

○長崎 英樹¹⁾、○荒 武^{1,2)}、大澤祥子¹⁾、福島 敦史³⁾、高橋 みき子³⁾、
小林 紀郎⁴⁾、藤澤 貴智⁵⁾、時松 敏明⁵⁾、櫻井 望⁵⁾、金谷 重彦⁶⁾、
平川 英樹¹⁾、有田 正規^{3,5)}

- 1) かずさDNA研究所ゲノム情報解析施設
- 2) 京都大学生存圏研究所
- 3) 理化学研究所環境資源科学研究センター
- 4) 理化学研究所情報統合本部
- 5) 国立遺伝学研究所生命情報・DDBJセンター
- 6) 奈良先端科学技術大学院大学情報科学領域

要旨

多様な代謝物をより多く同定することを目指すメタボローム解析は実験データだけでなく、実験手法を記載したメタデータも複雑化しており、世界的にデータの標準化が遅れている。我々は、公共リポジトリ MetaboBankにおけるメタボローム解析メタデータをResource Description Framework (RDF)というデータ形式でアーカイブ化を行っている。

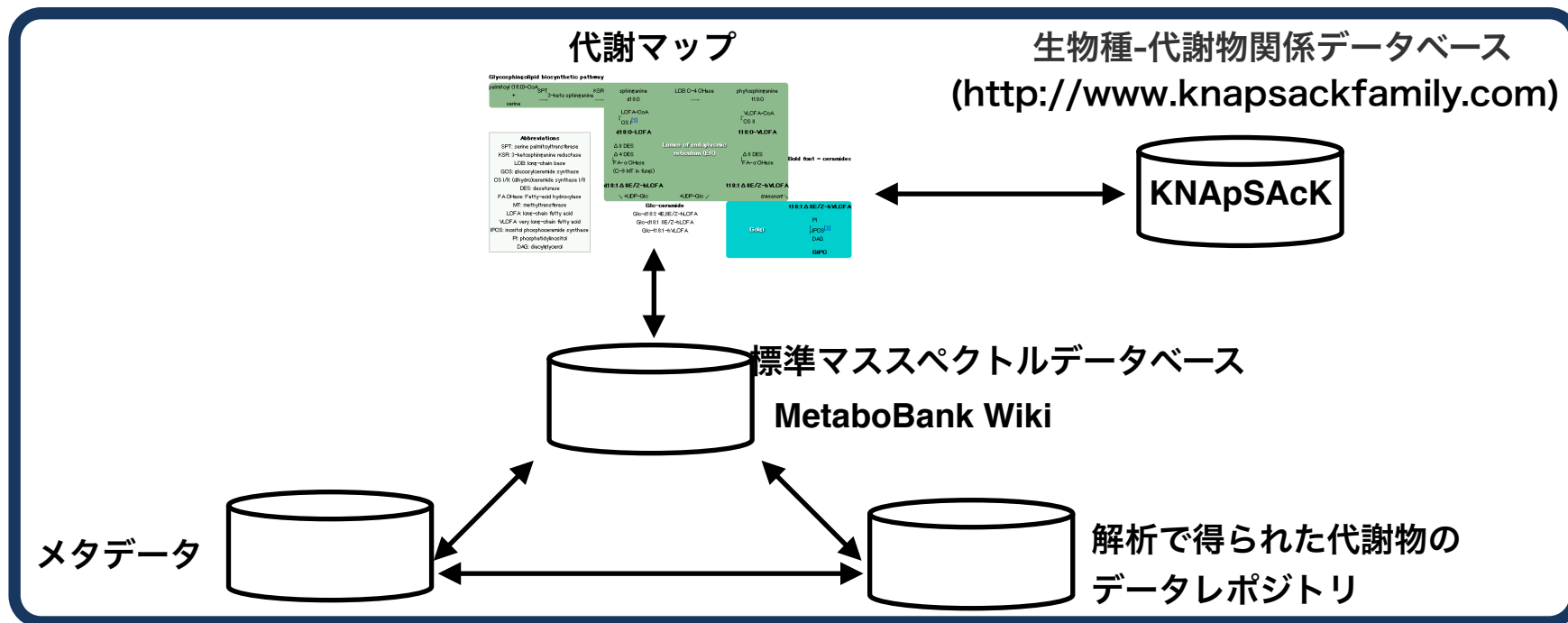
現在かずさDNA研究所(KDRI)と理化学研究所(理研)にて測定された、植物の解析メタデータを対象にRDFのデータモデルを作成している。また他の生物種や実験条件にもデータモデルを拡張中である。

また、現在奈良先端大のKNASAcKなど他のメタボローム関連データベースやKDRIのPlant GARDENなどゲノム情報関連のデータベースとのRDFを介した連携を進めているとともに、KDRIと理研では生データの再解析を行っている。

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/物質循環を考慮したメタボロミクス情報基盤

メタボローム 統合データベースMetaboBank

国立遺伝学研究所 (NIG)



かずさDNA研究所 (KDRI)



実験生データとメタデータを提供

理科学研究所 (理研)



データ連携協定によりKDRIのデータセットは理研にも提供される。



Metabolonoteの特徴

Metabolonoteの書式自由度が高く、データ間の紐付けが緩いリテラルのメタデータ

KDRIのメタボローム解析時のサンプル情報や分析情報(メタデータ)を専門に取り扱うデータベース

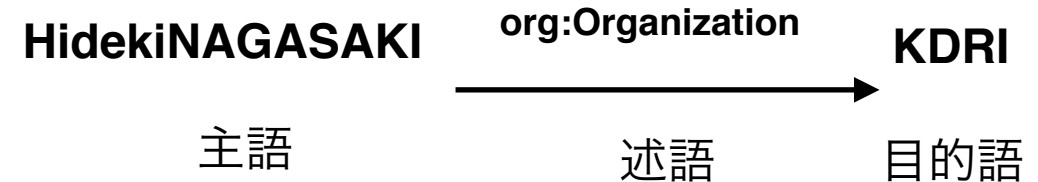
Wikiで作られていて作成や編集が容易

メタデータをXMLで一括出力可

Description

Harvested sample is frozen by liquid N2 and resulting powder (100mg) are solved in 300uL 80% methanol solution. 20uL sample is injected into HPLC after 0.2um membrane filter treatment. HPLC conditions: Agilent 1100 series (Agilent), Column: TSKgel-100V (4.6 x 250 mm, 5 micrometer; TOSOH), Solvent: A; 0.1% formic acid aq. B; ACN (addition 0.1% formic acid fc.), Gradient: (B);3 to 30% (0.0 to 25.0 min), 30 to 90% (25.0 to 40.0 min), 90% (40.0 to 45.0 min), 95% (45.1 to 50.0 min), 3% (50.1 to 57.0 min), Column temp.: 30 degree C, Flow rate=0.5mL/min, PDA: 200-650 nm (2 nm step). FT-ICR-MS conditions: Filter 1: FTMS + c norm !corona !pi res=50000 o(200.0-1500.0); 2: ITMS + c norm !corona !pi Dep MS/MS Most intense ion from (1);,3: ITMS + c norm o(200.0-1500.0)., Rejected mass=266.0000;294.0000;391.0000.

MetaboBank は Resource Description Framework (RDF) という形式で構築。



MetaboBank用に、定義づけをやり直す。

<http://metabolonote.kazusa.or.jp/>

Ara *et al.* Front Bioeng Biotechnol. 2015 Apr 7;3:38.



基本的にデータを手入力

Analytical Method Details Information

ID	MS1
Title	LC-FT-ICR-MS ESI positive method 1
Instrument	Agilent1100 HPLC (Agilent), LTQ-FT (Thermo Fisher Scientific)
Instrument Type	LC-FTICR-MS
Ionization	ESI
Ion Mode	Positive
Description	Harvested sample is frozen by liquid N2 and resulting powder (100mg) are solved in 300uL 80% methanol solution. 20uL sample is injected into HPLC after 0.2um membrane filter treatment. HPLC conditions: Agilent 1100 series (Agilent), Column: <u>TSKgel-100V (4.6 x 250 mm, 5 micrometer; TOSOH)</u> , Solvent: A; 0.1% formic acid aq. B; ACN (addition 0.1% formic acid fc.), Gradient: (B);3 to 30% (0.0 to 25.0 min), 30 to 90% (25.0 to 40.0 min), 90% (40.0 to 45.0 min), 95% (45.1 to 50.0 min), 3% (50.1 to 57.0 min), Column temp.: <u>30 degree C</u> , Flow rate=0.5mL/min, PDA: 200-650 nm (2 nm step). FT-ICR-MS conditions: Filter 1; FTMS + c norm !corona !pi res=50000 o(200.0-1500.0); 2: ITMS + c norm !corona !pi Dep MS/MS Most intense ion from (1).;3: ITMS + c norm o(200.0-1500.0)., Rejected mass=266.0000;294.0000;391.0000.
Extraction	
HPLC	
Mass	

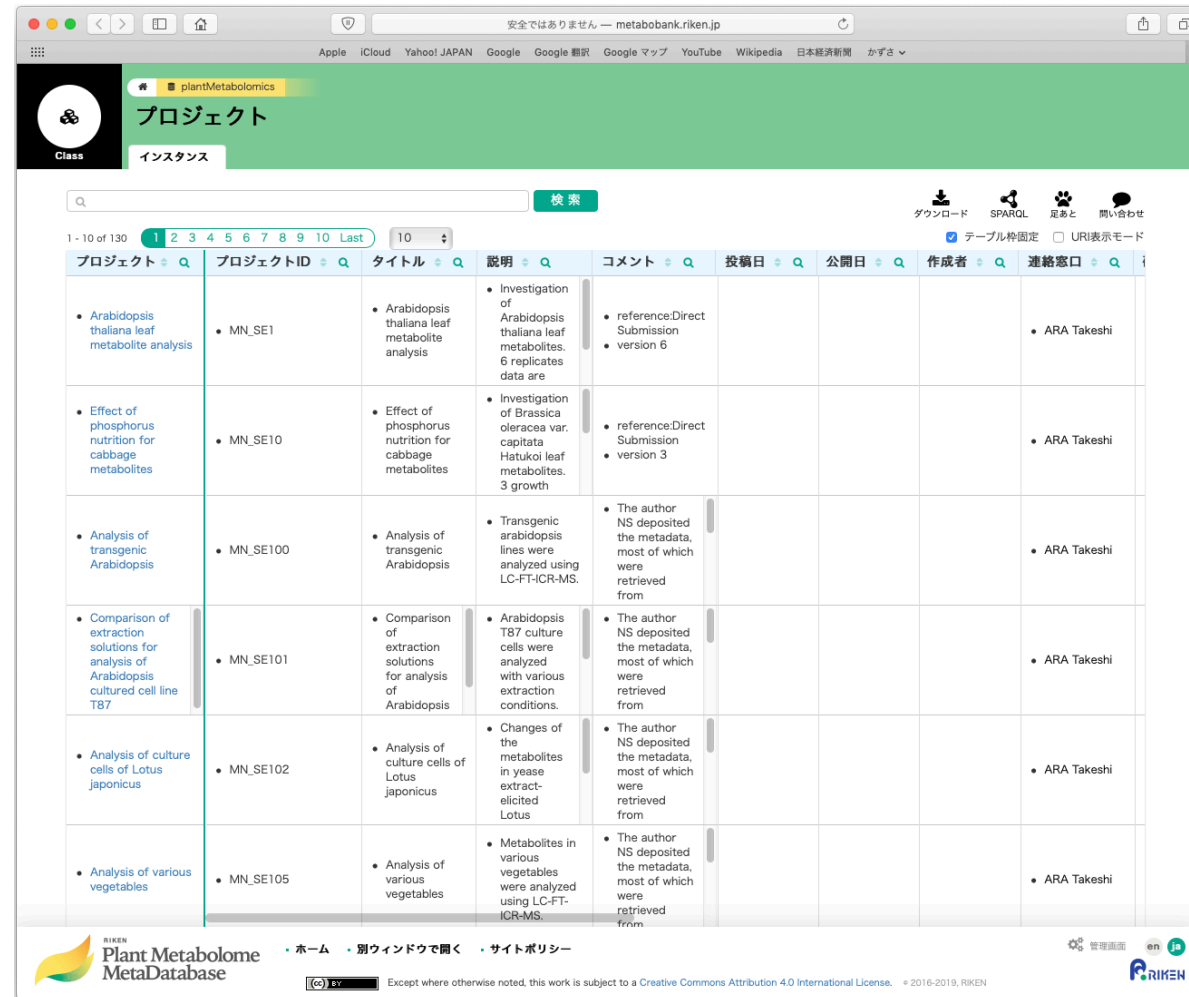
植物等のかずさ独自	93スタディ
食品メタボローム	16スタディ
その他(東大など)	6スタディ
合計	115スタディ

共通の項目(クラス:述語)でリンクされた
50シートのRDF変換用エクセルファイル

Chromatography	comment	description	temperature gradient	column type	column temperature	column pressure	column name
クロマトグラフィ	コメント	説明	温度勾配	カラムの種類	カラム温度	カラム圧力	カラム名
Chromatography	rdfs:comment	dcterms:description	temperatureGradient	columnType	columnTemperature	columnPressure	columnName
Chromatography	rdf:langString	rdf:langString	xsd:string	xsd:string	Temperature	Pressure	xsd:string
pm_chromato:MN_SE2_HPLC	"Elute monitoring by PDA equipped with Agilent1100 HPLC in 2 nm step."	@en	[temp:30C]	"TSKgel-100V (4.6 x 250 mm, 5			

=>エクセルファイルに収めたデータをRDF化 (turtle ファイル作成)

理研とのデータ連携協定に基づき理研のPlantMetabolomeMetaDatabase (PMM) から植物のみで構成された88スタディが現在公開されている。



The screenshot shows a web browser displaying the PlantMetabolomics database interface. The page title is "プロジェクト" (Project) and it shows a list of 130 studies. The table below summarizes the visible data:

プロジェクト	プロジェクトID	タイトル	説明	コメント	投稿日	公開日	作成者	連絡窓口
• Arabidopsis thaliana leaf metabolite analysis	• MN_SE1	• Arabidopsis thaliana leaf metabolite analysis	• Investigation of Arabidopsis thaliana leaf metabolites. 6 replicates data are	• reference:Direct Submission • version 6				• ARA Takeshi
• Effect of phosphorus nutrition for cabbage metabolites	• MN_SE10	• Effect of phosphorus nutrition for cabbage metabolites	• Investigation of Brassica oleracea var. capitata Hatukoi leaf metabolites. 3 growth	• reference:Direct Submission • version 3				• ARA Takeshi
• Analysis of transgenic Arabidopsis	• MN_SE100	• Analysis of transgenic Arabidopsis	• Transgenic arabidopsis lines were analyzed using LC-FT-ICR-MS.	• The author NS deposited the metadata, most of which were retrieved from				• ARA Takeshi
• Comparison of extraction solutions for analysis of Arabidopsis cultured cell line T87	• MN_SE101	• Comparison of extraction solutions for analysis of Arabidopsis	• Arabidopsis T87 culture cells were analyzed with various extraction conditions.	• The author NS deposited the metadata, most of which were retrieved from				• ARA Takeshi
• Analysis of culture cells of Lotus japonicus	• MN_SE102	• Analysis of culture cells of Lotus japonicus	• Changes of the metabolites in yease extract-elicited Lotus	• The author NS deposited the metadata, most of which were retrieved from				• ARA Takeshi
• Analysis of various vegetables	• MN_SE105	• Analysis of various vegetables	• Metabolites in various vegetables were analyzed using LC-FT-ICR-MS.	• The author NS deposited the metadata, most of which were retrieved from				• ARA Takeshi

<http://metabobank.riken.jp/pmm/db/plantMetabolomics>

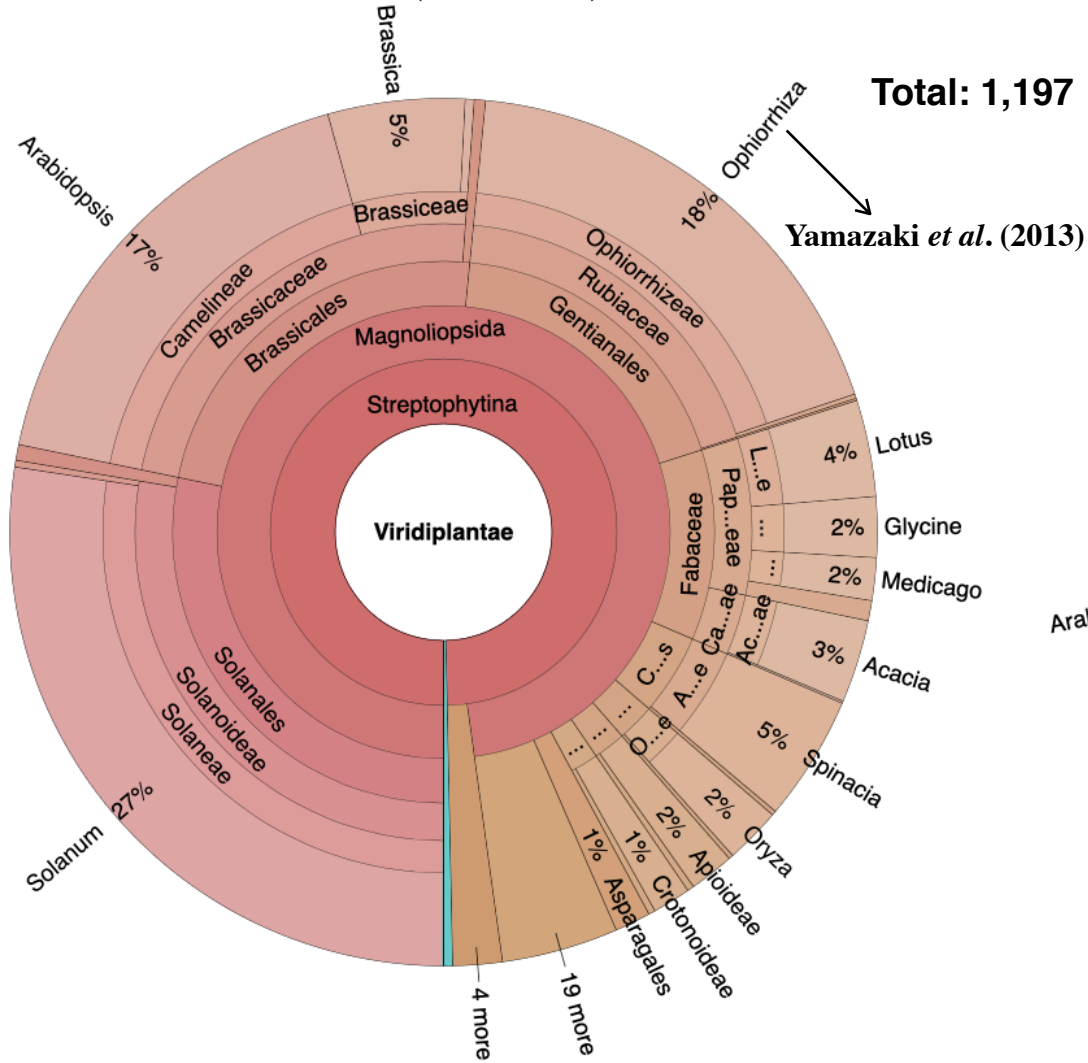
Licensed under a Creative Commons 表示4.0国際ライセンス

©2021長崎英樹 (かずさDNA研究所)

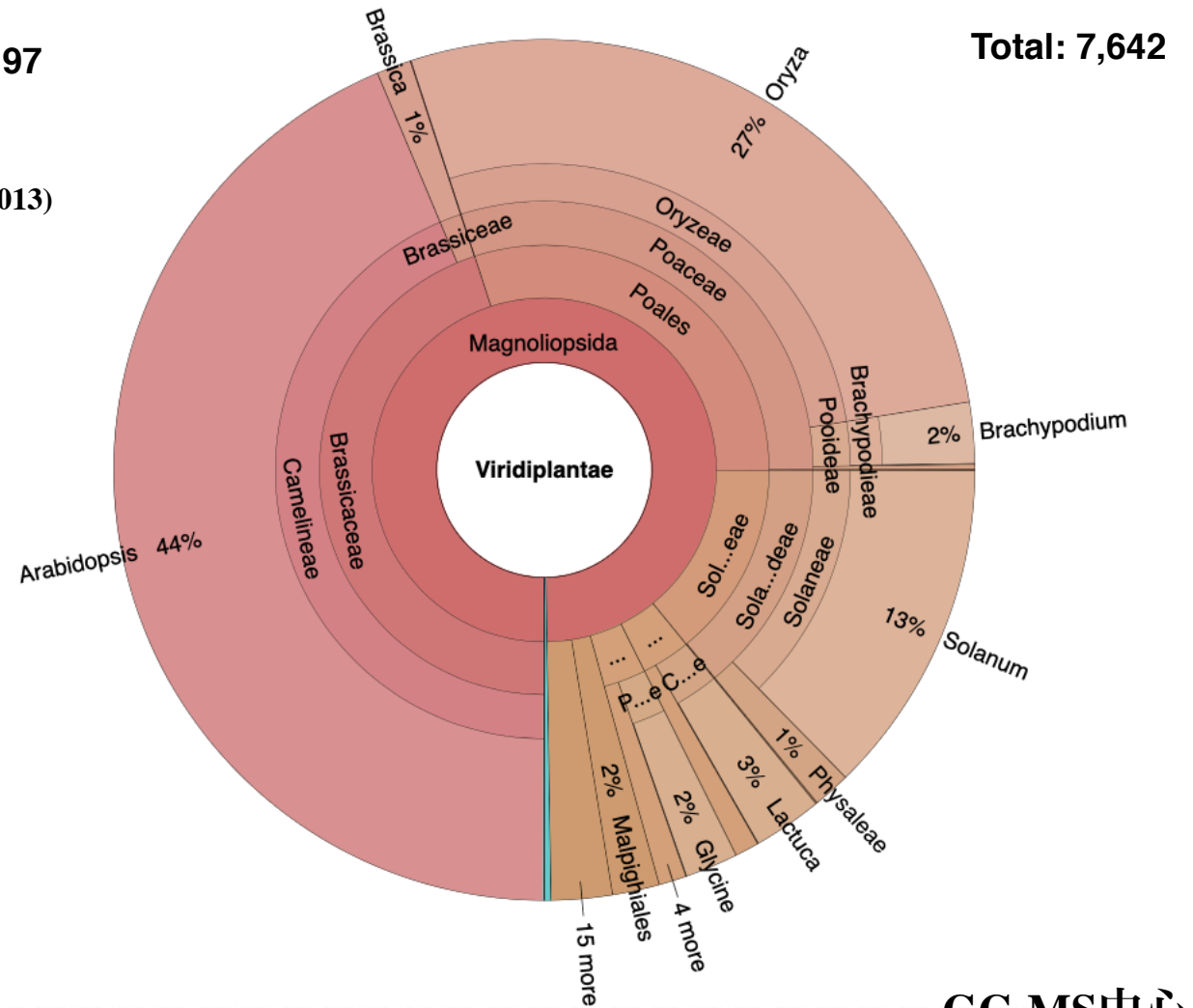


PMM中の植物サンプルの内訳
MetaboBankにおいても同様のデータを格納中

Metabolonote (2021年8月)



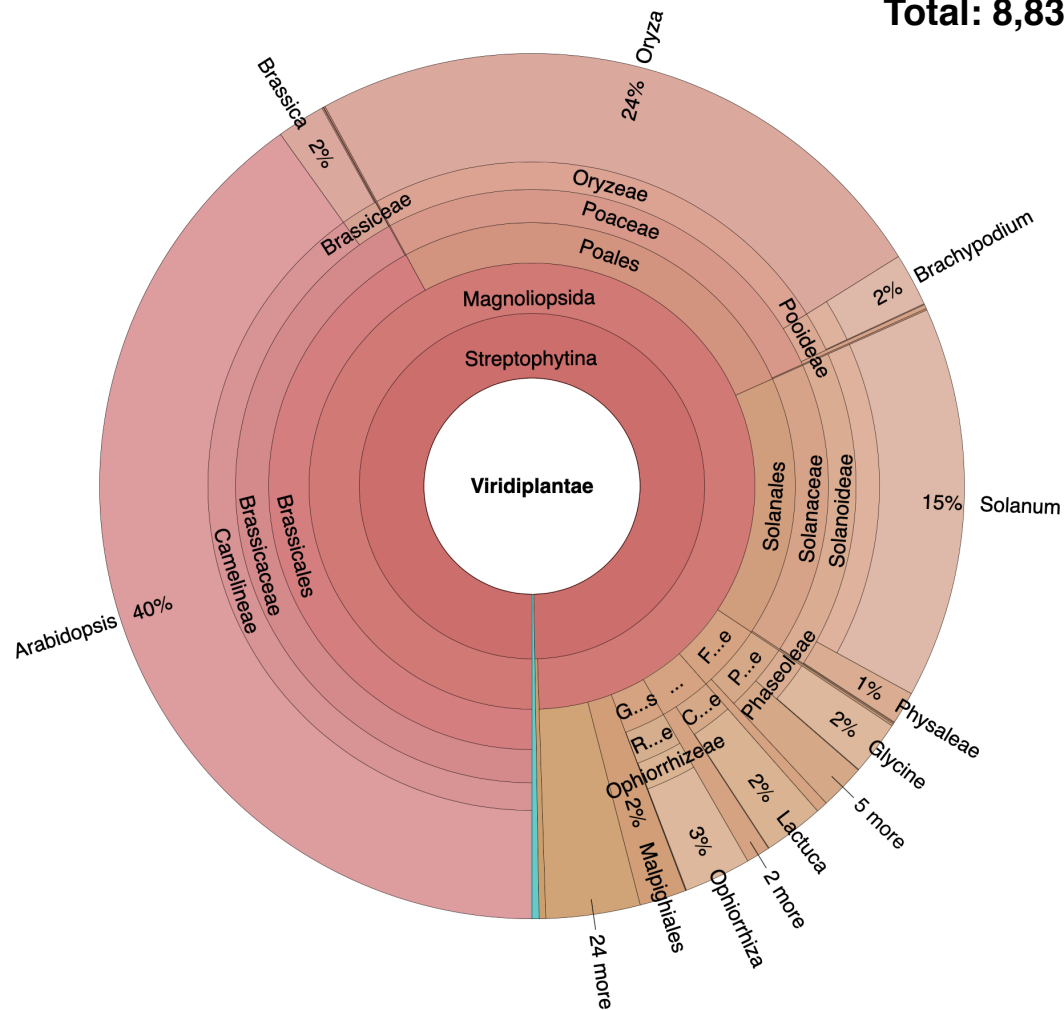
RIKENの植物データ (2021年8月)



メタボローム データベース中の植物サンプルの内訳

PMM (2021年7月)

Total: 8,839

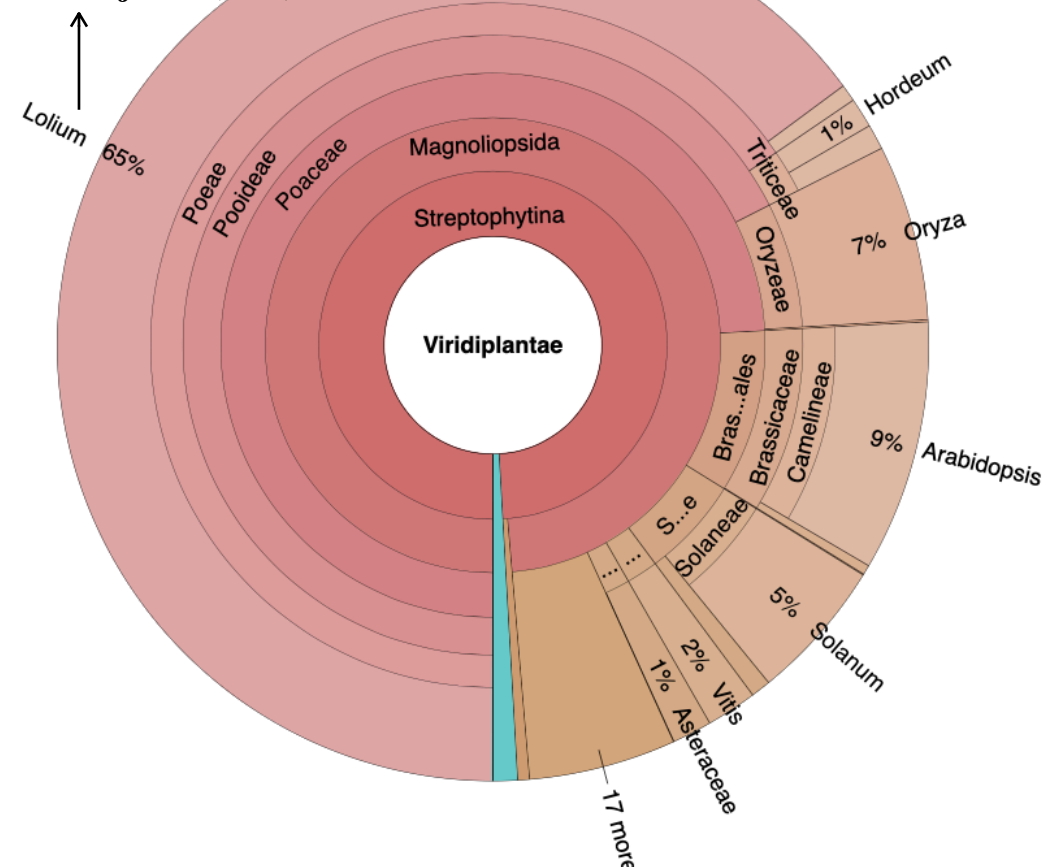


MetaboLightsの植物データ (2021年7月)

(<https://www.ebi.ac.uk/metabolights/>)

Total: 49,496

Subbraj *et al.* (2019)



PMM(MetaboBank)は突出して多い植物種はなくバランスがよい。

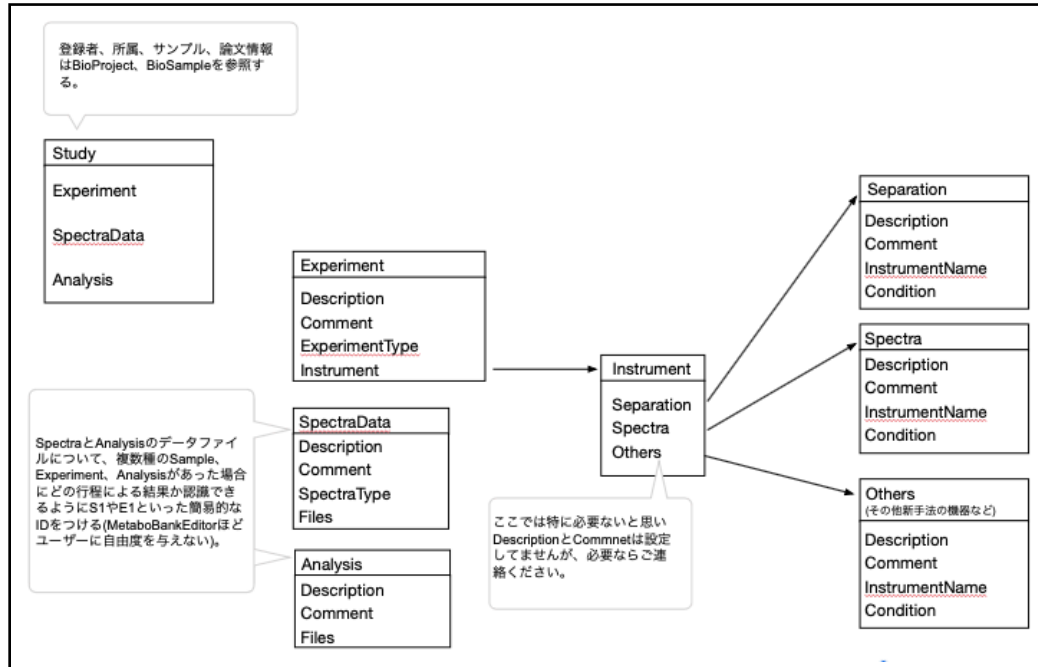
今後MetaboBankは、植物以外のヒト、食品や一般ユーザーからのデータも格納する計画である。

Licensed under a Creative Commons 表示4.0国際ライセンス

©2021長崎英樹 (かずさDNA研究所)



RDFスキーマを簡略化し、入力必須項目を削減した Ver. 2 を作成中 2021トーゴーの日リリース予定。

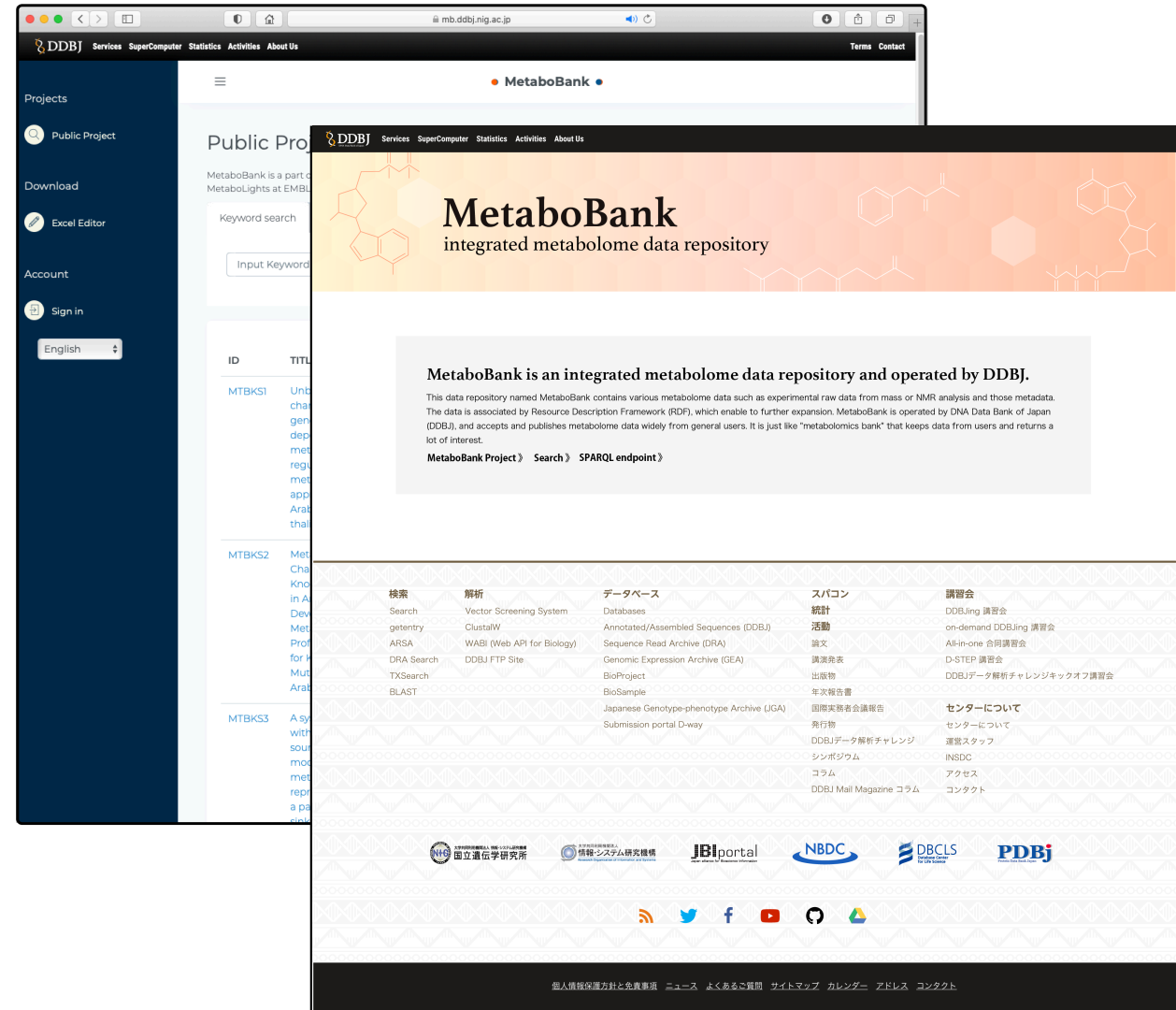


MetaboLights等とほぼ同じ条件。

NMRなどKDRI、理研の実験で使用されなかった解析機器にも対応。

ユーザアカウントだけでなく、DDBJ BioProject、BioSampleデータベースと連携して、統合オミックス解析に対応。

配列データベースと共通する串刺し検索が可能。

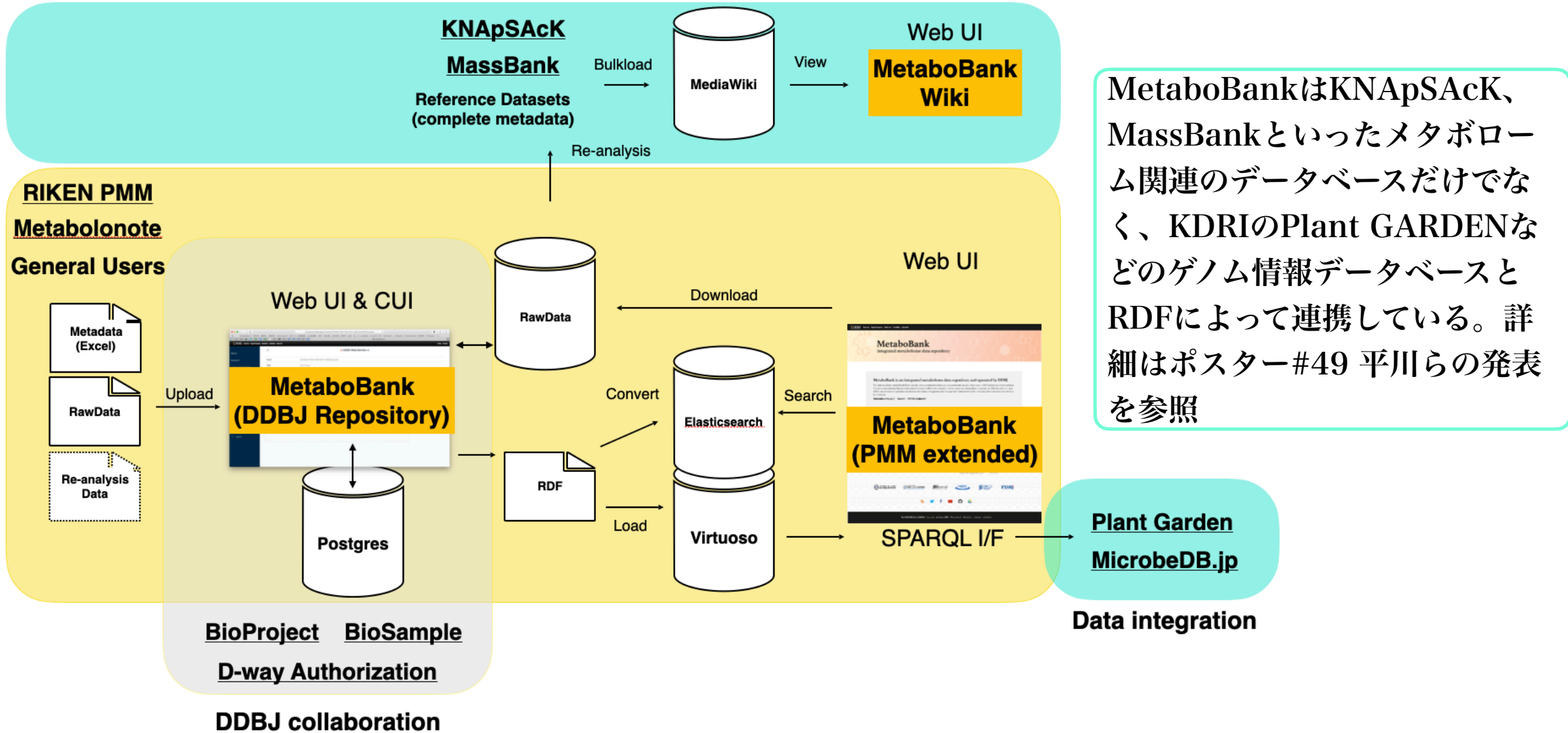


Licensed under a Creative Commons 表示4.0国際ライセンス

©2021長崎英樹 (かずさDNA研究所)



MetaboBankの構造



実験生データ再解析のためのPowerGetBatchの実行環境整備

京大・生存圏研 荒先生との共同研究

PowerGetBatch (<http://www.kazusa.or.jp/komics/software/PowerGetBatch>)

液体クロマトグラフィー(LC)-高分解能質量分析(MS)のデータから、ピーク抽出、アダクト判定、サンプル間のアラインメントを行い、化合物データベース検索による一次アノテーションを行う。

MetaboBankに移植するKNpSAcKなどのデータを用いて化合物の構造推定の向上が期待できる。

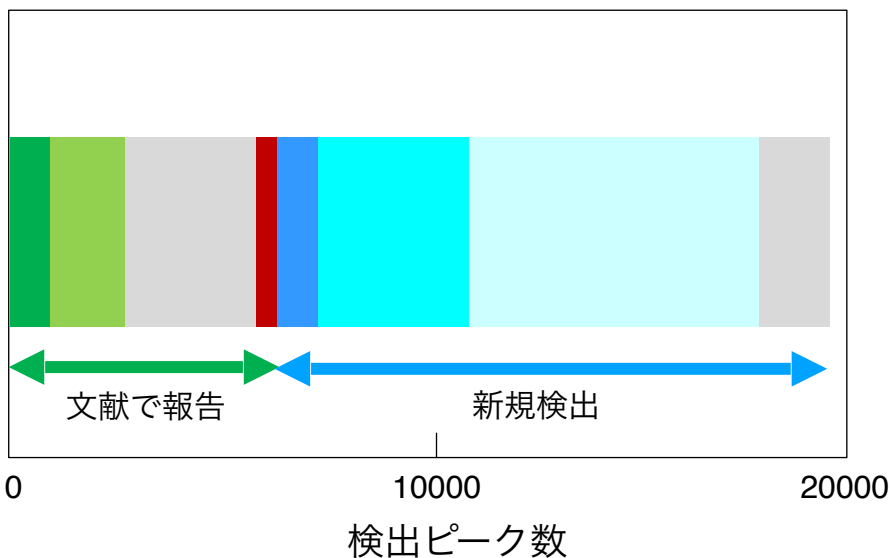
PowerGetBatchはβ版でパラメータ設定、マニュアルの整備や使いやすさの改善が必要。



スパコンへの実装

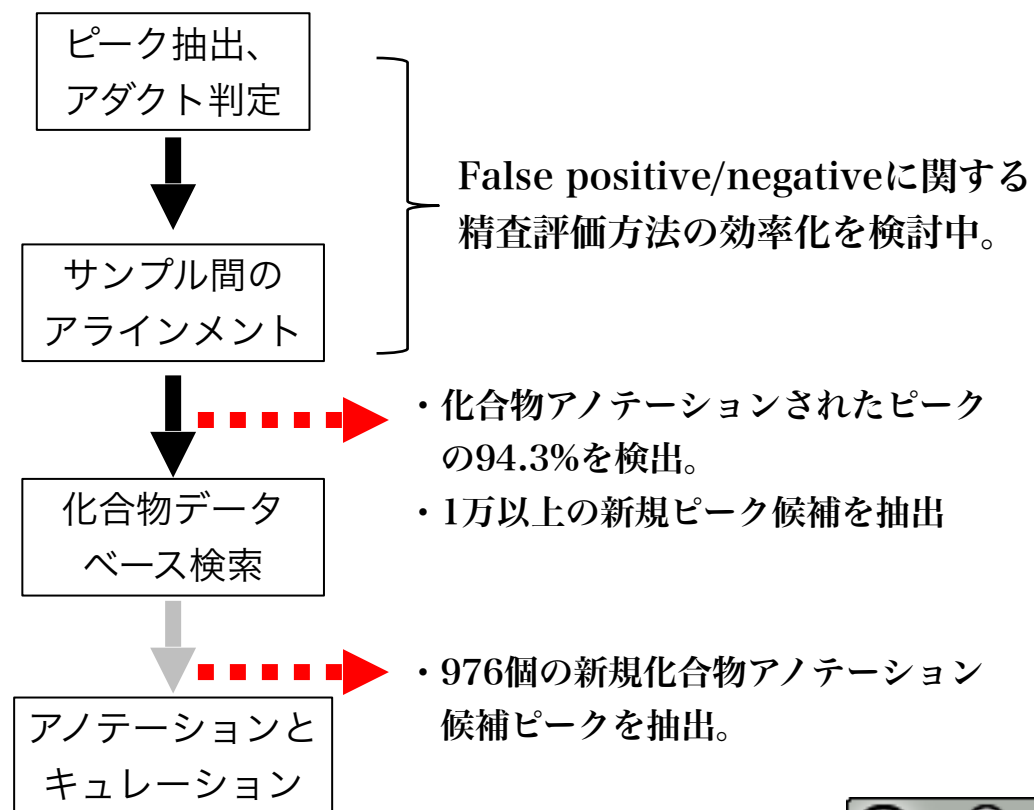
- Linux上でqsubコマンドによる並列処理を実行した。
- テストデータとして、*Jatropha cruceus* (Sano *et al.*, 2012)の分析データ (<http://metabolonote.kazusa.or.jp/SE15:/>)を使用した。

再解析結果 (暫定版)



- 化合物分類情報あり
- 組成式情報あり
- アノテーションなし
- 今回不検出
- 化合物分類情報候補あり
- 化合物DBヒット情報あり
- 組成式情報候補あり
- アノテーションなし

計算処理のワークフロー



まとめ

メタボローム統合データベースMetaboBank Ver.2をリリースする。
現在KDRIや理研の植物由来のデータ中心だが、今後それ以外の他の生物種や実験条件にもRDFデータモデルを拡張中である。

DDBJ BioProject、BioSampleデータベースと連携して、統合オミックス解析に対応。
ユーザーの入力必須項目を削減し、広くデータ受け入れる体制を整備、公共リポジトリとして稼働を目指す(来年度)。

KNASAcKなど他のメタボローム関連データベースやKDRIのPlant GARDENなどゲノム情報関連のデータベースとのRDFを介した連携を進めている。

現在生データ再解析を行っておりPowerGetBatch結果とキュレーション結果とを比較し、パラメータ設定中である。これらの再解析結果もMetaboBankのプロジェクトとしてアーカイブ化の予定である。