



MetaboBank

メタボロミクス再解析のためのリポジトリ

国立遺伝学研究所 有田正規

協力者： 奈良先端大学院大学 金谷重彦
理研環境資源科学研究センター 福島敦史
(現 京都府立大学)
かずさDNA研究所 櫻井望 (現 遺伝研)
かずさDNA研究所 平川英樹
東京農工大学 津川裕司

公共リポジトリ MetaboBank

DDBJセンターが担当。 2021年よりデータ受付を開始。現在3プロジェクト登録済（MTBKS201－207）、申込み3件。

（JST-NBDC「統合化推進」、AMED-CREST「微生物叢」）

DDBJ Services SuperComputer Statistics Activities About Us

MetaboBank

integrated metabolome data repository

MetaboBank is an integrated metabolome data repository and operated by DDBJ.

This data repository named MetaboBank contains various metabolome data such as experimental raw data from mass or NMR analysis and those metadata. The data is associated by Resource Description Framework (RDF), which enable to further expansion. MetaboBank is operated by DNA Data Bank of Japan (DDBJ), and accepts and publishes metabolome data widely from general users. It is just like "metabolomics bank" that keeps data from users and returns a lot of interest.

[MetaboBank Project](#) [Search](#) [SPARQL endpoint](#)

Search	Analysis	Databases	NIG SuperComputer
getentry	Vector Screening System	Annotated/Assembled Sequences (DDBJ)	NIG SuperComputer
ARSA	WABI (Web API for Biology)	Sequence Read Archive (DRA)	
DDBJ Search	DDBJ FTP Site	Genomic Expression Archive (GEA)	
TXSearch		BioProject	
		BioSample	
		MetaboBank	
		Japanese Genotype-phenotype Archive (JGA)	
		Submission portal D-way	

リポジトリの意義

多くの学術誌が「**公共リポジトリ**」への生データ登録を義務付け

「公共」であるためにはパブリック・ドメインであることに加え、登録者に対して公平（中立）であることが重要。

- 研究再現性の担保 （←詳細メタデータが必要）
- 研究不正への対応 （←自機関からの公開は駄目）

メタボロミクスにおける公共リポジトリ

1. MetaboLights (EBI)
2. MetabolomicsWorkbench & massive (UC San Diego)

アジアで初めてのメタボロミクス向け公共リポジトリがMetaboBank。

どうしたら再解析が可能になるのか？ → メタデータの充実

登録手順

手順はNGSデータの登録を踏襲（ウェブページに詳述）

NGSデータを登録した経験のある人なら、ストレスなく登録できるはず

D-way, JGA サーバ 停止 9月5日 10:00-12:00

(9/5-7).getentry と一部の anonymous ftp の更新停止

MetaboBank

Home

Submission ▾

FAQ

Search

Download

Contact

Metabobankで検索

登録の流れ

アクセッション番号

データ公開

Reviewer access

更新

補足: MD5 値

ホーム > metabobank > MetaboBank への登録

MetaboBank への登録

登録の流れ

MetaboBank は、塩基配列や遺伝子発現データと関連づけられるように BioProject と BioSample と連携しています。メタデータは実験種別毎に用意されたエクセルファイル（以下に記載するMAGE-TAB の IDF と SDRF）に記載します。

解析済みのデータ（同定・推定された化合物に関する情報）は所定の形式である Metabolite assignment file (MAF) に記載します。

1. 登録アカウントの取得

- [D-way_登録アカウント](#)を作成します。[マニュアル](#)
- [公開鍵と center name](#) をアカウントに登録し MetaboBank 登録を可能にします。

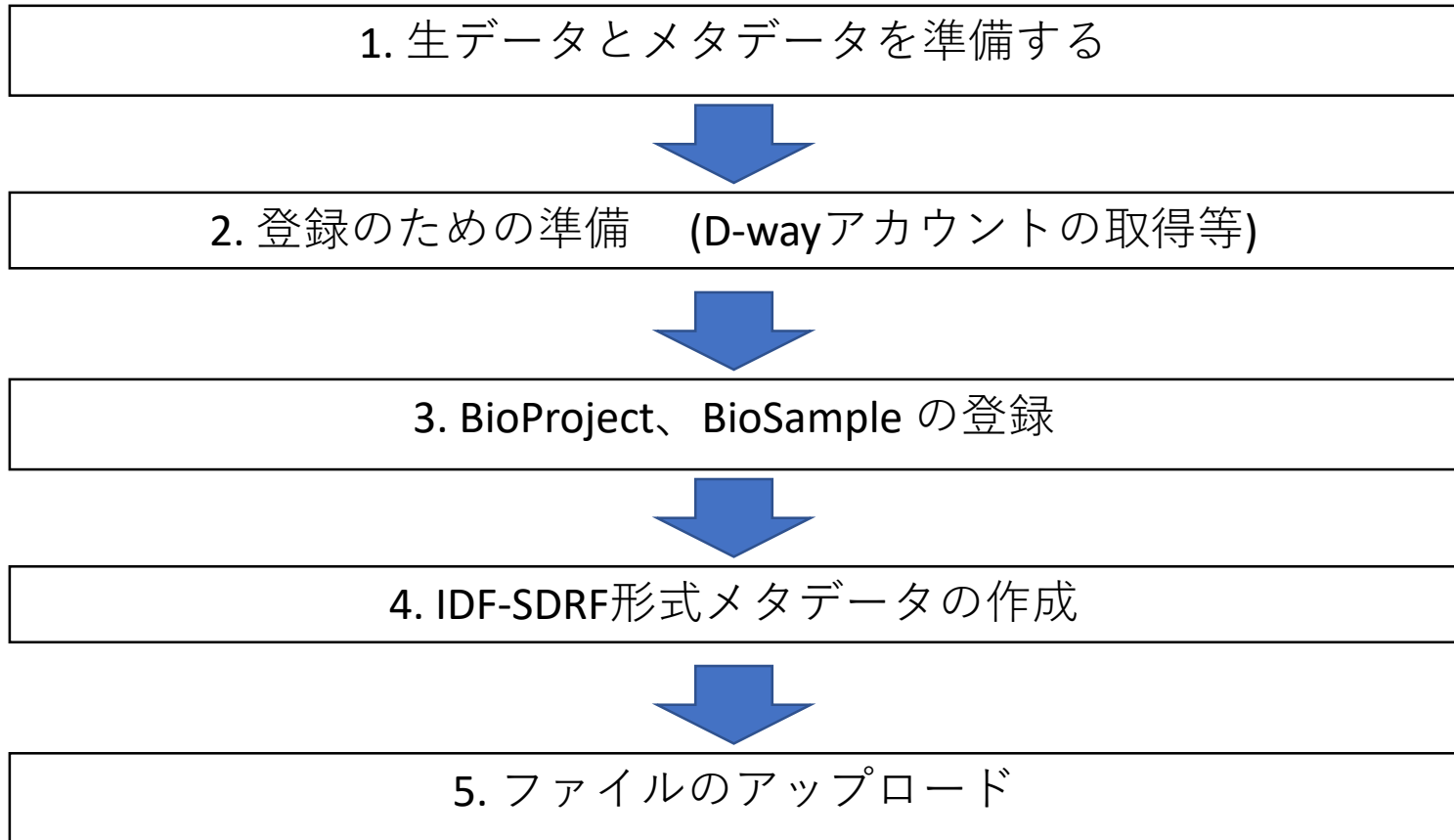
2. 登録申し込み

[MetaboBank 登録申し込みフォーム](#) [☒](#) から登録を申し込みます。申し込み内容に応じて担当者が個別に登録方法をご案内します。

フォームにアクセスできない場合、[申し込みファイル](#)をダウンロードし、必要事項を記入のうえ metabobank@ddbj.nig.ac.jp にメールで申し込んでください。

3. BioProject の登録

登録の流れ



初心者の場合、いくつかのステップに難所あり。(PuTTYでの鍵の作成、WinSCPを用いたファイル転送など)

ただ、チュートリアル等は充実しているので、ウェブの説明を丁寧に読んでもらえば大丈夫なはず。

BioProject, BioSample は全データ共通

BioProject

- DDBJ BioProject Handbookに従い、計画概要を登録

BioSample

- DDBJ BioSample Handbookに従い、情報を登録する
- 測定したサンプルの来歴（種、使用部位、発生段階、生育時の処理等）

	A	B	C	D	E	F	G	H	I	J	K	L
	biosample_accession	*sample_name	*sample_title	description	*organism	taxonomy_id	bioproject_id	tissue_type	sample_type	serotype	serovar	sex
1	SAMDC				Homo sapiens	9606	PSUBC	plasma	tissue sample			
2	SAMDC				Homo sapiens	9606	PSUBC	plasma	tissue sample			
3	SAMDC				ens	9606	PSUBC	plasma	tissue sample			
4	SAMDC				ens	9606	PSUBC	plasma	tissue sample			
5	SAMDC				ens	9606	PSUBC	plasma	tissue sample			
6	SAMDC				ens	9606	PSUBC	plasma	tissue sample			
7	SAMDC				ens	9606	PSUBC	plasma	tissue sample			
8	SAMDC				Homo sapiens	9606	PSUBC	plasma	tissue sample			
9	SAMDC				Homo sapiens	9606	PSUBC	plasma	tissue sample			
37	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
38	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
39	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
40	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
41	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
42	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
43	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
44	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
45	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			
46	SAMDC				Homo sapiens	9606	PSUBC	feces	tissue sample			

入力はタブ区切りtxt
(編集にExcelを使用できる)

MetaboBankに必要なメタデータ

必須

1. **STUDY:** 測定情報。IDF (Investigation Description Format) に記載
実験デザイン、登録者情報、プロトコールなど。
2. **ASSAY:** データとの関係。SDRF (Sample and Data Relationship Format) に記載
試料毎に、細胞タイプや測定装置・条件の詳細、対応する生データ。

強く推奨するものの、必須ではない要素

3. **Metabolite Annotation File:** いわゆるピークテーブルあるいはデータ行列
同定結果、 m/z 、イオンタイプ、同定基準等を記載
(*MassBank*に登録されていた内容は、これに相当)

キュレータとやり取りしながら登録。査定期間はおおよそ2週間程度。

同定される代謝物のリストは最重要事項ではない！

IDF, SDRF用のエクセルファイル

MetaboBank

- Home
- Submission
- FAQ
- Search
- Download
- Contact



MAGE-TAB
メタデータエクセル
IDF
SDRF

このメニューで
「metadata」を
選択

下にいくと例が
あります

ホーム > [metabobank](#) > メタデータ

メタデータ

MAGE-TAB

MicroArray Gene Expression Tabular (MAGE-TAB) は機能ゲノミクスデータを構造化・標準化された方法で記述するために開発された形式で、ArrayExpress と GEA で使用されています。MAGE-TAB はプロテオミクス分野でも使われ始めており、オミックス分野における国際標準になりつつあります。

MAGE-TAB は研究全体を記述する IDF、および、サンプルとデータファイルの関係性を記述する SDRF から構成されています。IDF と SDRF はプロトコルで、メタデータとデータファイルは SDRF を介してリンクします。

IDF

Study Title	Title ...		
Person Last Name	Makino	Fuji	
Person First Name	Hirotaka	Tero	
PubMed ID	20156332		
Protocol Name	Sample collection	Extraction	--
Protocol Parameters		Post extraction, Derivatization	

SDRF

Source Name	Characteristics (Tissue)	Protocol REF	Protocol REF	Parameter Value (Post extraction)	Raw Data File	Processed Data file	Metabolite Assignment File
Sample 1	leaf	Sample collection	Extraction	chloroform/methanol 1:1 (w/v)	raw1.ncML	processed1.xls	maf.tsv
Sample 2	leaf	Sample collection	Extraction	chloroform/methanol 1:1 (w/v)	raw2.ncML	processed2.xls	maf.tsv
Sample 3	root	Sample collection	Extraction	chloroform/methanol 1:1 (w/v)	raw3.ncML	processed3.xls	maf.tsv



IDF, SDRF, 生・解析済みデータファイル・MAF の関係

IDF, SDRF用のエクセルファイル

- Mass spectrometry, chromatography
 - Liquid chromatography-mass spectrometry (LC-MS, [download](#))
 - Liquid chromatography, diode array detector-mass spectrometry (LC-DAD-MS, [download](#))
 - Gas chromatography-mass spectrometry (GC-MS, [download](#))
 - Two dimensional gas chromatography-mass spectrometry (GC/GC-MS, [download](#))
 - Gas chromatography, flame ionization detector-mass spectrometry (GC-FID-MS, [download](#))
 - Capillary electrophoresis-mass spectrometry (CE-MS, [download](#))
- Mass spectrometry, direct injection
 - Direct infusion-mass spectrometry (DI-MS, [download](#))
 - Flow injection analysis-mass spectrometry (FIA-MS, [download](#))
 - Matrix-assisted laser desorption-ionisation mass spectrometry (MALDI-MS, [download](#))
- Mass spectrometry imaging (MSI, [download](#))
- Nuclear magnetic resonance spectroscopy (NMR, [download](#))

IDF: Investigation Description Format

	A	B	C	D
1	# Study (IDF) fields			
2	# Submission type: MS Chromatography - GC (GC-MS)			
3				
4	# Fields without line boxes can have multiple values in corresponding columns.			
5	MAGE-TAB Version	1.1		
6	Comment[MetaboBank accession]			
7	Study Title	Metabolomic Characterization of Knockout Mutants in Arabidopsis: Development of a Metabolite Profiling Database for Knockout Mutants		
8	Study Description	Despite recent intensive research efforts in functional genomics, the functions of only a limited number of Arabidopsis (<i>Arabidopsis thaliana</i>)		
9				
10	Experimental Design	genotype design	<div style="border: 1px solid black; padding: 5px; text-align: center;"> 必須項目(項目名のセルがオレンジ色)を中心に入力。項目によって選択肢あり。 </div>	
11	Experimental Factor Name	ecotype		
12	Experimental Factor Type			
13				
14	Person Last Name	Takahashi	Fukushima	Saito
15	Person First Name	Mikiko	Atsushi	Kazuki
16	Person Mid Initials			
17	Person Email	mikiko.takahashi@riken.jp	atsushi.fukushima@a.riken.jp; afukushima@kpu.ac.jp	kazuki.saito@riken.jp
18	Person Affiliation	RIKEN Center for Sustainable Resource Science	RIKEN Center for Sustainable Resource Science; Department of Applied Chemistry, Faculty of Engineering, Kansai University	RIKEN Center for Sustainable Resource Science
19	Person Roles	Submitter	Submitter	Submitter
20				
21	PubMed ID	24828308		
22	Publication DOI			
23				
24	Protocol Name	Sample collection	Extraction	Chromatography
25	Protocol Type	Sample collection	Extraction	Chromatography
		Sterilized seeds were stratified at 5 degreeC for 2 d, and were successively sown on Murashige and Skoog (MS) medium containing 1% sucrose. Plants were cultivated in controlled growth chambers at 22	Each sample was extracted with a concentration of 25 mg fresh weight of tissues per ml extraction medium (methanol/chloroform/water [3:1:1 v/v/v]) containing 10 stable isotope reference compounds using a Retsch	Column name: 30 m x 0.25 mm inner diameter fused-silica capillary column with a chiral stationary phase (CSP) bonded on 0.25 microliter film D

SDRF: Sample & Data Relationship Format

A	B	C	D	E	F	G
1 # Assay (SDRF) columns						
2 # Submission type: MS Chromatography - GC (GC-MS)						
3						
Source Name	Characteristics[biosample_accession]	Characteristics[organism]	Characteristics[ecotype]	Characteristics[isolate]	Characteristics[tissue]	Characteristics[light_cycle]
Alkane mixture		unidentified				
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col-0		shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col-0		shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col-0		shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col	aba1-5	shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col	aba1-5	shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col	aba1-5	shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col	aba1-5	shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col	aba1-5	shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col	aba1-5	shoot system	16hr (daylight time)/8hr (dark time)
Arabidopsis thaliana_shoot system		Arabidopsis thaliana	Col	aba2-3	shoot system	16hr (daylight time)/8hr (dark time)

BioSampleのaccession番号を入力し、紐づける
 (このプロジェクトはBioSample未登録のため
 空欄)

Characteristics[light]	Sample Name	Protocol REF	Parameter Value[Post extraction]	Parameter Value[Derivatization]	Extract Name	Protocol REF	Parameter Value[Chromatography instrument]
	Alkane_2	Extraction			Alkane_2	Chromatography	Agilent 6890N Gas Chromatograph (Agilent Te
approximately 80 phol	Cont_02_1	Extraction		methoximation and silylation	Cont_02_1	Chromatography	Agilent 6890N Gas Chromatograph (Agilent Te
approximately 80 phol	Cont_03_1	Extraction		methoximation and silylation	Cont_03_1	Chromatography	Agilent 6890N Gas Chromatograph (Agilent Te
approximately 80 phol	Cont_04_1	Extraction		methoximation and silylation	Cont_04_1	Chromatography	Agilent 6890N Gas Chromatograph (Agilent Te
approximately 80 phol	Cont_05_1	Extraction		methoximation and silylation	Cont_05_1	Chromatography	Agilent 6890N Gas Chromatograph (Agilent Te
approximately 80 phol	Cont_06_1	Extraction		methoximation and silylation	Cont_06_1	Chromatography	Agilent 6890N Gas Chromatograph (Agilent Te

サンプル調製や測定機器の情報を記載す
 る

必須項目(項目名のセルがオレンジ
 色)を中心に入力

SDRF つづき

AH	AI	AJ	AK	AL	AM	実験因子となる項目を記載				
1										
2										
3										
4	Comment[Raw Data File	Comment[Raw Data File md5]	Protocol REF	Processed	Comment[Proc	Protocol REF	Metabolite Assign	Comment[maf_value_unit]	Factor V	
5	Alkane_2.cdf	fc33a8d3072db88a0d5569e82d3412f9	Data processing			Metabolite identification				
6	Cont13_1.cdf	bf020534f1a1336842c010e446ae2251	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col-0
測定生データのファイル情報を記載						Metabolite identification	MTBKS2.MAF.txt	AU		Col-0
						Metabolite identification	MTBKS2.MAF.txt	AU		Col-0
						Metabolite identification	MTBKS2.MAF.txt	AU		Col-0
						Metabolite identification	MTBKS2.MAF.txt	AU		Col-0
11	Cont_01_1.cdf	3632c73124af68da6b10d4854068910b	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col
12	Cont_02_1.cdf	a9ade0024209c8dd4560fefdfab4055	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col
13	Cont_03_1.cdf	abe3c221813de994bfc0ebb9d93ee288	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col
14	Cont_04_1.cdf	b256c63ed11b6358f448dc6cd7aefc61	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col
15	Cont_05_1.cdf	8a68868aa90be09f8df1227ade464d1e	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col
16	Cont_06_1.cdf	db2a9cde21d9edc018180869dd4fa805	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col
17	Cont_07_1.cdf	0089d72605c6f30bdc6602501412a324	Data processing			Metabolite identification	MTBKS2.MAF.txt	AU		Col

公開できるアノテーション結果がある場合は、Metabolite assignment file (MAF)の情報を記載。

MAFの例 (テンプレートのExcelファイルに別途記載)

Metabolite Assignment File) - MS										T	U	V	W	
Identifier	chemical_formula	smiles	inchi	metabolite_identification	metabolite_class	mass_to_charge	fragmentation	modifications	ch	# From here, enter sample names and measured v	R00_1-Neg	R00_2-Neg	R00_3-Neg	R01_1-N
C9H10CINO2		N[C@@H](CC1=CC=CC2=Chloro-L-phenylalanine		Phenylalanine and derivat		181.0133	137.02408:28 181	[M-H]-NH3]		223	295	220		
C9H16O4		O=C(O)CCCCCCCC(=C Azelaic acid		Dicarboxy fatty acids		187.0972	120.46558:16 125	[M-H]-		811	828	700		
C6H8O7		O=C(O)CC(O)(C(=O)O) Citric acid		Tricarboxylic acids and de		191.0189	85.02881:66 87.0	[M-H]-		1859	2092	1881		
C10H8O4		COC1=C(O)C=C2OC(=6-Methoxy-7-hydroxycoumarin		7-hydroxycoumarins		191.034	81.88715:16 104.	[M-H]-		3947	3379	3685		
C10H8O4		COC1=C(O)C=C2OC(=6-Methoxy-7-hydroxycoumarin		7-hydroxycoumarins		191.0341	104.0271:214 104	[M-H]-		2999	2652	2822		
C10H8O4		COC1=C(O)C=C2OC(=6-Methoxy-7-hydroxycoumarin		7-hydroxycoumarins		191.0342	52.73787:31 104.	[M-H]-		14151	18316	17233		
C10H8O4		O=C1OC2=CC(O)=C(O) Scopoletin		Coumarin and derivatives		191.0345	104.02554:18 120	[M-H]-		1091	904	840		
C10H10O4				droxycinnamic acids		193.0494	134.03557:52 135	[M-H]-						
C10H10O4				ulic acid and derivative		193.0496	134.03284:109 13	[M-H]-						
C10H12O4				yl-phenylketones		195.0654	96.95388:18 149.	[M-H]-						
C11H12N2O2		O=C(O)C(N)CC2=CNC1 Tryptophan		Indolyl carboxylic acids ar		203.0818	74.02122:19 116.	[M-H]-						
C10H8O3S		OS(=O)(=O)C1=CC2=C1 2-Naphthalenesulfonic acid; LC-ESF-2-naphthalene sulfonates		2-naphthalene sulfonates		207.0113	135.04758:19 143	[M-H]-						

検出された化合物の情報

各測定サンプルにおける検出量

なぜ共有する？ 再解析のメリット

MTBKS104 (トマト)

検出ピーク数: 2647→5800

MSI level3以上ピーク数: 256→1084 (4.2倍)

MSI level3以上ピーク数 / 検出ピーク数割合: 9.6%→18.6%

新たに判明した化合物の例

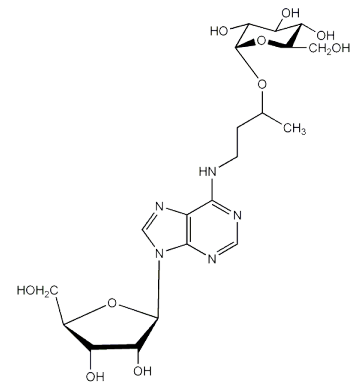
1. Hydroxypentadecanoyl-CoA, putative

水酸基の位置は複数あって詳細は不明 ただし CoA 化合物をメタボロ
でみられることは重要

2. Dihydrozeatin riboside-O-glucoside, putative

3. beta2-Tomatine, putative

サイトカイニン (植物ホルモン) や、アルカロイド
の測定につながる



登録よろしくお願ひします

MetaboBank リポジトリ

DDBJサービスの一環として、EBI MetaboLightsと連携したサービスを実施していきます。

<https://www.ddbj.nig.ac.jp/metabobank/>

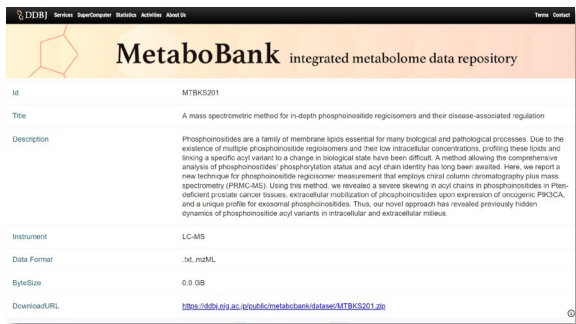
データ登録・公開までの内部フロー

BioProject, BioSample を最初に登録。IDF, SDRFとの重複もあるため、キュレータが相談に応じます。

外部ユーザ



DDBJ内の処理



IDF, SDRFエクセルに記載。

BioProject/BioSample 登録
MTBKS201～ 番号発行

IDF/SDRF tsv
データファイルの受け取り



公開用 RDF に変換
mb2 検索系に取り込み。エンバーゴあり



BioProject/BioSample 公開、NCBI/EBI と共有
BioProject には、MetaboBank ftp のリンクを追加
MetaboBankの検索系、およびMetaboBank Wiki に記載

再解析の現状

MTBKS102 (シロイヌナズナ)

検出ピーク数: 773→2071

MSI level3以上ピーク数: 79→353 (4.4倍)

MSI level3以上ピーク数 / 検出ピーク数割合: 10.2%→17%

新たに判明した化合物の例

1. p-Coumaroyl-CoA, putative

カルコンの部品で存在はわかっているが、通常見えにくい

2. Neoglucobrassicin, putative

3. Indolylmethylthiohydroximate, putative

グルコシノレートで存在は既知
ただしメタボロミクスで見えると
わかることは重要

