

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)

研究開発中間報告書

「物質循環を考慮したメタボロミクス情報基盤」

有田正規 | 国立遺伝学研究所 教授



©2019 有田 正規(情報・システム研究機構) Licensed under CC BY 4.0

§1. 研究開発実施の概要

本研究は4つのグループにより構成される。

【遺伝研グループ】メタボローム・リポジトリの開設と運用

- MetaboBank リポジトリおよび入力エディタ開発
公的リポジトリ MetaboBank を遺伝研 DDBJ サービスとして開始
データ登録用のエクセルマクロ MetaboBank エディタを作成
- 食品／植物メタボローム・リポジトリの作成と公開
比較可能なように同一条件で測定した食品 222 サンプル、植物 28 サンプルのデータを容易に比較解析できるインターフェースを開発。データもダウンロード可能。

【理研グループ】植物メタボローム・アノテーション高度化

- 理研メタボロームメタデータベースの構築と公開
理化学研究所 CSRS で過去 10 年以上にわたり測定されてきた 56 プロジェクトのデータを精査し、RDF 形式でメタデータを作成。
- スペクトルライブラリの作成と公開
117 脂質クラスに対応する理論スペクトルライブラリ、オールイオンフラグメンテーション (AIF) に対応する基礎代謝物ライブラリ等を作成。
- MS-DIAL の開発
代謝物ネットワークの解析機能および AIF データの解析機能を追加
イオンモビリティ CCS データの解析機能を追加

【かずさグループ】食品メタボローム・アノテーション高度化

- 植物メタボロームデータの整備
かずさ DNA 研究所で過去 10 年以上にわたり測定されてきた 104 プロジェクトのデータを精査し、RDF 形式でメタデータを作成。
- PowerGetBatch を用いたパイプライン構築
整理したプロジェクトと紐づけされた解析生データを再解析するためのパイプライン構築

【奈良グループ】生理活性オントロジーの構築

- 代謝物情報の蓄積と活性オントロジーの整理
KNApSack データベースのコア情報を 3 千代謝物、2 万 6 千ペア作成。
生理活性オントロジーとして 140 クラス 1 万データ整理。
- メタ代謝マップの作製
アルカロイドの生合成マップを中心に 32 枚のメタ代謝マップを構築。
- 深層学習を用いた生合成開始物質予測やインシリコ・スクリーニングを実施

§2. 研究開発実施体制

1. 各グループの担当項目（主な開発スタッフおよび直接雇用者）

(1) 遺伝研グループ(研究代表者グループ) 全体統括:有田正規

MetaboBank リポジトリおよび入力エディタ開発

櫻井望, 時松敏明

食品/植物メタボローム・リポジトリの作成と公開

櫻井望

奈良グループと連携した代謝物情報整理, マップ作成

佐藤允治

理研グループと連携した MS-DIAL の開発

多田一風太

(2) 理研グループ(主たる共同研究者グループ(1))

理研メタボロームメタデータベースの構築と公開

福島敦史, 高橋みき子, 小林紀郎, 学生アルバイト

スペクトルライブラリの作成と公開

津川裕司, 高橋みき子, 長崎英樹 (2020)

MS-DIAL の開発

津川裕司, 高橋みき子

(3) かずさグループ(主たる共同研究者グループ(2))

植物メタボロームの整備

平川英樹, 櫻井望, 長崎英樹(2018-2019)

PowerGetBatch を用いたパイプライン構築

櫻井望, 長崎英樹(2018-2019)

(4) 奈良グループ(主たる共同研究者グループ(3))

代謝物情報の蓄積と活性オントロジーの整理, メタ代謝マップの作製

金谷重彦, 森田晶, 大橋美名子

深層学習を用いた生合成開始物質予測とスクリーニング

金谷重彦, 黄銘, 小野直亮, 学生アルバイト

2. 有識者会議等

(1) 会議概要

名称	DNA データ研究利用委員会
目的	委員会は、センターが実施している日本 DNA データバンク事業を推進

	<p>するため、国内の研究機関との連携、ENA/EBI（欧州）、GenBank /NCBI（米国）との国際連携協力のもとに次の各号に掲げる事項について審議する。</p> <p>一 センターの事業計画に関すること</p> <p>二 センターの関係事業経費に関すること</p> <p>三 センターの総合的な評価・調整に関すること</p> <p>四 その他 DNA データに関する調査・検討等に関すること</p>
委員数	11人(2020)

(2) 開催歴

年月日	場所	主な議題・指摘事項等
2018年 2月20日	国立遺伝学研究所 (当時の DDBJ センター長 は高木利久教授)	メタボローム・リポジトリの試験運用と題し、DDBJ サービスとして MetaboBank の運用を打診。理研 CSRS、かずさ DNA 研、遺伝研の3研究所間での「生命科学データの共有に関する覚書」締結を報告。予算制約のなかでの運営への懸念。その他、指摘事項は特になし。
2019年 2月20日	同上 (センター長 有田正規)	2018年11月のメール審議における、メタボローム・リポジトリの開設承認を確認。理研およびかずさのメタボローム・データを整理中であることを報告。公的リポジトリで欧米とのデータ交換を確認。
2020年 2月13日	同上 (センター長 有田正規)	メタボローム・データのリポジトリである MetaboBank の公開準備を報告。DDBJ ホームページで Wiki を運営する困難も説明。Wiki は学会等に変更させなければ可能ではという指摘(伊藤剛委員)。

§3. 研究開発の目的、実施内容及び成果

1. 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
MetaboBank	MetaboBank	DDBJ センター事業の一環として実施するメタボローム・データリポジトリ。Metabonote および RIKEN PMM に記載されるメタデータを含む。

上記以外のもの

正式名称	略称	概要
Metabonote	メタボノート	メタボローム研究のメタデータを MediaWiki を用いて整理したデータベース。自然言語で記述できるためフレキシブル http://metabonote.kazusa.or.jp/
MassBase	マスベース	メタボローム研究の分析生データを集積したデータベース http://webs2.kazusa.or.jp/massbase/
RIKEN Plant Metabolomics MetaDatabase	RIKEN PMM	理研 CSRS で取得した植物メタボロームデータを体系的に整理して公開するサービス。本データベースでは、メタデータの世界標準技術であるセマンティックウェブや Resource Description Framework (RDF) と呼ぶ枠組みに基づいてデータを公開している。 http://metabobank.riken.jp/
食品・植物メタボロームリポジトリ	食レポ・植レポ	食品と植物にあるメタボロームピークを検索、確認できるデータベース http://metabolites.in/foods および /plants
KNApSack	ナップザック	様々なデータを含むが、本研究ではアルカロイド等二次代謝物の生合成マップおよび植物界における二次代謝物の分布をデータベース化。

(2) ツール等

正式名称	略称	概要
MS-DIAL MS-FINDER		質量分析データ解析用ソフトウェア。MS-DIAL は代謝物由来イオンの検出、同一化合物由来イオンのグルーピング、代謝物アノテーション等の一連のデータ処理を行う。MS-FINDER は未知のスペクトルから構造解析 (structure elucidation) するためのソフトウェアツール。
PowerGetBatch		メタボロームデータ解析ツール http://www.kazusa.or.jp/komics/ja/tool-ja/235-powergetbatch.html

※データベース、ツールの詳細は別紙参照。

2. 達成目標及び実施計画

(1) 当初の実実施計画・達成目標

研究開始当初計画

【2018年度】

- ・ MassBank wiki の開始 (遺伝研)
- ・ MassBank 公共リポジトリの設計 (遺伝研)
- ・ 食品メタボローム情報のリポジトリ登録 (かずさ)
- ・ 代謝物構造と生理活性の相関解析 (奈良)
- ・ GC/MS 測定データのメタデータ化とリポジトリ登録 (理研)

【2019年度】

- ・ MassBank データのバージョン化とリポジトリ開設、スペクトルセットの頒布 (遺伝研)
- ・ Metabolonote 等を MassBank ファミリーとして統合 (遺伝研)
- ・ 植物二次代謝物のスペクトル・ライブラリ作成 (かずさ)
- ・ メタ代謝マップの MassBank wiki 連携 (奈良)
- ・ MassBank における GC/MS セクションの開始 (理研)

【2020年度】

- ・ リポジトリデータの MetabolomeXchange 登録と国際連携 (遺伝研)
- ・ リポジトリの支援体制構築 (遺伝研)
- ・ 二次代謝物ライブラリのバージョン化と、スペクトルセットの頒布 (かずさ)
- ・ メタ代謝マップを用いた植物-昆虫共生系の解析 (奈良先)
- ・ GC/MS 登録情報のバージョン化と、スペクトルセットの頒布 (理研)

(2) 期間中に追加・削除・変更した実施計画・達成目標

当初からの計画変更

2018年6月 変更点

かずさグループ代表の櫻井望が国立遺伝学研究所に移籍。かずさグループ代表を平川英樹とし、櫻井は遺伝研グループおよびかずさグループの両方に貢献。

2018年12月 変更点

スパコンレンタル費用200万円を理研グループにサーバ備品代として移管

理由

理研 PMM の物理サーバを理研情報システム本部所有のサーバより、本課題で購入する新サーバに移すため。

背景

研究開始当初は情報システム本部の所有するマシン上で稼働していたが、機器が老朽化したため。購入したサーバは理研システム本部 (和光) に設置する。

2019年4月 変更点

主な研究開発対象データベースの名称・URLを「Massbank (<http://massbank.jp>)」から「MetaboBank (<http://mb.ddbj.nig.ac.jp>)」に変更。

理由

サービス名称の MassBank およびドメイン名の Massbank.jp について、日本質量分析学会へ譲渡することとなったため。

3. 実施内容

(1) 実施内容

【遺伝研グループ】

MetaboBank リポジトリおよび入力エディタ開発

DDBJ サービスの一環として共通バナーを用いてリポジトリを2020年10月に公開した (Fig.1)。初期データとして理研およびかずさグループで整理してきた情報を収載し、塩基配列と重ならないアクセッション番号を付与している。EBI/NCBI にも開発状況は適時連絡している。一般からのデータ投稿は現在構築中の DDBJ 統

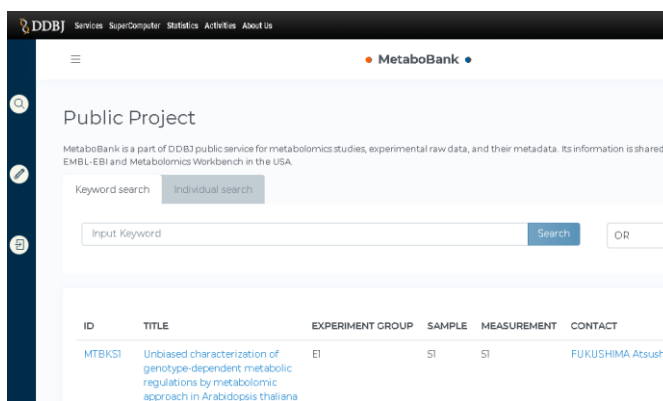


Figure 1 MetaboBank のトップページ
<http://mb.ddbj.nig.ac.jp>

合アカウントシステムを通して実施し、同アカウントからはゲノム等、他のデータ投稿も可能である。システムでは論文投稿時のレビュー用公開 URL 設定や、一般公開への切り替えといった操作をユーザ自身が設定する。一般公開を選択した場合、DDBJ の担当キュレータが内容を確認し、必要項目が正しく入力されていることを確認したうえで公開に至る。

データ登録用には「MetaboBank エディタ」と呼ぶ Microsoft Excel のマクロを用意している (Fig. 2; MetaboBank ページからダウンロード可能)。メタボローム研究のメタデータはゲノム・プロテオームに比して圧倒的に多く、対象も幅広い。ユーザが長期にわたり同じプロジェクトにデータを追加しやすくするため、入力方式は (データを長期に保持できない) ブラウザ形式ではなく、Excel を採用した。利用法として 70 ページ超におよぶエディタ利用マニュアル (日本語) も作成した。サイトへの登録はエディタか出力する JSON フォーマットを用いており、プログラムを書けるユーザに登録方法の選択肢を多く残している。また国際的なメタボローム・リポジトリ MetaboLights と MetabolomicsWorkbench との違いは、国際メタボロミクス学会の標準イニシアティブ (MSI) が推奨するガイドラインを完全に満

たしている、要するにメタデータが詳しい点である。

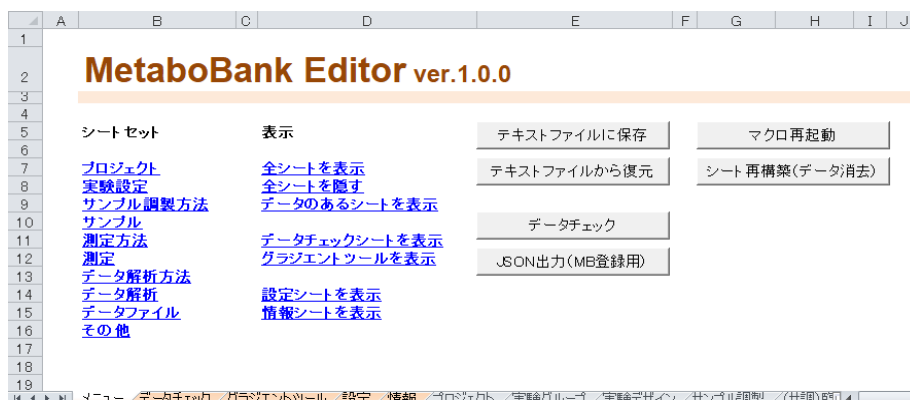


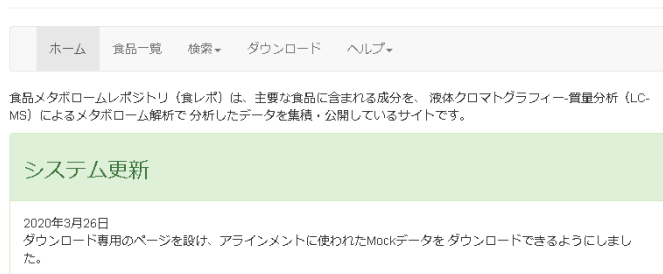
Figure 2

MetaboBank エディタのメインメニュー。日本語表示だが言語設定で英語表記にもなる。

このエディタで特徴的なのはグラジエントツールと呼ばれるクロマトグラフィー条件を GUI 形式で入力・登録させるエクセルマクロである。論文等では文字情報で書かれ概要を掴みにくい条件を視覚的に確認できる。このツールを用いた入力はフォーマットが揃うため、レポジトリ上でグラジエントの類似性を比較・検索することが可能になる。

食品/植物メタボローム・リポジトリの作成と公開

食品メタボロームレポジトリ



食レポの目的：未知成分のリファレンス

Figure 3 食品および植物メタボローム・リポジトリ

食品や植物におけるメタボロームをクロマトグラフィー結果から比較可能なように同一条件で分析し、精密質量やスペクトルで検索可能にして再解析を容易にしたサイトを作成した (Fig. 3)。食品は加工品も含めて 222 サンプル、植物は 14 種 28 サンプルにおける二次元クロマトグラムおよびスペクトルを全て公開している。このデータに含まれるスペクトルの 8 割近くは未同定である。このサイトで特定の植物種や科で測定されるピーク (スペクトル) を解析することで、代謝物を効率的に同定できる (論文リスト 13: オカラミンの発見)。またこのサイトは API も用意しているため、プログラムを用いた未知物質の同定も可能になる。例えばフラボノイドのピークに注目することでブドウに含まれるメチオニンが付加したフラボノイドや、キク科やミカン科、ハウレンソウ (アカザ科) に特徴的なフラボノイドを見つけれられている (国内招待講演 7-14)。このサイトは検索可能なリポジトリの有効性を示す好例であり MetaboBank が目指すモデルとして需要である。

【理研グループ】

理研メタボロームメタデータベースの構築と公開

植物の GC/MS データを中心にメタデータを作成、登録・公開した。RDF 形式のメタデータはウェブ国際標準規格に沿っており(論文リスト

10)、生データおよび解析済みデータ

は理研植物メタボロームメタデータベースにも格納している(Fig. 4)。2020年11月時点で、56件のメタボロームデータセットを公開し、その概要は国内招待講演等で発表した(国内招待講演2-4)。理研 PMM からは、かずさ DNA 研究所にて RDF 化した Metabolonote のデータセット 82 プロジェクトも公開し、同じデータは MetaboBank からも公開している。そのうち比較解析が可能な GC/MS データについてアップデートした MS-DIAL ソフトウェアその他を用いて再解析している。

スペクトル・ライブラリの作成と公開

代謝物の同定率を向上させるにはライブラリの充実が重要である。リポドミクスにおける同定率向上のためセラミドを中

心に量子化学計算でスペクトル解析を実施し、水酸基がβ位に配置された脂肪酸を含むセラミド等のスペクトルを明

らかにした。その後既知の構造にアシル基が付加された脂質を多数発見した。代表的なスペクトルをもとに脂肪酸の長さを変更した理論スペクトル・ライブラリを 117 脂質クラスについて作成した(Fig. 5; 論文リスト 21)。脂質クラスの分類や名称は脂質の国際コンソーシ



Figure 4 理研メタボロームメタデータベース (理研 PMM)



Figure 5 マウスの臓器に特異的な脂質分子 (理研 2020年6月プレスリリースより)

アムとともに標準化し (論文リスト 1,7), それらの存在分布をマウス臓器および培養細胞について検証して, 国際共著論文として発表した。

MS-DIAL の開発

ライブラリの構築はそれを扱えるソフトウェアプログラムと連動している。MS-DIAL プログラムは 2019 年に Version3.0, 2020 年に Version4.0 を論文発表した。それぞれ大きな開発項目は以下の通り。

MS-DIAL3.0 代謝物ネットワークの解析機能, All-Ions-Fragmentation の解析機能とそれに基づくスペクトルライブラリ構築 (論文リスト 6, 8, 9, 20)

MS-DIAL4.0 Ion-Mobility CCS の解析機能 (論文リスト 21)

【かずさグループ】

植物メタボロームデータの整備

MassBase および Metabolonote データベースに収載された 3,428 の解析生データと紐付けされた 157 生物種、104 プロジェクトのメタデータを Resource Description Framework (RDF) の書式に変更した。内訳は

植物, 微生物 : 82 プロジェクト

食品 : 16 プロジェクト

その他 : 6 プロジェクトとなる。

そのうち植物について解析生データ 2,586 個と紐付けもおこなった。

Metabolonote のメタデータは, 本プロジェクトの開始前調査研究により (2017 年度), 国際的なデータ連

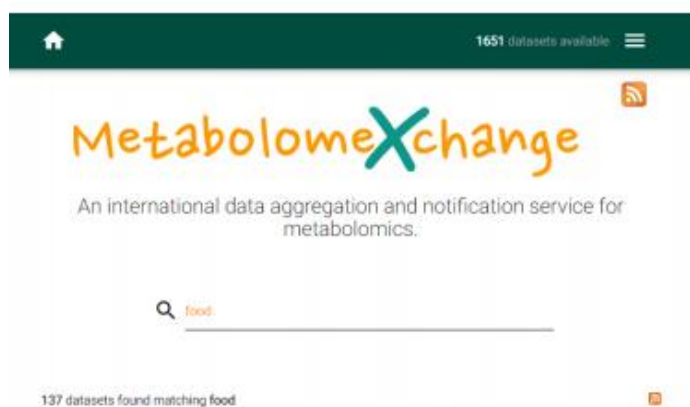


Figure 6 メタデータ登録サイト
携サイト MetabolomeXchange へ MetabolomeXchange.org

の登録が認められている (Fig. 6)。これらのデータは全て MetaboBank に収載されている。

PowerGetBatch を用いた大量処理パイプライン構築

Java で開発されたソフトウェア PowerGetBatch を使うコマンドラインを工夫し, サーバにおける大量処理パイプラインを構成した。これを用いてかずさの大量データを再解析している。(国内ポスター4,5)

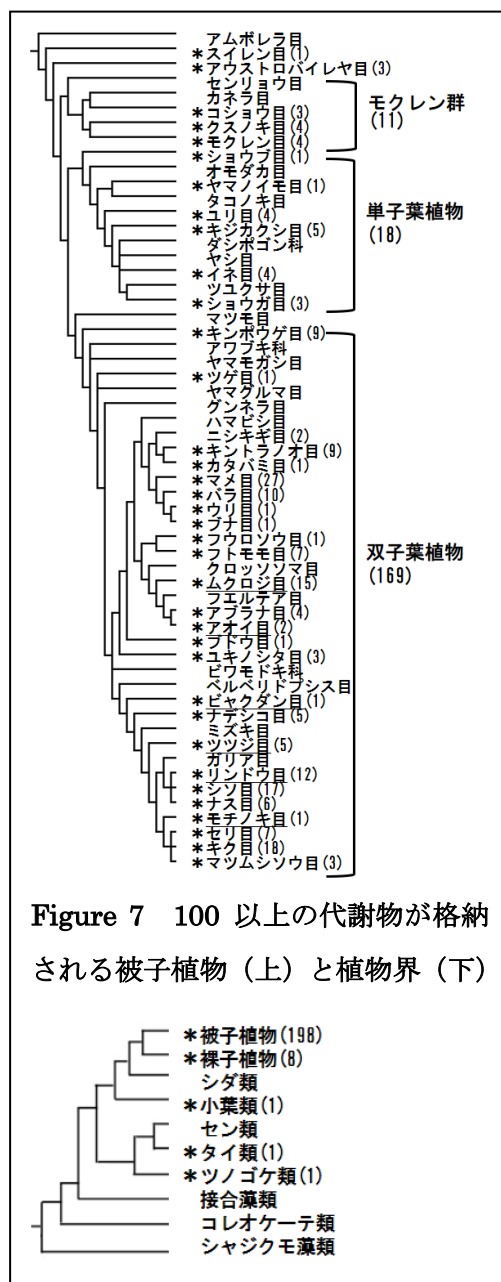
【奈良グループ】

代謝物情報の蓄積と活性オントロジーの整理 (Core DB、Metabolite Activity DB)

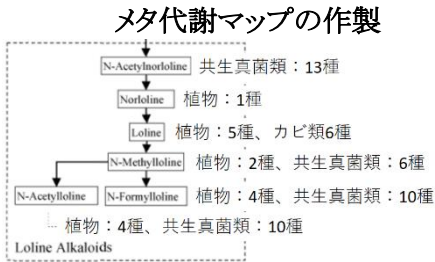
メタボローム研究の進展に伴い、様々な生物における代謝物が同定されるに至った。KNAPSAcK Core DB は、本プロジェクト開始前時点で 50,054 種の代謝物を収載し、それらを生合成する生物との関係として、102,005 組の生物種-代謝物種の関係として、102,005 組の生物種-代謝物種の間を整理してあった。本プロジェクトの開始とともに、本 DB の充実を図り、また、コロナ禍でのテレワークにおいても研究を遂行できるシステムを構築しデータの充実を図った。その結果、2020 年 10 月時点で、代謝物数は 53,032 種となった。また、生物種と代謝物の関係は、約 2 万 6 千組増えて、128,771 組となった。

メタボローム研究では、通常 50 程度の代謝物に関わる遺伝子発現、代謝経路解析がなされることが多い。植物に限ってみると、KNAPSAcK Core DB に登録された植物種について、植物におけるすべての分類目に分布するに至っており、植物メタボローム研究における中核 DB としての役割を担っている。植物メタボローム以外についても、ヒト(347 種)、海洋生物としては *Alcyonacea* 属(124 種)など 4 種、カビ類では *Aspergillus*(583 種)、*Penicillium* 属(569 種)など 6 種、細菌類では *Streptomyces* 属(1903 種)を含む 7 属について 80 以上の代謝物が蓄積されている。

これらの代謝物情報をもとに、オントロジーを作成し、活性情報の集積を、KNAPSAcK Metabolite Activity DB(<http://www.knapsackfamily.com/MetaboliteActivity/top.php>)に進めた。2020 年時点、活性種として 10852 件の代謝物と活性の関係を整理し、140



の包括的活性種情報に分類した（論文リスト 4）。



生合成経路を、生物間相互作用を含む物質循環として示したマップを作成し公開した。現在までに、67報の原著論文をもとに515種のアлкаロイドをも

とに生合成に関するメタマップ 32 枚を作製した。この情報をもとに、共生菌種と植物についての新たな関係性を取得できる。

Group of SS	β-phospho glycerate		Pyruvate		Phosphoenol pyruvate		Oxaloacetate		alpha-Ketoglutarate		Terpenes		TCA cycle		Fatty acid		Nucleic acids		
	Utp	L-Ser	D-Ser	L-Cys	D-Cys	L-Ala	L-Leu	L-Trp	L-Phe	L-Ile	L-Val	DMAAPP	Biosynthesis	OTPP	Acetyl CoA	Acetyl CoA	Methyl CoA	Acetyl CoA	Adenine
Arg, Phe, Asp, Pyrrolidine/Pyrrolidone Alkaloids	1																		
Trp, Ser, Alkaloids	2																		
Trp, Monoterpenoid Indole Alkaloids	3																		
Trp + DMAAPP, Trp + Phe, Trp + Oxaloacetate, Trp (Simple Indole Alkaloids), Trp + His	4																		
Trp + DMAAPP, Trp + Phe, Trp + Oxaloacetate, Trp (Simple Indole Alkaloids), Trp + His	5																		
Trp + DMAAPP, Trp + Phe, Trp + Oxaloacetate, Trp (Simple Indole Alkaloids), Trp + His	6																		
Trp + DMAAPP, Trp + Phe, Trp + Oxaloacetate, Trp (Simple Indole Alkaloids), Trp + His	7																		
Trp + DMAAPP, Trp + Phe, Trp + Oxaloacetate, Trp (Simple Indole Alkaloids), Trp + His	8																		
Trp + DMAAPP, Trp + Phe, Trp + Oxaloacetate, Trp (Simple Indole Alkaloids), Trp + His	9																		
Trp + DMAAPP, Trp + Phe, Trp + Oxaloacetate, Trp (Simple Indole Alkaloids), Trp + His	10																		
Trp, Seroton group, Trp + Phe, Oxidative	11																		
Trp, Isoquinoline Alkaloids (Benzylisoquinoline Alkaloids)	12																		
Tyr + Serotonia (Tetrahydroisoquinoline monoterpene alkaloids), Tyr + Ser	13																		
Tyr + Serotonia (Tetrahydroisoquinoline monoterpene alkaloids), Tyr + Ser	14																		
Care, Quinazolin, Piperidine Alkaloids, Isoindoline Alkaloids, Lycoperdin Alkaloids	15																		
Phe + Ser	16																		
Ala + Acetyl CoA + Malonyl CoA	17																		
His, Indole alkaloids/indoline derivatives	18																		
Anthranilate/Oxindole Alkaloids	19																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	20																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	21																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	22																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	23																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	24																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	25																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	26																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	27																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	28																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	29																		
Anthranilate + Trp + Anthranilate + Trp + Ala Anthranilate + Phe (PKS)	30																		
Strophanth alkaloids	31																		
Putrescine alkaloids	32																		

Figure 8 アルカロイドの生合成とアミノ酸の関係

(2) 「対象とするデータベース条件」のうち採択時に未達成であった項目の対応状況

開発対象とするデータベースは採択時 MassBank であった。当該システムおよびドメインは日本質量分析学会に譲渡し、新たに公的レポジトリ MetaboBank を構築した。MetaboBank レポジトリは公募要領に示した「対象とするデータベース条件」 a~i の全てを満たしている。

(3) 統合化推進プログラムの他のチームや DBCLS との連携

糖鎖科学グループ(木下), 植物ゲノム情報解析グループ(田畑), プロテオームグループ(石濱)とは, DBCLS が主催するスパークルゾンにおいて実務者どうしが情報交換し, スパークル問い合わせ文の開発やデータフォーマットにおいて連携している。特にプロテオームグループとは, より密に連携するための会合を 2020 年 3 月 3 日に遺伝研で実施した。

タンパク質構造グループ(栗栖)とは, DBCLS, NBDC とともにジャパンバイオサイエンス情報ポータル(JBI ポータル)を構築し, 利用講習会を合同で実施するなどデータベース利用者を増やすための活動で連携している。

微生物統合グループ(黒川)とは, 将来的に DDBJ として統合・データ集約することを前提に DD BJ 統合アカウントシステムにともなう準拠する形式でレポジトリを構築している。とりわけ開発者の森とは日頃から情報交換を実施, チュートリアル等も一緒におこなう。メタデータ形式を作成する際も, 将来の統合検索を踏まえたすり合わせを実施している。

§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

表 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)

遺伝研グループ： MetaboBank

種別	2018年度	2019年度	2020年度 (9月末時点)
ユニーク IP 数	NA	NA	NA (10月5日公開)
ページ数	NA	NA	NA (10月5日公開)

理研グループ： RIKEN PMM

種別	2018年度	2019年度	2020年度 (9月末時点)
ユニーク IP 数	NA	618	600
ページ数	NA	4,349	9,065

かずさグループ： Metabolonote

種別	2018年度	2019年度	2020年度 (9月末時点)
ユニーク IP 数	1,214	1,879	1061
ページ数	14,664	19,703	18,868

かずさグループ： MassBase

種別	2018年度	2019年度	2020年度 (9月末時点)
ユニーク IP 数	113	128	152
ページ数	1,485	2,746	2,135

奈良グループ： KNA p SAcK

種別	2018年度	2019年度	2020年度 (9月末時点)
ユニーク IP 数	6,208	15,032	11,184
ページ数	134,918	281,068	167,161

(2) 分析

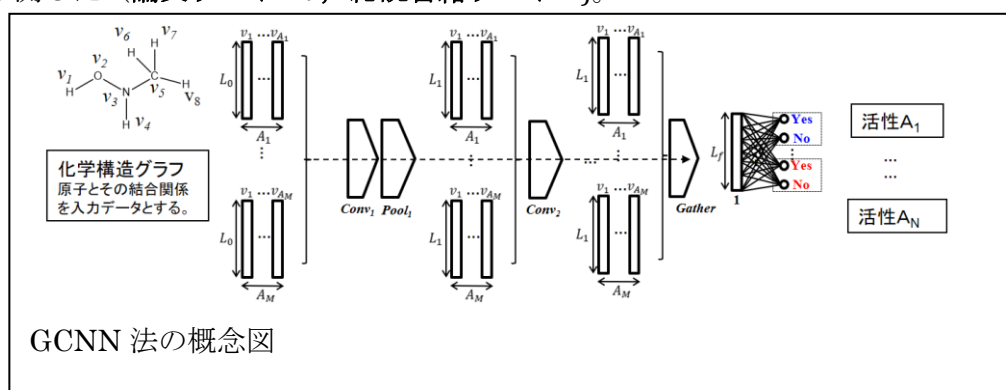
遺伝研 MetaboBank は 2020 年に開設で統計結果が無い。理研 PMM, Metabolonote, MassBase は、ここ数年利用が増加している。アクセスが圧倒的に多いのは KNApSAcK である。

2. データベースを利用して得られた研究成果・産業応用の例

インシリコ・スクリーニングにおける KNApSAcK Core DB の活用

(3a) 深層学習(グラフコンボリューション)による生合成開始物質の予測

代謝物ディープラーニングにおける重要課題は化学構造式を入力データに活用することである。グラフコンボリューションニューラルネットで、原子を頂点、結合を辺と捉える化学グラフを入力データとして活用し、アルカロイド化合物 515 種の生合成開始物質を予測した (論文リスト 19, 総説書籍リスト 4)。



また、代謝物の構造リストがインシリコ・スクリーニングに用いられている。代表的応用例として、Sars-Cov-2 における薬物設計などを実施した (論文リスト 17,18)。

3. その他

- ・ MS-DIAL ソフトウェアのダウンロード数はバージョン毎におおよそ 1000 件。2019 年度に本プロジェクト関係者以外による論文で MS-DIAL が言及されたものは、PubMed データベースによると 9 報。
- ・ KNApSAcK データベースは現在 knapsackfamily.com というミラーサイトにも委嘱され、同じコンテンツが奈良先端大とミラーサイトの両方からアクセスできる。上記のアクセス数は奈良先端大のサーバに基づく。

§5. 今後の展開

- ・ 研究開発の後半では、これまで開発してきたスペクトル・ライブラリにバージョンを付加して DVD-ROM 化し、販売する目処をつけたい。MetaboBank の登録を受け付ける専任キュレータの person 費を何らかの自己収入によって工面したいと考えている。
- ・ DDBJ サービスはアカデミアのみならず企業からの需要も大きい。その際、メタボローム解析に関しては企業から測定サンプルを受け取って解析結果を返すサービスを事業として展開したい。そうでないと上記の person 費を含めた持続性が担保できない。

§6. 自己評価

日本はメタボローム研究を早期に開始した国の一つだが、これまで生データを収集する公的リポジトリは存在しなかった。データ科学において比較可能な情報を収集することは極めて重要なため、国際メタボロミクス学会の学会標準に準拠したリポジトリ **MetaboBank** の開設は極めて意義のある試みと考えている。

DDBJ サービスの一部としてこの事業を開始できたことも、国際的な知名度や信頼を得る上で非常に重要なステップであった。しかし本プログラムを含めた外部資金による運営になっているため、今後数年間のうちに自立できる目処をつけなくてはならない。

本来こうした事業は競争的資金により研究者が実施する内容ではなく、国のデータベース戦略に基づいて持続的予算を配分して実施すべき内容である(参考:学術会議提言「持続可能な生命科学のデータ基盤の整備に向けて」2019年)。しかし現状ではそうした予算を期待できないため、本プログラムで実施せざるを得なかった。ソフトウェアやスペクトル・ライブラリは研究成果として作成されているが、その維持管理および品質向上に必要なリソースは研究予算で賄うことが難しい。その道筋をどうつけるかを、残り期間で解決せねばならない。

§7. 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)	21 件
未発行論文 (accepted, in press 等)	国内(和文)	0 件
	国際(欧文)	0 件

(2) 論文詳細情報 (PI, 学生, 直接雇用スタッフをボールド体)

1. Burla B, Arita M, **Arita M**, Bendt AK, Cazenave-Gassiot A, et al. “MS-based lipidomics of human blood plasma: a community-initiated position paper to develop accepted guidelines” *J Lipid Res* 59(10), 2001-2017, 2018 (DOI: 10.1194/jlr.S087163)
2. **Fukushima A**, Hikosaka S, Kobayashi M, Nishizawa T, Saito K, et al. “A Systems Analysis With “Simplified Source-Sink Model” Reveals Metabolic Reprogramming in a Pair of Source-to-Sink Organs During Early Fruit Development in Tomato by LED Light Treatments”, *Front Plant Sci.* 9:1439, 2018 (DOI: 10.3389/fpls.2018.01439)
3. Kitazaki K, **Fukushima A**, Nakabayashi R, Okazaki Y, Kobayashi M, et al. “Metabolic reprogramming in leaf lettuce grown under different light quality and intensity conditions using narrow-band LEDs”, *Sci Rep.* 8(1):7914, 2018 (DOI: 10.1038/s41598-018-25686-0)
4. **K Liu, AH Morita, S Kanaya** and Md. Altaf-Ul-Amin “Metabolite-content-guided prediction of medicinal/edible properties in plants for bioprospecting” *Curr Res Complement Altern Med* 130, 130, 2018 (DOI: 10.29011/CRCAM-130/100030)
5. **Eguchi R**, Karim MB, Hu P, Sato T, Ono N, **Kanaya S**, Altaf-Ul-Amin M “An integrative network-based approach to identify novel disease genes and pathways: a case study in the context of inflammatory bowel disease” *BMC Bioinform* 19(1), 264, 2018 (DOI: 10.1186/s12859-018-2251-x)
6. **Tada I**, Tsugawa H, Meister I, Zhang P, Shu R et al. “Creating a Reliable Mass Spectral-Retention Time Library for All Ion Fragmentation-Based Metabolomics” *Metabolites* 9(11), 251, 2019 (DOI: 10.3390/metabo9110251)
7. Lipidomics Standards Initiative Consortium “Lipidomics needs more standardization” *Nat Metabolism* 1(1), 745-747, 2019 (DOI: 10.1038/s42255-019-0094-z)
8. Tsugawa H, Satoh A, Uchino H, Cajka T, Arita M, **Arita M** “Mass Spectrometry Data Repository Enhances Novel Metabolite Discoveries with Advances in Computational Metabolomics” *Metabolites* 9(6), 119, 2019 (DOI: 10.3390/metabo9060119)
9. * Tsugawa H, Nakabayashi R, Mori T, Yamada Y, **Takahashi M**, et al. “A chemoinformatics approach to characterize metabolomes in stable-isotope-labeled organisms” *Nat Methods* 16(4), 295-298, 2019 (DOI: 10.1038/s41592-019-0358-2)

代謝物スペクトルのネットワークを用いて多様な二次代謝物の構造予測を実施した。本業績をもとに理研の高橋(直接雇用)・山田と津川がそれぞれ CSRS 奨励賞, 理研梅峰賞を受賞。

10. Vos RA, Katayama T, Mishima H, Kawano S, Kawashima S, et al. "BioHackathon 2015: semantics of data for life sciences and reproducible research" *F1000 Res* 9(1), 1-36, 2020 (DOI: 10.12688/f1000research.18236.1)
11. **A Fukushima**, T Kuroha, K Nagai, Y Hattori, M Kobayashi, et al. "Metabolite and Phytohormone Profiling Illustrates Metabolic Reprogramming as an Escape Strategy of Deepwater Rice during Partially Submerged Stress", *Metabolites* 10, 68 (2020). (DOI: 10.3390/metabo10020068)
12. NT Vu, **K Kamiya**, **A Fukushima**, S Hao, W Ning, et al. "Comparative co-expression network analysis extracts the SHSP70 gene affecting to shoot elongation of tomato", *Plant Biotechnol* 36, 143-153, 2019 (DOI: 10.5511/plantbiotechnology.19.0603a)
13. **Sakurai N**, Mardani-Korrani H, Nakayasu M, Matsuda K, Ochiai K, et al. "Metabolome analysis identified okaramines in the soybean rhizosphere as a legacy of hairy vetch" *Front Genet* 11, 114 2020 (DOI: 10.3389/fgene.2020.00114)
14. Ohbuchi K, **Sakurai N**, Kitagawa H, Sato M, Suzuki H, et al. "Differential annotation of converted metabolites (DAC-Met): Exploration of Maoto (Ma-huang-tang)-derived metabolites in plasma using high-resolution mass spectrometry" *Metabolomics* 16 (6), 63, 2020 (DOI: 10.1007/s11306-020-01681-3)
15. Hiraga Y, Ara T, Nagashima Y, Shimada N, **Sakurai N**, et al. "Metabolome analysis using multiple data mining approaches suggest luteolin biosynthesis in *Physcomitrella patens*" *Plant Biotechnol (Tokyo)* 37(3), 377-381, 2020 (DOI: 10.5511/plantbiotechnology.20.0525b)
16. Kusano M, **Fukushima A**, Tabuchi-Kobayashi M, Funayama K, et al. "Cytosolic GLUTAMINE SYNTHETASE 1;1 modulates metabolism and chloroplast development in roots", *Plant Physiol* 182(4):1894-1909 (2020). (DOI: 10.1104/pp.19.01118)
17. **Miyazaki Y**, Ono N, Huang M, Md.Altaf-Ul-Amin, **Kanaya S**, "Comprehensive exploration of target-specific ligands from compounds using a graph convolution neural network" *Mol Inf*, 39:e1900095 (2020) doi: 10.1002/minf.201900095
18. Hijikata A, Shionyu-Mitsuyama C, Nakae S, Shionyu M, Ota M, **Kanaya S**, Shirai T. "Knowledge-based Structural Models of SARS-CoV-2 Proteins and Their Complexes With Potential Drugs" *FEBS Lett*, 594: 1960-1973 (2020) doi: 10.1002/1873-3468.138
19. Ono N, **Eguchi R**, **Morita AH**, Katsuragi T, Nakamura S, Huang M, Md Altaf-Ul-Amin, **Kanaya S** "Classification of alkaloids according to the starting substances of their biosynthetic pathways using graph convolutional neural networks" *BMC Bioinf*, 20, 380, (2019) doi: 10.1186/s12859-019-2963-6
20. ***Tada I**, Chaleckis R, Tsugawa H, Meister I, et al. "Correlation-based deconvolution (corrdec) to generate high-quality MS2 spectra from data independent acquisition in multisample studies" *Anal Chem* 92(16), 11310-11317, 2020 (DOI: 10.1021/acs.analchem.0c01980)
AIF形式のノンターゲットデータからMS2スペクトルを抽出する手法を開発。本業績をもとに総研大の多田(当時大学院生)が遺伝研森島奨励賞, 総研大研究科長賞を受賞。
21. Tsugawa H, Ikeda K, Takahashi M, Satoh A, Mori Y, et al. "A lipidome atlas in MS-DIAL4" *Nat Biotechnol* 38(10), 1159-1163, 2020 (DOI: 10.1038/s41587-020-0531-2)

2. その他の著作物（総説、書籍など）

1. 神谷 健, 草野 都, 福島 敦史, “メタボロームデータ解析および解釈に資する可視化手法”, 日本化学会情報化学部会誌, 2019 年 37 巻 3 号 p. 72-. (DOI: 10.11546/cicsj.37.72)
2. 桑畑 和明, 佐久間 柚衣, 川島 雪生, 福島 敦史, 長嶋 雲兵, 草野 都, 立川 仁典, “量子化学計算による植物が生産する UV-B 防御物質の物性値予測”, Journal of Computer Chemistry, Japan 18 巻:108-114 (2019). (DOI: 10.2477/jccj.2019-0002)
3. Ichihashi Y, Fukushima A, Shibata A, Shirasu K, “High Impact Gene Discovery: Simple Strand-Specific mRNA Library Construction and Differential Regulatory Analysis Based on Gene Co-Expression Network”, Methods Mol Biol.1830:163-189 (2018). (DOI: 10.1007/978-1-4939-8657-6_11)
4. Kanaya S, Md. Altaf-Ul-Amin, Morita HA, Huang M., Ono N “Databases for natural product research” In *Comp. Nat. Prod. III* Volume 7, 222-238 (2020), Elsevier
5. Arita M “Bioinformatics – the power of integrated platforms for omics mining” In *Comp Nat Prod III* Volume 7, 211-221 (2020), Elsevier

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	14 件
	国際	4 件
口頭発表	国内	3 件
	国際	3 件
ポスター発表	国内	8 件
	国際	3 件

(2) 招待講演

〈国内〉

1. 発表者、タイトル、学会名、場所、月日
2. 福島 敦史, "RIKEN Plant Metabolome MetaDatabase の開発", 日本農芸化学会 20 年度福岡大会 シンポジウム「精密な食機能デザイン研究を切り拓く先端オミクスソリューション」、福岡、日本、2020 年 3 月 27 日
3. 福島敦史, "理研植物メタボロームメタ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 2019 年年会 第 8 回生命医薬情報学連合大会 IIBMP2019 BoF セッション「質量分析データベース・リポジトリの最前線」、東京、日本、2019 年 9 月 9 日
4. 福島敦史, 「質量分析を用いた植物メタボロミクスデータの情報解析と共有」、質量分析インフォマティクス研究会・第 4 回ワークショップ、東京、2019 年 3 月 19 日
5. 有田 正規 「バイオインフォマティクスが推し進めるバイオの世界」バイオインフォマティクスフォーラム, 8/24, 那覇, 2019

6. Masanori Arita "Repository and Academic Journals: the challenge of MetaboBank for data reuse", 京都生体質量分析研究会シンポジウム, 02/19, 2020
7. 櫻井 望「メタボローム解析の紹介」臨海ハッカソン DDBJing, 6/11, 隠岐, 2019
8. 櫻井 望「ノンターゲットメタボローム解析のマススペクトルデータ」第 46 回 BMS コンファレンス, 7/8, 札幌, 2019
9. 櫻井 望「メタボロームデータベースのこれまでとこれから」第 46 回 BMS コンファレンス, 7/10, 札幌, 2019
10. 櫻井 望「公共メタボロームデータレポジトリ MetaboBank の開発」IIBMP2019 BoF セッション「質量分析データベース・レポジトリの最前線」, 9/9, 東京, 2019
11. 櫻井 望「有用成分探索のための未知成分リファレンス構築の重要性～食品メタボロームレポジトリを通じて～」2019 年度第一回脂質駆動学術産業創生研究部会, 2/7, 京都, 2020
12. Nozomu Sakurai, Kunihiro Suda "Food Metabolome Repository: a reference of metabolite peaks for discovery and utilization of unknown chemicals" The 4th KBMS Symposium, Kyoto, Japan, February 19, 2020
13. Nozomu Sakurai "Construction of a reference to unknown metabolites for the discovery of unused plant derived specialized metabolites" JSPS Symposium 13 "New Trends in Plant Chemical Research by the Interconnection between Chemical Biology and Metabolite Chemistry", Osaka, Japan, March 20, 2020
14. 櫻井 望, 秋元 奈弓「新規フラボノイドを探索する FlavonoidSearch ツール」日本農芸化学会 2020 年度大会シンポジウム, 3/27, 福岡, 2020

〈国際〉

発表者、タイトル、学会名、場所、月日

1. Arita M "Cheminformatics for Predicting Structures from Mass Spectra" 1st Annual Conference of Chinese Society of Metabolomics (plenary), Shanghai, China, April 20 (19-21), 2019
2. Arita M "Data sharing and the power of omics integration" 6th Global Forum of Leaders for Agricultural Science and Technology (GLAST), Chengdu, China, Nov 12, 2019
3. Nozomu Sakurai "Construction of a metabolome database for uncovering the chemical world in foods", International Symposium on a New Era in Food Science and Technology 2019, Gifu, Japan, October 9, 2019
4. Nozomu Sakurai, Nayumi Akimoto "Discovery of novel flavonoids in nature by metabolome analyses and the FlavonoidSearch system" The 9th International Conference on Polyphenols and Health, Kobe, Japan, November 29, 2019

(3) 口頭講演

〈国内〉

発表者、タイトル、学会名、場所、月日

第 13 回メタボロームシンポジウム 筑波大学、2019 年 10 月 16-18 日

1. 多田一風太, 津川裕司, 有田正規 "MS-DIAL4.0 による統合解析"
2. 津川裕司, 池田和貴, 高橋みき子, 有田正規, 有田誠 "MS-DIAL4.0 による lipidomics standards initiative に基づく 120 脂質クラスの包括的解析"
3. **Atsushi Fukushima** and Kozo Nishida, "MSEAp: Development of a metabolite set enrichment analysis toolkit for plant metabolomics community", 第 36 回日本植物細胞分子生物学会(金沢)大会、金沢、2018 年 9 月 6 日(金)～8 日(日)

〈国際〉

発表者、タイトル、学会名、場所、月日

The 15th International Conference of the Metabolomics Society, Hague, June 23-27, 2019

1. Atsushi Fukushima, Mikiko Takahashi, Nozomu Sakurai, Toshiaki Tokimatsu, Hideki Nagasaki, Hideki Hirakawa, Takeshi Ara, Makoto Kobayashi, Miyako Kusano, Kazuki Saito, Masanori Arita, Norio Kobayashi, "RIKEN Plant Metabolome MetaDatabase: an integrated plant metabolome data repository based on the semantic web"
2. Hiroshi Tsugawa, Yoshifumi Mori, Yasuhiro Higashi, Aya Satoh, Sven Meyer, Kazuki Saito, Masanori Arita "MS-DIAL 4.0: a computational workflow for ion mobility tandem mass spectrometry data in metabolomics"
3. Ipputa Tada, Romanas Chaleckis, Hiroshi Tsugawa, Isabel Meister, Pei Zhang, Craig E. Wheelock, Masanori Arita "Correlation-based deconvolution (CorrDec) method for data-independent acquisition mass spectrometry"

(4) ポスター発表

〈国内〉

発表者、タイトル、学会名、場所、月日

1. 福島敦史, 高橋みき子, 櫻井望, 時松敏明, 長崎英樹, 平川英樹, 荒武, 小林誠, 草野都, 斉藤和季, 有田正規, 小林紀郎, "RIKEN Plant Metabolome MetaDatabase の開発", 第13回メタボロームシンポジウム 筑波大学, 2019年10月16-18日
2. 福島敦史ほか, 「植物メタボロームデータの再解析・アノテーション高度化に向けた情報基盤整備」, トーゴの日シンポジウム2018, 東京, 2018年10月5日(金)
3. 福島敦史ほか, 「植物メタボロームデータの再解析・アノテーション高度化に向けた情報基盤整備」, 第12回メタボロームシンポジウム, 鶴岡, 2018年10月17日(水)~19日(金)
4. 長崎英樹, 植物・食品メタボローム解析メタデータのRDF化と測定生データの再解析, 第13回メタボロームシンポジウム, 筑波大学, 2019年10月16-18日
5. 長崎英樹, 植物・食品メタボローム解析メタデータのRDF化および測定生データの再解析に向けて, トーゴの日シンポジウム2020, オンライン開催, 2020年10月5日

第13回メタボロームシンポジウム 筑波大学, 2019年10月16-18日

6. 長崎英樹, 大澤祥子, 荒武, 福島敦史, 高橋みき子, 小林紀郎, 櫻井望, 平川英樹, 有田正規 「植物・食品メタボローム解析メタデータのRDF化と測定生データの再解析」
7. 福島敦史, 高橋みき子, 櫻井望, 時松敏明, 長崎英樹, 平川英樹, 荒武, 小林誠, 草野都, 斉藤和季, 有田正規, 小林紀郎, "RIKEN Plant Metabolome MetaDatabase の開発"
8. 川島武士 "進化発生生物学にメタボローム解析を取り込む"

〈国際〉

発表者、タイトル、学会名、場所、月日

1. Atsushi Fukushima, et al. "RIKEN Plant Metabolome MetaDatabase: An integrated plant metabolome data repository based on the semantic web", SWAT4HCLS 2018, Antwerp, December 4-5, 2018

The 15th International Conference of the Metabolomics Society, Hague, June 23-27,

2019

2. Katsuya Obuchi, Nozomu Sakurai, Hiroyuki Kitagawa, Hirotaka Kushida, Akinori Nishi, Masahiro Yamamoto, Kazuhiro Hanazaki, Masanori Arita “DAC-Met : Exploring the metabolic fate of the herbal components in “maoto” decoction through a differential annotation strategy”
3. Nozomu Sakurai, Kunihiro Suda “Food Metabolome Repository: A database for cross-sample specificity-based peak prioritization in untargeted metabolomics”

4. 知財出願

(1) 出願件数

種別		件数
特許出願	国内	0 件
	国外	0 件
その他の知的財産出願		0 件

(2) 一覧

なし

5. 受賞・報道等

(1) 受賞

1. 受賞名、受賞者、月日
 - ・ 理研 CSRS 奨励賞、高橋みき子、山田豊、2020 年 6 月 10 日
 - ・ 理研梅峰賞、津川裕司、2020 年 3 月 12 日 https://www.riken.jp/pr/news/2020/20200319_1/index.html
 - ・ 国立遺伝学研究所森島奨励賞、多田一風太、2020 年 9 月 29 日
 - ・ 総合研究大学院大学研究科長賞、多田一風太、2020 年 9 月 29 日

(2) メディア報道

該当なし

(3) その他

該当なし

§8. 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
2018年 5月1日	チーム内ミーティング(非公開)	NAIST 金谷研究室	15人	メタボロームリポジトリの会議
2018年 7月20日	チーム内ミーティング(非公開)	東京連絡事務所会議室3	8人	同上
2018年 7月27日	チーム内ミーティング(非公開)	遺伝研有田研究室	6人	同上
2018年 9月26日	チーム内ミーティング(非公開)	東京連絡事務所会議室3	8人	同上
2018年 10月12日	チーム内ミーティング(非公開)	東京連絡事務所会議室3	8人	同上
2018年11 月7日	チーム内ミーティング(非公開)	東京連絡事務所会議室3	7人	同上
2018年12 月18日	チーム内ミーティング(非公開)	東京連絡事務所会議室3	7人	同上
2019年2 月19日	チーム内ミーティング(非公開)	東京連絡事務所会議室1	7人	同上
2019年 5月30-31 日	チーム内ミーティング(非公開)	遺伝研	7人	メタボロームリポジトリ MetaboBankの研究開発進捗報告のためのミーティング
2019年 7月18日	チーム内ミーティング(非公開)	かずさDNA 研究所	9人	研究進捗報告のためのミーティング
2019年 7月26日	チーム内ミーティング(非公開)	かずさDNA 研究所	4人	研究進捗報告のためのミーティング
2019年 8月29日	チーム内ミーティング(非公開)	理研東京連絡事務所	6人	メタボロームリポジトリ MetaboBankの研究開発進捗報告のためのミーティング
2019年 9月25日	チーム内ミーティング(非公開)	理研横浜	8人	同上
2019年 11月28日	チーム内ミーティング(非公開)	遺伝研	8人	同上
2020年 2月4日	チーム内ミーティング(非公開)	理研横浜	8人	同上

年月日	名称	場所	参加人数	目的・概要
2020年 4月21日	チーム内ミーティング(非公開)	オンライン開催	9人	同上
2020年 5月21日	チーム内ミーティング(非公開)	オンライン開催	9人	同上
2019年 6月25日	チーム内ミーティング(非公開)	オンライン開催	10人	同上
2020年 8月24日	チーム内ミーティング(非公開)	オンライン開催	10人	同上

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
2019年10月 17日	13回メタボロームシンポジウム 特別セミナー	筑波大学	80名	NBDC 統合プロジェクトにおけるメタボローム活動の紹介
2019年6月 22日	LipoQuality International Workshop on Lipid omics	理研横浜	60名	新学術領域リポクオリティ, シンガポール国立大学と合同でリポドミクス技術に関する国際ワークショップ
2020年9月 2日	日本バイオインフォマテイクス学会年会 (IIBMP2020)チュートリアル	オンライン	各セッション 200名	DDBJ 主催の利用チュートリアル(3セッション)においてメタボロミクスも紹介
2020年10月 22-29日	日本インドネシアバイオインフォマテイクス・バイオリソースワークショップ	オンライン	30名	DDBJ 主催の利用チュートリアル(英語。2時間ずつ5日)においてメタボロミクスも紹介

以上