. 利用状況の合計値(年度別)

名称	種別	平成 25(2013) 年度	平成 26(2014) 年度	平成 27(2015) 年度	平成 28(2016) 年度
KomicMarket, MassBase, Metabolonote	訪問者数	1,402	1,495	1,208	1,284
	訪問数	3,550	2,563	2,484	3,546
	ページ数	22,540	12,739	11,264	26,318

表 2-2. 利用状況の合計値(月間平均)

2. データベースを利用して得られた研究成果事例

- Nature.com のサイトでキーワード検索すると、過去二年間に MassBank を引用している論文数は 42 報(レビュー、ニュース、共同研究グループの論文含む)あった。
<http://www.nature.com/search?q=massbank>
- UC Davis 大学の Fiehn 教授および、世界中のリポジトリ管理者との共同研究において、MassBank データを活用した。他のマススペクトルデータベースと情報を統合し、スペクトルから代謝物構造を予測するソフトウェア MS-FINDER や、スペクトルの ID である SPLASH の開発につながった。
- KNApSAcK family DB に関わる論文の被引用件数は、現在、270 を越え、被引用論文として、Nature 関連誌に 3 回引用されるに至っている。また、KNApSAcK family DB の活用事例として、メタボロミクス、ネットワーク・バイオロジー、新規データベース開発における活用と多岐にわたっている。
 1. 料理レシピの栄養学ならびにヘルスケアの観点から医学・栄養学により本データベースの活用がなされた。
関連：ビッグデータ・フード・サイエンス：料理レシピと世界の食品アクセスにおけるデータ・サイエンス、明日の食品産業 2016 年 3 月号。
 2. InsectInDB(<http://insect-plant.org/insectindb>)においてチョウと食草の関係を二次代謝物により体系化を目指したデータベースであり、食草と二次代謝物の関係に関する情報は KNApSAcK Core DB が活用された。
関連：文献 38
 3. International Movement and Nutrition Society をドイツレーゲンスブルク大学と主宰し、研究員1名が現在もレーゲンスブルク大学で研究を遂行。2015 年より毎年国際集會も開催している。
<http://www.uni-regensburg.de/psychologie-paedagogik-sport/psychologie-lange/international-movement-nutrition-society/index.html>
 4. 統合化推進プログラムで前期に開発した高速に組成式演算や化合物データベース検索を行う MFSearcher ウェブサービスが、ワーゲニンゲン大学(オランダ)のメタボローム

解析用 Galaxy システムに組み込まれ、利用されるようになった。

関連: Boekel J, Chilton JM, Cooke IR, Horvatovich PL, Jagtap PD, Käll L, Lehtiö J, Lukasse P, Moerland PD and Griffin TJ, “Multi-omic data analysis using Galaxy”, Nature Biotechnology vol. 33, pp.137-139, 2015 (DOI:10.1038/nbt.3134)

3. その他

コンソーシアムと MassBank of North America (MONA) の発足

UC Davis において MassBank を含む複数サーバのデータを統合したデータベース MONA が始動した。 <http://mona.fiehnlab.ucdavis.edu/>

これによりスペクトルデータ全体をヨーロッパ Norman MassBank のみならず米国にもダウンロード形式で配置し、メインサーバにアクセスしなくても利用できる状態になった。 SPLASH 論文(文献 28)に記載したように、そのスペクトルは GNPS や HMDB といった他のスペクトルサーバからも (MassBank ID ではなく、それぞれのデータベース ID を通して) 利用されている。そのため、MassBank 単独のアクセス統計や利用指標では、利用の実態を把握できなくなった。

KNApSAcK の引用数と評価

本申請ではデータの充実を図った KNApSAcK family DB のアクセス件数は、1年あたり 3,492,931 件(2016年)であり、世界からアクセスされるに至っている。学術的には、たんぱく質におけるリガンド探索、リガンドとしての代謝物の構造類似性研究、さらには、昆虫-食草関係における二次代謝物による解析など、国内外の研究者に活用されるに至っている。本 DB の充実を図り、かつ、検索効率を上げる努力を進めた結果、KNApSAcK family DBG を過去に紹介した論文が、トムソン・ロイター社において上位 1% 以内の被引用論文として評価された。このことは、本 DB が世界のメタボローム研究ならびに周辺研究者が活用できる世界標準として認知されつつあることを示しており、さらなる充実を図る計画を進めているところである。

§5. 研究開発期間中に得られた科学・技術や産業に対する波及効果

産業界への影響は特に無し。

§6. 今後の展開

MassBank wiki の整備と旧 MassBank からの移行

平成 28 年 3 月から日本質量分析学会の持つ旧 MassBank から、新規に作成した MassBank wiki へのリンクを貼る形で運用を開始し、1~2 年をかけて順次ユーザを MassBank wiki に移す予定である。同時に質量分析学会と折衝を進め、学会側にも維持管理の分担を求めると同時に、長期的には NBDC に全てのデータを集約するように働きかける。

KNApSack, KomicMarket2, Metabolomics.JP, LipidBank 等とのさらなる連携

国内では、本プロジェクトを通じて全国の研究者と情報交換、連携できた。海外との連携も本格化したため、今後は海外の連携先を増やしていく予定である。

昆虫等、植物以外のメタボローム統合

現在のメタボローム情報は個々の植物毎であり、それ以外の生物との相互作用という観点からは整理していない。今回は、文献から生物種間のインタラクション情報を収集できた。今後はメタボローム計測においても生物種間代謝物相互作用、または循環を明らかにするよう実験デザインを組み、代謝物循環を明らかにするデータベースの構築を目指したい。

マススペクトル、代謝物の標準化

今回作成したスペクトルや代謝物リストはあくまで実測に基づいており、ノイズも除去しきれていない。今後より多くのデータを集約、比較し理論的なスペクトル、代謝物リストの作成を目指したい。

§7. 自己評価

国内外に散逸するマススペクトルデータの共有を推進するため、二つの国際標準化を実現した。一つは分子構造の InChIKey、もう一つはマススペクトルの SPLASH である。これらを通じて世界中のスペクトルデータベースがデータの重複を検出できるようになった貢献は極めて大きい。実際、MassBank データは MONA, Norman MassBank だけでなく HMDB, GNPS といったサーバからも公開され、SPLASH により一致判定ができるようになった。

アノテーションソフトウェアの開発は、PowerGet, MS-DIAL, MS-FINDER 等の成果を生み、データ処理量や精度からみても世界水準であることは間違いない。本プロジェクトの成果は世界に大きなインパクトを及ぼしている。

しかしそれらの成果を大きく発信すべき MassBank-wiki の開発が遅れてしまった。学協会との調整不足やサーバ事故も原因の一部ではあるが、研究計画が楽観的であった点は否めない。またメタボロームによる代謝物同定結果と、ゲノムによる代謝物推定結果、そして文献からの情報が殆んど一致しないことも想定外であった。

MassBank wiki サーバ公開の問題点は平成29年中に解決できるはずで、大きな問題ではないと考えている。しかしオミックスデータの齟齬は今後の研究上、大きな問題になりうる。多くの研究者がオミックス統合を掲げているが、そうした成果や知識は研究者の間ですらうまく共有されていない。今後そうした齟齬や問題点を明らかにするデータベース、つまりオミックス統合をオープンプロブレムとして提示するデータベースの構築が望まれる。

統合化推進プログラム全体へは、メタボロームという観点、および学協会との連携という観点から、重要な貢献をしたと考えている。また今回、DBCLS のスパークルソンと化推進プログラムの各研究開発課題とのリンクを作れた点も良かったと考えている。

§8. 外部発表等

1. 原著論文発表

(1) 論文数概要

種別	国内外	件数
発行済論文	国内(和文)	0 件
	国際(欧文)	37 件
未発行論文 (accepted, in press 等)	国内(和文)	0 件
	国際(欧文)	1 件

(2) 論文詳細情報

1. Nakamura Y, Afendi FM, Parvin AK, Ono N, Tanaka K, Hirai Morita A, 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(1), e7, 2014 (doi: 10.1093/pcp/pct176)
2. Tsugawa H, Kanazawa M, Ogiwara A, Arita M “MRMPROBS suite for metabolomics using large-scale MRM assays” *Bioinformatics* 30(16), 2379-2380, 2014 (doi: 10.1093/bioinformatics/btu203)
3. Tsugawa H, Ohta E, Izumi Y, Ogiwara A, Yukihiro D, Bamba T, Fukusaki E, Arita M “MRM-DIFF: data processing strategy for differential analysis in large scale MRM-based lipidomics studies” *Front Genet*, 5, 471, 2015 (doi: 10.3389/fgene.2014.00471)
4. Fukushima A, Kanaya S, Nishida K “Integrated network analysis and effective tools in plant systems biology” *Frontiers in Plant Sci.*, 5, 1-9, 2014 (doi:10.3389/fpls.2014.00598)
5. Amin MA, Katsuragi T, Sato T, Ono N, Kanaya S “An Unsupervised Approach to Predict Functional Relations between Genes Based on Expression Data, *BioMed Res. International* 154594, 1-8, 2014 (doi: 10.1155/2014/154594)
6. Amin MA, Afendi FM, Kiboi SK, Kanaya S “Systems Biology in the Context of Big Data and Networks” *BioMed Res. International*, 428570, 1-11, 2014 (doi: 10.1155/2014/428570)
7. Kibinge N, Ikeda S, Ono N, Amin MA, Shigehiko Kanaya S “Integration of Residue Attributes for Sequence Diversity Characterization of Terpenoid Enzymes” *BioMed Res. International*, 753428, 1-10, 2014 (doi: 10.1155/2014/753428)
8. 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 Res. International*, 194812, 1-13, 2014 (doi: 10.1155/2014/194812)
9. Ohtana Y, Abdullah AA, Amin MA, Huang M, Ono N, Sato T, Sugiura T, Horai H, Nakamura Y, Morita AH, Lange KW, Kibinge NK, Katsuragi T, Shirai T, Kanaya S “Clustering of 3D-Structure Similarity Based Network of Secondary Metabolites Reveals Their Relationships with Biological Activities” *Mol. Inf.* 33, 790-801, 2014 (doi: 10.1002/minf.201400123)
10. Maruyama K, Urano K, Yoshiwara K, Morishita Y, Sakurai N, Suzuki H, Kojima M, Sakakibara H, Shibata D, Saito K, Shinozaki K, Yamaguchi-Shinozaki K “Integrated analysis of the effects of cold and dehydration on rice metabolites, phytohormones, and gene transcripts” *Plant Physiol.*, 164(4), 1759-1771, 2014 (doi: 10.1104/pp.113.231720)
11. Mano J, Khorobrykh S, Matsui K, Iijima Y, Sakurai N, Suzuki H, Shibata D “Acrolein is formed from trienoic fatty acids in chloroplast: A targeted

- metabolomics approach” *Plant Biotechnology* 31(5):535-543, 2014 (doi: 10.5511/plantbiotechnology.14.1112a).
12. Mannen K, Matsumoto T, Takahashi S, Yamaguchi Y, Tsukagoshi M, Sano R, Suzuki H, Sakurai N, Shibata D, Koyama T, Nakayama T “Coordinated transcriptional regulation of isopentenyl diphosphate biosynthetic pathway enzymes in plastids by phytochrome-interacting factor 5” *Biochemical and biophysical research communications* 443(2):768-774, 2014 (doi: 10.1016/j.bbrc.2013.12.040)
 13. Kera K, Ogata Y, Ara T, Nagashima Y, Shimada N, Sakurai N, Shibata D, Suzuki H “ShiftedIonsFinder: A standalone Java tool for finding peaks with specified mass differences by comparing mass spectra of isotope-labeled and unlabeled data sets” *Plant Biotechnology* 31(3):269-274, 2014 (doi: 10.5511/plantbiotechnology.14.0609c)
 14. Aoki Y, Takahashi S, Takayama D, Ogata Y, Sakurai N, Suzuki H, Asawatreratanakul K, Wititsuwannakul D, Wititsuwannakul R, Shibata D, Koyama T, Nakayama T “Identification of laticifer-specific genes and their promoter regions from a natural rubber producing plant *Hevea brasiliensis*” *Plant science* 225:1-8, 2014 (doi: 10.1016/j.plantsci.2014.05.003)
 15. Okada T, Afendi FM, Yamazaki M, Chida KN, Suzuki M, Kawai R, Kim M, Namiki T, Kanaya S, Saito K, Informatics framework of traditional Sino-Japanese medicine (Kampo) unveiled by factor analysis, *J Nat Med.* 70, 107-114, 2016 doi 10.1007/s11418-015-0946-0
 16. Lange KW, Hauser J, Nakamura Y, Kanaya S, “Dietary seaweeds and obesity” *Food Science and Human Wellness*, 4, 87-96, 2015 doi:10.1016/j.fshw.2015.08.001
 17. Abdullah AA, Altaf-Ul-Amin M, Ono N, Sato T, Sugiura T, Morita AH, Katsuragi T, Muto A, Nishioka T, Kanaya S “Development and mining of a volatile organic compound database” *Biomed Research International* 2015:139254, 2015 doi: 10.1155/2015/139254
 18. Hirose Y, Suda K, Liu YG, Sato S, Nakamura Y, Yokoyama K, Yamamoto N, Hanano S, Takita E, Sakurai N, Suzuki H, Nakamura Y, Kaneko T, Yano K, Tabata S, Shibata D “The Arabidopsis TAC Position Viewer: a high-resolution map of transformation-competent artificial chromosome (TAC) clones aligned with the Arabidopsis thaliana Columbia-0 genome” *Plant Journal* 83(6), 1114-1122, 2015 doi: 10.1111/tpj.12949
 19. Tanaka K, Arita M, Sakurai H, Ono N, Tezuka Y "Analysis of Chemical Properties of Edible and Medicinal Ginger by Metabolomics Approach" *Biomed Research International* 2015:671058, 2015 doi: 10.1155/2015/671058
 20. Suzuki M, Nakabayashi R, Ogata Y, Sakurai N, Tokimatsu T, Goto S, Suzuki M, Jasinski M, Martinoia E, Otagaki S, Matsumoto S, Saito K, Shiratake K. “Multiomics in grape berry skin revealed specific induction of the stilbene synthetic pathway by ultraviolet-C irradiation” *Plant Physiology*, 168(1), 47-59, 2015 doi: 10.1104/pp.114.254375
 21. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, VanderGheynst J, Fiehn O, Arita M "MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis" *Nature Methods* 12(6), 523-526, 2015 doi: 10.1038/nmeth.3393
 22. Li D, Ono N, Sato T, Sugiura T, Altaf-Ul-Amin M, Ohta D, Suzuki H, Arita M, Tanaka K, Ma Z, Kanaya S "Targeted Integration of RNA-Seq and Metabolite Data to Elucidate Curcuminoid Biosynthesis in Four Curcuma Species" *Plant Cell Physiology*, 56(5), 843-851, 2015 doi: 10.1093/pcp/pcv008
 23. Ara T, Enomoto M, Arita M, Ikeda C, Kera K, Yamada M, Nishioka T, Ikeda T,

- Nihei Y, Shibata D, Kanaya S, Sakurai N "Metabolonote: a wiki-based database for managing hierarchical metadata of metabolome analyses" *Frontiers in Bioengineering and Biotechnology*, 3:38, 2015 doi: 10.3389/fbioe.2015.00038
24. Tanaka K, Arita M, Li D, Ono N, Tezuka Y, Kanaya S "Metabolomic Characterization of a Low Phytic Acid and High Anti-oxidative Cultivar of Turmeric" *Natural Product Communications*, 10(2), 329-334, 2015
 25. Maravi DK, Kumar S, Sharma PK, Kobayashi Y, Goud VV, Sakurai N, Koyama H, Sahoo L "Ectopic expression of AtDGAT1, encoding diacylglycerol O-acyltransferase exclusively committed to TAG biosynthesis, enhances oil accumulation in seeds and leaves of *Jatropha*" *Biotechnol Biofuels*, 9:226, 2016 doi: DOI: 10.1186/s13068-016-0642-7
 26. Sriyudthsak K, Mejia RF, Arita M, Hirai MY "PASMmet: a web-based platform for prediction, modelling and analyses of metabolic systems" *Nucleic Acids Res*, 44(W1):W205-11, 2016 doi: 10.1093/nar/gkw415
 27. Tsugawa H, Kind T, Nakabayashi R, Yukihiro D, Tanaka W, Cajka T, Saito K, Fiehn O, Arita M "Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software" *Anal Chem*, 88(16):7946-58, 2016 doi: 10.1021/acs.analchem.6b00770
 28. Wohlgemuth G, Mehta SS, Mejia RF, Neumann S, Pedrosa D, Pluskal T, Schymanski EL, Willighagen EL, Wilson M, Wishart DS, Arita M, Dorrestein PC, Bandeira N, Wang M, Schulze T, Salek RM, Steinbeck C, Nainala VC, Mistrik R, Nishioka T, Fiehn O "SPLASH, a hashed identifier for mass spectra" *Nat Biotechnol*. 34(11):1099-1101, 2016 doi: 10.1038/nbt.3689
 29. Sakamori K, Ono N, Ihara M, Suzuki H, Matsuura H, Tanaka K, Ohta D, Kanaya S, Matsuda K "Selective regulation of pyrethrin biosynthesis by the specific blend of wound induced volatiles in *Tanacetum cinerariifolium*" *Plant Signal Behav*. 11(4):e1149675, 2016 doi: 10.1080/15592324.2016.1149675.
 30. Kibinge N, Ono N, Horie M, Sato T, Sugiura T, Altaf-Ul-Amin M, Saito A, Kanaya S "Integrated pathway-based transcription regulation network mining and visualization based on gene expression profiles" *J Biomed Inform*, 61:194-202, 2016 doi: 10.1016/j.jbi.2016.04.002.
 31. Okada T, Takahashi H, Suzuki Y, Sugano S, Noji M, Kenmoku H, Toyota M, Kanaya S, Kawahara N, Asakawa Y, Sekita S "Comparative analysis of transcriptomes in aerial stems and roots of *Ephedra sinica* based on high-throughput mRNA sequencing" *Genom Data*, 10:4-11, 2016 doi: 10.1016/j.gdata.2016.08.003
 32. Wada Y, Wada K, Iwasaki Y, Kanaya S, Ikemura T. "Directional and reoccurring sequence change in zoonotic RNA virus genomes visualized by time-series word count" *Sci Rep*, 6:36197, 2016 doi: 10.1038/srep36197
 33. Wijaya SH, Afendi FM, Batubara I, Darusman LK, Altaf-Ul-Amin M, Kanaya S "Finding an appropriate equation to measure similarity between binary vectors: case studies on Indonesian and Japanese herbal medicines" *BMC Bioinformatics*, 17(1):520, 2016 doi: DOI: 10.1186/s12859-016-1392-z
 34. Liu, K, Abdullah, AA., , Huang, M, Altaf-Ul-Amin, M, Kanaya, S, "Novel Approach to Classify Plants Based on Metabolite-Content Similarity", *BioMed Research International*, 2017 (2017), Article ID 5296729, 12 pages, doi.org/10.1155/2017/5296729
 35. Yamamoto N, Ara T, Sakurai N, Suzuki H, Shibata D, Tsugane T "An improved MatchedIonsFinder algorithm for refining ion feature assignments among chromatograms in liquid chromatography-mass spectrometry" *Journal of Life Sciences and Technologies*. 4(2):75-8, 2016 doi: 10.18178/jolst.4.2.75-78.

36. Maravi DK, Kumar S, Sharma PK, Kobayashi Y, Goud VV, Sakurai N, Koyama H, Sahoo L “Ectopic expression of AtDGAT1, encoding diacylglycerol O-acyltransferase exclusively committed to TAG biosynthesis, enhances oil accumulation in seeds and leaves of *Jatropha* ” *Biotechnology for Biofuels* 9(1):226, 2016 doi: 10.1186/s13068-016-0642-7.
37. Sato M, Sakurai N, Suzuki H, Shibata D, Kino K “Enzymatic carboxylation of hydroxystilbenes by the γ -resorcylic acid decarboxylase from *Rhizobium radiobacter* WU-0108 under reverse reaction conditions” *Journal of Molecular Catalysis B: Enzymatic*, 122:348-52, 2015 doi: 10.1016/j.molcatb.2015.10.006.
38. Muto-Fujita A, Takemoto K, Kanaya S, Nakazato T, Tokimatsu T, Matsumoto N, Kono M, Chubachi Y, Ozaki K, Kotera M, Data integration aids understanding of butterfly-host plant networks (2017) *Scientific Rep.*, accepted.

2. その他の著作物(総説、書籍など)

1. Altaf-Ul-Amin M, Katsuragi T, Sato T, Kanaya S “A Glimpse to Background and Characteristics of Major Molecular Biological Networks” *Biomedical Research International*, 2015:540297, 2015 doi: 10.1155/2015/540297
2. Kanaya S, Altaf-Ul-Amin M, Kiboi SK, Afendi FM “Big Data and Network Biology 2015” *Biomedical Research International*, 2015: 604623, 2015 doi: 10.1155/2015/604623
3. 平井(森田)晶, 中村 由紀子, 黄 銘, 佐藤 哲大, 小野 直亮, 西岡 孝明, 白井剛, 金谷 重彦, “ビッグ・データ・バイオロジー 医食同源と生態学の体系化に向けて,” *化学と生物*, vol.53, no.9, pp600-607, 2015.8
4. 櫻井 望, 有田正規 「バイオインフォマティクス入門」日本バイオインフォマティクス学会, 慶應義塾大学出版会 (2015年) 執筆分担
5. Rocca-Serra P, Salek RM, Arita M, Correa E, Dayalan S, Gonzalez-Beltran A, Ebbels T, Goodacre R, Hastings J, Haug K, Koulman A, Nikolski M, Oresic M, Sansone SA, Schober D, Smith J, Steinbeck C, Viant MR, Neumann S "Data standards can boost metabolomics research, and if there is a will, there is a way" *Metabolomics* 12(1):14, 2016
6. Carroll AJ, Salek RM, Arita M, Kopka J, Walther D “Editorial: Metabolome Informatics and Statistics: Current State and Emerging Trends” *Front Bioeng Biotechnol.* 4:63, 2016 doi: 10.3389/fbioe.2016.00063
7. Salek RM, Arita M, Dayalan S, Ebbels T, Jones AR, Neumann S, Rocca-Serra P, Viant MR, Vizcaino JA. "Embedding standards in metabolomics: the Metabolomics Society data standards task group" *Metabolomics*, 11(4), 782-783, 2015
8. 津川裕司, 有田正規 「生体内の低分子代謝産物を網羅的に捉えるための新技術」*化学と生物*, vol. 54, no.3, pp.151-153, 2016.2
9. 有田正規, 津川裕司 「化学とメタボロミクス」*化学*, vol.72, no.2, pp.26-30, 2017.2

3. 国際学会発表及び主要な国内学会発表

(1) 概要

種別	国内外	件数
招待講演	国内	26 件
	国際	3 件
口頭発表	国内	9 件
	国際	11 件
ポスター発表	国内	6 件
	国際	11 件

(2) 招待講演

〈国内〉

1. 金谷重彦, 「バイオビッグデータに挑む: 医・薬・食用植物の機能性 DB の構築とマイニング」、7 月 28 日、北海道大学理学部セミナー、札幌
2. 金谷重彦, 「ビッグデータバイオロジーとしての生態学 医食同源の体系化」、日本農芸化学会、奈良、2014 年 9 月 19 日
3. 金谷重彦, 「機能性機能食品データベース構築と活用」IT x バイオ連携促進セミナー、(株)道銀地域総合研究所、2014 年 9 月 10 日、札幌
4. 金谷重彦, 「健康に向けた食品の栄養成分データベース」、健康と長寿に貢献する遺伝学と情報学、日本遺伝学会、2014 年 9 月 20 日、長浜バイオ大学
5. 有田 正規 津川 裕司 「藻類メタボロミクス: 植物との違い」日本植物学会第 78 回大会シンポジウム「バイオリソースとゲノム情報から考える藻類研究の未来形」、東京 2014 9/12
6. 有田 正規 津川 裕司 「ノンターゲット MS/MS による藻類脂質の網羅的解析とデータベース化」健康・長寿研究談話会 (旧ホスファチジルセリン研究会)、品川 2014 11/7
7. Arita M "Impact of Metabolome Database in Plant Science" The 38th Naito Conference "Molecule-based biological systems", October 7-10, Sapporo, Japan, 2014
8. 櫻井望 「メタボロミクスにおける化合物推定」CAC フォーラムー泊研修会、湯河原 2014/10/7
9. 櫻井望 「食品メタボロミクスを拓く最先端技術」食品開発セミナー、千葉、2014/11/17, 19
10. 有田 正規 「Comprehensive lipidomics and Mass/LipidBank databases」日本分子生物・生化学会大会 (BMB2015) シンポジウム「リポドミクスから見てきた脂質の新機能 - 基礎から臨床まで」神戸, 12/4 2015
11. 有田 正規 「科学者から見た学術のオープン化」林和弘氏、有田正規氏講演会「オープンな知がイノベーションを生む - オープンサイエンスの潮流と図書館の可能性 -」(東大新図書館トークイベント 14) 東京, 10/17 2015
12. 櫻井望 「Sirius2 をつかった組成式の推定」第 63 回質量分析総合検討会, 6/17, 2015
13. 櫻井望 「かずさ DNA 研究所におけるメタボローム解析」千葉県バイオ・ライフサイエンス・ネットワーク会議平成 27 年度 総会・事例報告会, 幕張, 7/14, 2015
14. 櫻井望 「地場産業振興のためのメタボローム解析による食品分析」千葉県バイオ・ライフサイエンス・ネットワーク会議, 幕張, 1/29, 2015
15. 櫻井望 「メタボローム解析の紹介」AJACS 津軽, 弘前, 9/3, 2015
16. 櫻井望 「メタボローム解析の紹介」AJACS 薩摩, 弘前, 1/27, 2016
17. 櫻井望 「メタボローム解析の技法とデータベース構築の実際」植物バイオ研究会第 5 回会合, 東京, 2/23, 2016
18. 金谷 重彦 「ビッグデータを活用した新規ビジネスの可能性」、第 2 創業・新事業創出セミナー、東京 2015.11

19. 金谷 重彦「ビッグ・データ・バイオロジー～医食同源と生態学の体系化に向けて」、関西文化学術研究都市 8 大学連携 市民公開講座, 京都, 9/15, 2015
20. 有田 正規 津川裕司「MS2 スペクトルのフラグメンテーション解析と前駆体構造予測」日本分子生物学会年会 シンポジウム「生命システムを俯瞰するための質量分析情報解析技術とデータベースの活用」 11/30 横浜 2016
21. 秋元奈弓「経験的な開裂ルールを適用した化合物同定システムの開発」JSBi 年会 BoF 質量分析インフォマティクス 9/30 お台場 2016
22. 小野直亮「エコロジーとヘルスケアにおける代謝物データベースの構築」JSBi 年会 BoF 質量分析インフォマティクス 9/30 お台場 2016
23. 櫻井望「質量分析による化合物同定のための情報処理」第 5 回生命医薬情報学連合大会 BoF 質量分析インフォマティクス 9/30 お台場 2016
24. 櫻井望「メタボロームでここまで可能になった！」 医農連携セミナー, 京都, 6/3, 2016
25. 櫻井望「メタボローム解析の紹介」AJACS 安芸, 広島, 7/6, 2016
26. 櫻井望「メタボローム解析の紹介」AJACS 京都2, 京都, 9/1, 2016

〈国際〉

1. Arita M "Comprehensive Acquisition of MS/MS Spectra Benefits Database Research" 6th International Singapore Lipid Symposium, Singapore, December 1 (Nov30-Dec1), 2015
2. Kanaya S, "KNApSAcK Family DB: Data-driven Nutrition Science from Accessibility of Edible Species to Cuisine," 第 4 回ケモインフォマティクス秋の学校, 東京, 11/25. 2015
3. Kanaya S, "KNApSAcK Edible Species Database and Data-Driven Nutrition Science," International Conference on Movement and Nutrition in Health and Disease, ドイツ、レーゲンスブルグ、6/11. 2015

(3) 口頭講演

〈国内〉

1. 金谷重彦、オミックスプラットフォームに基づく二次代謝物情報データベース KNApSAcK Family の研究開発、第 37 回情報科学討論会、2015 年 11 月 28 日、豊橋
2. 有田正規 「メタボロームデータベースの構築と活用」 第9回メタボロームシンポジウム 三島 9 月 30 日
3. 津川裕司 「MS-DIAL プログラムを用いたリポクオリティ解析とデータベースへの連携」 第9回メタボロームシンポジウム 三島 9 月 30 日
4. 津川裕司 ほか 「マスシフト則を適用することによる自動 MS/MS 生成および未知化合物の同定戦略」9回メタボロームシンポジウム 三島 10 月 2 日
5. 福島敦史ほか 「AtMetExpress: Development of metabolite-profiling database in Arabidopsis」 9回メタボロームシンポジウム 三島 10 月 2 日
6. 櫻井望ほか「植物が生産する代謝産物の基本台帳整備」第 33 回日本植物細胞分子生物学会 東京大学 8 月 11 日
7. 秋元奈弓ほか「MSMS Flavonoid Search : MSMS スペクトルの経験的な開裂予測に基づくフラボノイド化合物アノテーションシステムの構築」第 33 回日本植物細胞分子生物学会 東京大学 8 月 11 日
8. 秋元奈弓ほか「フラボノイドを網羅的に検出する FlavonoidSearch システムの開発」日本食品化学工学会第 63 回大会, 名古屋 2016 年 8 月 27 日
9. 木越景子、櫻井望ほか「マイタケの加熱方法による成分変化」日本きのこ学会第 20 回大会, 静岡, 2016 年 9 月 8 日

〈国際〉

1. Arita M “Metabolome Database”, 1st Symposium on International Collaboration between NAIST and UC Davis: Data-driven genomics science towards opening a NAIST satellite laboratory at UC Davis. UC Davis, 9th Feb. 2015
2. Okazaki A “Metabolic pathway of carbon dioxide incorporation during wax ester fermentation in Euglena”, 1st Symposium on International Collaboration between NAIST and UC Davis: Data-driven genomics science towards opening a NAIST satellite laboratory at UC Davis. UC Davis, 9th Feb. 2015
3. Fukushima A “Development of a metabolite profiling database of Arabidopsis,” 1st Symposium on International Collaboration between NAIST and UC Davis: Data-driven genomics science towards opening a NAIST satellite laboratory at UC Davis. UC Davis, 9th Feb. 2015
4. Tsugawa H, “Untargeted metabolomics strategy via data independent MS/MS acquisition,” 1st Symposium on International Collaboration between NAIST and UC Davis: Data-driven genomics science towards opening a NAIST satellite laboratory at UC Davis. UC Davis, 9th Feb. 2015
5. Kanya S, “KNApSack Family Database towards Big Data”, PRAGMA28, Software defined e-Science, Apr,9, (2015) Nara, Japan
6. Arita M "MassBank Past, Present and Future" Norman Massbank Workshop, Duebendorf, Switzerland, 2014 9/17
7. Masanori Arita “Toward a true reference: establishing spectral standards”, “Compound identification and databases” workshop in Metabolomics2015, San Francisco, 29 June 2015
8. Atsushi Fukushima “Development of metabolite-profiling database in Arabidopsis: AtMetExpress – Meta-analysis of metabolome data” Metabolomics 2015, San Francisco, 30 June 2015
9. Hiroshi Tsugawa “MS-FINDER: Integrated Strategy for Structure Elucidation on LC-MS/MS by using Chemo- and Bioinformatics Resources” Metabolomics 2015, San Francisco, 1 July 2015
10. Wijaya SH, Tanaka Y, Katsuragi T, Morita AH, Afendi MF, Batubara I, Ono N, Altaf-UI-Amin MD, Takahashi Y, Darusman LD, Kanaya S, “Utilization of KNApSack Family Databases for Developing Herbal Medicine Systems,” Indonesia, Jakarta, PRAGMA 29, Oct. 2015
11. Li D, Ono N, Sato T, Sugiura T, Altaf-UI-Amin MD, Arita M, Tanaka K, Ma Z and Kanaya S, “Comparison of Curcuminoid Biosynthesis of Curcuma longa and its Cultivars Using a Pathway Based RNA-Seq Analysis Method,” Metabolomics, 米国、サンフランシスコ、2015, Jun. 2015

(4) ポスター発表

〈国内〉

1. 櫻井望、「実験条件の専用管理システム Metabolonote によるメタボロミクス実験データソースの統合化」、トーゴの日シンポジウム、東京、2014年10月5日
2. 有田正規ほか「代謝物リストの見本を載せるデータベース」トーゴの日シンポジウム 東京 10月5～6日
3. 櫻井望ほか「安定同位体を用いた植物代謝成分の基本台帳整備」第9回メタボロームシンポジウム 三島 10月1～2日
4. 秋元奈弓ほか「MSMS スペクトルの経験的な開裂予測に基づくフラボノイド化合物アノテーションシステム MSMS Flavonoid Search の構築と展開」第9回メタボロームシンポジウム 三島 10月1～2日

5. 櫻井望ほか「メタボロミクス技術による成分基本台帳整備と食品研究への応用」日本農芸化学会 2016 年度大会 札幌 3 月 29 日
6. 櫻井望、有田正規、金谷重彦「メタボロミクスを基盤としたエコロジーとヘルスケア・インフォマティクス」トーゴの日シンポジウム 2016、東京、2016 年 10 月 5～6 日
秋元奈弓ほか「網羅的フラボノイド同定システム FlavonoidSearch の開発」第 10 回メタボロームシンポジウム、鶴岡、2016 年 10 月 21 日

〈国際〉

1. Sakurai N, “Metabolonote: A Semantic MediaWiki-based database for metadata of metabolomics experiments”, Metabolomics 2014, Tsuruoka, Japan, 23-26th June 2014
2. Wijaya SN, Muraoka S, Andarini LA, Afendi FM, Hirai-Morita A, Latifah K, Md. Altaf-Ul-Amin, Sato T, Ono N, Sugiura T, Kanaya S “HerbsMed: Herbal medicine apps using integrated Jamu and Kampo formulas” Metabolomics 2014, Tsuruoka, Japan, 23-26th June 2014
3. Katsuragi T, Ono N, Sato T, Kanaya S “Dynamic simulation of metabolic network in Arabidopsis thaliana using parameters estimated by a genetic algorithm with modularity” Metabolomics 2014, Tsuruoka, Japan, 23-26th June 2014
4. Li D, Ono N, Sato T, Sugiura T, Md. Altaf-Ul-Amin, Arita M, Tanaka K, Ma Z, Kanaya S “Strategy of integration between RNA-seq and metabolome in curcuminoid biosynthesis pathways in cultivars of curcuma longa” Metabolomics 2014, Tsuruoka, Japan, 23-26th June 2014
5. Nakamura Y, Parvin AK, Ono N, Hirai-Morita A, Sato T, Sugiura T, Md. Altaf-Ul-Amin, Kanaya S “A new metabolomic information search using the KNApSAcK Metabolite Activity database” Metabolomics 2014, Tsuruoka, Japan, 23-26th June 2014
6. Azian Azamimi Abdullah et al. “KNApSAcK Metabolite Ecology Database for Investigating the Relationships Between VOCs and Biological Activities” Metabolomics2015, San Francisco, 29 June – 2 July
7. Donghan Li et al. “Comparison of curcuminoid biosynthesis of Curcuma longa and its cultivars using a pathway based RNA-Seq analysis method” Metabolomics2015, San Francisco, 29 June – 2 July
8. Saravanan Dayalan et al “Distributed data sharing node for mass spectra repositories in the Metabolomics Node Network” Metabolomics2015, San Francisco, 29 June – 2 July
9. Sakurai N et al., “Construction of a pipeline for the comprehensive annotation of plant metabolites by liquid chromatography-high-resolution mass spectrometry” Metabolomics2015, San Francisco, 29 June – 2 July
10. Akimoto N et al., “Flavonoid Search: Enhanced annotation of flavonoids through construction of empirical dissociation rules and a predicted mass fragment database” Metabolomics2015, San Francisco, 29 June – 2 July
11. Akimoto N et al., “FlavonoidSearch for high-throughput detection of flavonoids” 12th Annual Conference of the Metabolomics Society, Dublin, 27-30 June, 2016

4. 知財出願

なし

5. 受賞・報道等

(1) 受賞

1. 大棚 勇輝, 白井 剛, 金谷 重彦, Md. Altaf-Ul-Amin, 杉浦 忠男, 小野 直亮, 佐藤 哲大, 中村由紀子, 桂樹 哲雄、「植物における二次代謝物の効能と立体構造の関係に関する研究」第 37 回情報科学討論会、11 月 28 日、豊橋 (ポスター賞)
2. Ohtana Y, Abdullah AA, Amin MA, Huang M, Ono N, Sato T, Sugiura T, Horai H, Nakamura Y, Morita AH, Lange KW, Kibinge NK, Katsuragi T, Shirai T, Kanaya S, “Clustering of 3D-Structure Similarity Based Network of Secondary Metabolites Reveals Their Relationships with Biological Activities,” *Mol. Inf.* **33**, 790-801 (2014) (DOI: 10.1002/minf.201400123) 2014 Molecular Informatics Best Paper Award Jan, 2015
3. Best Young Investigator In Plant Metabolomics Award
Katsuragi T, Ono N, Sato T, Kanaya S “Dynamic simulation of metabolic network in *Arabidopsis thaliana* using parameters estimated by a genetic algorithm with modularity” 10th International Conference of the Metabolomics Society, June 23-26, Tsuruoka, 2014
4. 第 63 回質量分析討論会 最優秀賞 (口頭発表部門)、津川裕司、6/19, 2015
5. 理化学研究所研究奨励賞、津川裕司、4 月, 2016
6. 2014 年 Molecular Informatics 論文賞、Ohtana Y, Abdullah AA, Altaf-Ul-Amin MD, Huang M, Ono N, Sato T, Sugiura T, Horai H, Nakazawa Y, Morita A.H, Lange K, Kibinge N, Katsuragi T, Shirai T Kanaya S 1/1, 2015
7. Tobias Kind, Hiroshi Tsugawa CASMI 2016, 1st Prize in Category 3 (Best Automatic Structural Identification)
<http://www.casmi-contest.org/2016/results.shtml>

(2) メディア報道

なし

(3) その他

なし

§9. 研究開発期間中の活動

1. 進捗ミーティング

年月日	名称	場所	参加人数	目的・概要
2014-2017年	隔月のスパークルソン	柏または三島	20~30人	研究進捗および情報交換のためのミーティング http://wiki.lifesciencedb.jp/mw/index.php/SPARQLthon

2. 主催したワークショップ、シンポジウム、アウトリーチ活動等

年月日	名称	場所	参加人数	目的・概要
2014年6月23-26日	International Metabolomics Conference	鶴岡第一ホテル、ワシントンホテル	550人	国際メタボロミクス会議を主催。データ共有のサテライトワークショップ等も開催。 http://www.metabolomics2014.org/
2015年6月29-7月2日	International Metabolomics Conference	ヒルトンサンフランシスコ	1,000人超	国際メタボロミクス会議の国際メンバー。サテライトワークショップで MassBank 等紹介。
2015年9月30-10月2日	第9回メタボロームシンポジウム	三島	300人	実行委員会メンバーとして毎年主催するイベントを三島で開催
2016年6月27-30日	International Metabolomics Conference	ダブリン国際会議場	400人	国際メタボロミクス会議の国際メンバー。サテライトワークショップで MassBank 等紹介。
2016年1月9日	1st Symposium on International Collaboration between NAIST and UC Davis	UC Davis	30人	奈良先グループの主催で UC Davis との交流シンポジウム開催。
2016年9月30日	質量分析インフォマティクス	お台場プラザ平成	50人	日本バイオインフォマティクス年会のサテライトワークショップ。
2016年12月	生物とプログラミング	山口情報芸術センター	インスタレーションのみ	ゲノムが解読された生物だけを選定して食材に用いた「ゲノム弁当」販売、その他バイオインフォマティクスの紹介

以上

別紙 研究開発対象のデータベース等

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
1	MassBank	High Resolution Mass Spectral Database	代謝物質(基礎代謝、植物二次代謝標準物質)を測定したマスペクトルのデータベース。国内外の28機関のデータをまとめて検索、閲覧が可能であり、マニュアルも充実。	http://www.massbank.jp/	2006	継続・発展	データベース等	提供済	対象外	調整中	
2	Bio-MassBank		植物の組織や微生物試料をLC-, GC-, CE-MSで分析して得られたマスペクトルを収集。未同定代謝物のマスペクトルをその化学構造を表現する化学的descriptorとして利用することによって、異なる試料間で同じあるいは類似したマスペクトルがあれば同じ未同定代謝物が含まれていると考えることができます。	http://bio.massbank.jp/	2011.12	休止・閉鎖	データベース等	対象外	対象外	対象外	
3	KNAPSAcK Core DB		文献情報をもとに生物種と代謝物の関係を収集したデータベースです。代謝物名、分子式、C_ID、CAS_ID、INCHI-KEYや関連する生物種が収録されており、それぞれをキーワードとする検索が可能です。	http://kanaya.naist.jp/knapsack_jsp/top.html	2004.4	継続・発展	データベース等	提供済	対象外	調整中	34
4	KNAPSAcK Biological Activity (Natural Products)		生体由来の代謝物質にみられる生物活性のデータベースです。生物種と健康、薬用、効能の情報とが関連づけられています。学名による検索、全体表示によるリストからのブラウズが可能です。KNAPSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086) の1つです。	http://kanaya.naist.jp/BiologicalActivity/top.jsp	2011.8	継続・発展	データベース等	調整中	対象外	調整中	1
5	KNAPSAcK DietNavi (病気予防データベース)		食品や二次代謝物と健康との繋がりを整理したデータベースです。肥満、高血圧、高脂血症などの生活習慣病を防ぐにはどのような食材や成分が有効であるかを文献情報とともに提供しています。KNAPSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086) の1つです。	http://kanaya.naist.jp/DietNavi/top.jsp	2012.11	継続・発展	データベース等	調整中	対象外	調整中	Katsuragi T, Ono N, Hirai-Morita A, Yukiko N, Amin M, Kanaya S, Cuisine omics: fundamental structures of zouni and retrortable pouched pack of curry unveiled by multivariate analysis based on food ingredients, Foods Food Ingr. 218, 43-60, (2013)
6	KNAPSAcK FoodProcessor (加工食品データベース)		加工食品の成分に関するデータベースです。日本国内で製造されているレトルトカレーの材料について、成分のオントロジー、学名などを収録しています。成分リストからの検索や、製造者名による検索が可能です。県ごとの検索もできます。KNAPSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086) の1つです。	http://kanaya.naist.jp/FoodProcessor/top.jsp	2012.11	継続・発展	データベース等	提供済	対象外	調整中	Katsuragi T, Ono N, Hirai-Morita A, Yukiko N, Amin M, Kanaya S, Cuisine omics: fundamental structures of zouni and retrortable pouched pack of curry unveiled by multivariate analysis based on food ingredients, Foods Food Ingr. 218, 43-60, (2013)
7	KNAPSAcK JAMU (IndonesiaHerb データベース)		薬草を原料としたインドネシアの伝統的な生薬JAMUに関するデータベースです。JAMUのリストから検索することで、効能や配合生薬成分とその詳細情報を、また生薬のリストから検索することでJAMUの処方とその詳細を調べることができます。各生薬の代謝物情報へのリンクも提供されています。KNAPSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086) の1つです。	http://kanaya.naist.jp/jamu/top.jsp	2009.11	継続・発展	データベース等	提供済	対象外	調整中	33
8	KNAPSAcK KAMPO		日本薬局方に基づく漢方生薬配合および生薬に関するデータベースです。漢方名検索、あるいは、生薬名(またはグループ)検索により、関連漢方の生薬配合一覧、漢方薬一覧、また効能や様々な処方についての詳細情報が調べられます。KNAPSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086) の1つです。	http://kanaya.naist.jp/kampo/top.jsp	2008.8	継続・発展	データベース等	提供済	対象外	調整中	15, 33
9	KNAPSAcK WorldMap (世界の薬用植物データベース)		薬用植物とその使用国の情報を収録したデータベースです。植物ごとに、それぞれ種名や効果、文献情報を含んでおり、国名や種名からの検索が可能です。KNAPSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086) の1つです。	http://kanaya.naist.jp/KNAPSAcK_World/top.jsp	2009.6	継続・発展	データベース等	提供済	対象外	調整中	平井、西原、富田、小島、大西、稲村、生島、佐藤、小野、黄、鈴木、中村、加藤、金谷、ビッグデータ・フード・サイエンス: 料理レシピと世界の食品アクセスにおけるデータ・サイエンス、明日の食品産業 3, 33-44, 2016
10	KNAPSAcK Lunch Box (食用データベース)		食用生物の食履歴、効能などに関するデータベースです。収録されている情報は文献情報に基づいており、食用生物の学名、分類名、原産、日本伝来、食用、健康促進/薬用、文献情報、加工品に関する情報を調べることができます。KNAPSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086) の1つです。	http://kanaya.naist.jp/LunchBox/top.jsp	2008.7	継続・発展	データベース等	提供済	対象外	調整中	

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
11	KNApSAcK Biological Activity (Metabolite Activity)		代謝物質とそれらの生物活性の関係に関するデータベースです。代謝物質名、生物活性(機能)、ターゲットとなる生物種および参考文献の情報を収録しています。各データは代謝物質名、C_ID、CAS_ID、生物活性、生物種、分子式で探索することが可能です。KNApSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086)の1つです。	http://kanaya.naist.jp/MetaboliteActivity/top.jsp	2013.1	継続・発展	データベース等	提供済	対象外	調整中	17
12	KNApSAcK Metabolite Ecology		植物の代謝物を分布様式、生物活性ごとに分類し、代謝物、代謝物質の生態学的活性とその対象生物という関係を検索できるデータベースです。代謝物名、代謝物が存在する生物・組織、カテゴリー、文献情報等を収録しています。KNApSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086)の1つです。	http://kanaya.naist.jp/MetaboliteEcology/top.jsp	2015.2	新規	データベース等	調整中	対象外	調整中	17
13	KNApSAcK Motorcycle		薬用植物における有効成分の生合成過程を把握する目的で、植物の二次代謝に関わる酵素-酵素反応をまとめたデータベースです。文献情報をもとに、酵素のアミノ酸配列と代謝酵素反応を関連づけています。反応ID、酵素名、遺伝子名、生物種によるアミノ酸配列の検索の他、BLASTPを使用したアミノ酸配列データベースとの比較も可能です。KNApSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086)の1つです。	http://kanaya.naist.jp/motorcycle/top2.html	2011.8	継続・発展	データベース等	提供済	対象外	調整中	22, Afendi FM, Okada T, Yamazaki M, Hirai-Morita A, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Amin M, Darusman LF, Saito K, Kanaya S, KNApSAcK Family Databases: Integrated Metabolite? Plant Species Databases for Multifaceted Plant Research, Plant Cell Physiol (2011) 53 (2): e1., doi.org/10.1093/pcp/pcr165
14	KNApSAcK Tea Pot (ハーブデータベース)		健康促進や薬用効果のある植物のデータベースです。学名、一般名、原産地、日本伝来、効能、参考文献の情報が収録されています。種名による検索や一覧表示からのブラウズが可能です。KNApSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086)の1つです。	http://kanaya.naist.jp/TeaPot/top.jsp	2011.8	継続・発展	データベース等	提供済	対象外	調整中	
15	KNApSAcK YAKUZEN	薬膳データベース	食材が持つ効能(性味, 帰経など)と薬膳レシピのデータベースです。食材の効能検索では、効能の他、含有成分、標準成分表、アミノ酸組成表が収録されています。また食材、効能、体質九分類から薬膳レシピが検索できます。KNApSAcK Family (http://integbio.jp/dbcatalog/record/nbdc01086)の1つです。	http://kanaya.naist.jp/YAKUZEN/top.jsp	2015.9	新規	データベース等	調整中	対象外	調整中	平井、西原、富田、小島、大西、稲村、生島、佐藤、小野、黄、鈴木、中村、加藤、金谷、ビッグデータ・フード・サイエンス: 料理レシピと世界の食品アクセスにおけるデータ・サイエンス、明日の食品産業 3, 33-44, 2016
16	体質判定・改善アプリ 体質九分類判定・改善支援ツール		体質九分類による体質判定、体質改善にオススメの食材と薬膳料理のレシピを提案するサイトです。	http://taishitsu.e-growth.net/		新規	データベース等	対象外	対象外	対象外	山森、許、渡辺、黄、小野亮、佐藤、阿部、上馬場、川端、今西、Amin、朱、戴、王、金、太田、鈴木、多変量解析法にもとづいた日本語版中医体質調査票(CCMQ-J)における各質問間相関の評価ならびにBMI、年齢の推定モデルの構築, 13, 43-56, (2016)
17	metabolomics.jp		メタボロミクスを中心とした有田研究室の活動全般を対象としたウェブサイトです。フラボノイド、基礎代謝物、生薬、植物系統分類データベースのほか、ファイトレメディエーションを含めた放射線情報、講義資料も掲載しています。	http://metabolomics.jp/wiki/Main_Page		継続・発展	データベース等	対象外	対象外	対象外	
18	Metabolonote		メタボロミクス実験の詳細な実験手法に関する情報(メタデータ)のみを専門的に取り扱うデータベースです。セマンティックMediaWikiを利用したシステムにより、ユーザー登録(無料)をすることでだれでも気軽に各自のメタデータを記録・編集することができます。メタデータを実際のデータ(生データファイルや、ピークアノテーション情報、ピークのスペクトル情報など)と切り離して管理することにより、1) 実験後すぐに、さらには実験前であっても、メタデータを記載することができるため、実験設定の詳細を忘れてしまう前に記録に留めておくことができます。また、2) 一度記録したメタデータは、論文や実際のデータを管理するその他のデータベースから共通して参照できるというメリットがあります。Metabolonoteのコアシステムは公開されているため、ユーザー独自のMetabolonoteをLIMSや公開サイトとして構築できるほか、フォーマットを独自に定義することで、メタボロミクス以外のデータ管理についても使用することが可能です。	http://metabolonote.kazusa.or.jp/	2012.10	継続・発展	データベース等	提供済	対象外	提供済	23. Ara T, Enomoto M, Arita M, Ikeda C, Kera K, Yamada M, Nishioka T, Ikeda T, Nihei Y, Shibata D, Kanaya S, Sakurai N "Metabolonote: a wiki-based database for managing hierarchical metadata of metabolome analyses" Frontiers in Bioengineering and Biotechnology, 3:38, 2015 doi: 10.3389/fbioe.2015.00038

No.	正式名称	別称	概要	URL	公開日	状態	分類	生命科学系データベースアーカイブ	NBDCヒトデータベース	NBDC RDFポータル	関連文献 (論文リストに記載があれば、その番号でも可)
19	KomicMarket	Kazusa Omics Data Market	KomicMarket (Kazusa Omics Data Market)は、代謝物を含むシステムバイオロジーやマルチオミクス解析に強く興味を持つ研究者により開発、運用されています。 KomicMarketは、LC-FT-ICR-MS法(液体クロマトグラム-フーリエ変換-イオンサイクロトロン質量分析)で検出された代謝物のピークのアノテーション(精密質量に基づく組成式や化合物名候補など)を主に格納しています。	http://webs2.kazusa.or.jp/komicmarket/	39301	継続・発展	データベース等	対象外	対象外	対象外	8. 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 Res. International, 194812, 1-13, 2014 (doi: 10.1155/2014/194812)
20	MassBase	A comprehensive mass spectral tags archive for plant metabolomics	MassBaseは代謝産物(メタボライト)に関するマスペクトラルタグ(MST)のアーカイブです。主に植物の生体サンプルやその代謝産物から得られた生のクロマトグラムデータおよびクロマトグラムデータを解析したテキストデータを収録しています。MassBankやKNAPSAcKのような他のデータベースと組み合わせて興味のある代謝を理解するだけでなく、代謝物質ピークを抽出し、代謝物質をアノテイトするパイプライン・ソフトウェアを開発するためにも活用できます。全てのデータセットがダウンロード可能です。	http://webs2.kazusa.or.jp/massbase/index.php/	2008.1	継続・発展	データベース等	公開済	対象外	対象外	8. 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 Res. International, 194812, 1-13, 2014 (doi: 10.1155/2014/194812)
21	MFSearcher	Kazusa Molecular Formula Searcher	精密質量値から予測される組成式を高速に推定することができます。既存の化合物データベースに対する質量値からの検索も高速に行えます。	http://webs2.kazusa.or.jp/mfsearcher/	2010.11	継続・発展	データベース等	対象外	対象外	対象外	Sakurai N et al. An application of a relational database system for high-throughput prediction of elemental compositions from accurate mass values. Bioinformatics. 2013;29(2):290-1. doi: 10.1093/bioinformatics/bts660.
22	MS-MS Fragment Viewer		LC-FT/ICR-MS解析で得たフラグメントイオンの構造を予測したFT-MS、IT-MS/MS、FT-MS/MSのスペクトルデータを格納したメタボロミクスデータベースです。	http://webs2.kazusa.or.jp/msmsfragmentviewer/	2008.8	継続・発展	データベース等	対象外	対象外	対象外	8. 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 Res. International, 194812, 1-13, 2014 (doi: 10.1155/2014/194812)
23	KOMICS	Kazusa Metabolomics Portal	質量分析によるメタボローム解析とメタボロームを活用したオミクス解析のためにかずさDNA研究所でこれまで開発されてきた解析ツールやデータベースを公開しているポータルサイトです。LC-精密質量MSの強力な解析ツールPowerGetや、世界最大級のデータ量を持つ分析生データのレポジトリMassBase等、ユニークなリソースが整備されています。また、メタボロミクス関係のリソースへのリンク集など、これからメタボロミクス分野へ参入する研究者にとっても役立つ情報を提供しています。	http://www.kazusa.or.jp/komics/ja/	2012.3	継続・発展	データベース等	対象外	対象外	対象外	8. 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 Res. International, 194812, 1-13, 2014 (doi: 10.1155/2014/194812)