

# MetaboBank

## サイトビジット 2019 0712

有田正規 遺伝研  
金谷重彦 NAIST  
櫻井望 遺伝研  
平川英樹 かずさDNA  
(福島敦史 理研CSRS)



©2019 有田 正規(情報・システム研究機構) licensed under CC 表示 4.0 国際 ライセンス

## メンバー紹介

遺伝研: 有田 (MassBank)、櫻井 (メタデータ、サーバ)、川島、多田

直接雇用: 時松敏明 (メタデータ項目、BioSample摺合)

佐藤充治 (代謝マップエディタ)

奈良先: 金谷 (代謝マップ)、小野、他ラボメンバー

直接雇用: 森田晶(システム)、大橋美名子(補佐)、ほか学生

かずさ: 平川(統括)、櫻井(食品データ)

直接雇用: 長崎英樹(MassBase整理)、大澤祥子(補佐)

理研CSRS: 福島(PMM作成)、津川、小林

直接雇用: 高橋みき子(PRIMe整理)、ほか学生

発表は有田(10)、金谷(10)、櫻井(10)、有田(10)

# 開始時の問題点と目標

## データ再利用の難しさ

- ・ 生データの取得法、フォーマットが多様  
→国際標準化(Lipidomics)、ABFフォーマットの採用
- ・ データ統合が困難  
→CSRSかずさは統合、EBIと連携、BioSample等の利用

## 最終目標

- ・ 論文投稿時に登録するリポジトリ = 恒久的なID付与
- ・ 統合解析のできるデータサーバ
  - 生データに解析結果や測定法がぶら下がる構造 = 他DBとの連携  
(例: Metabolonoteの分類法)
  - 解析結果(例えばスペクトル、化合物リスト)の検索 = MassBank
  - データ登録者へのリワード

# 2020年までのマイルストーン

いずれも国際標準が前提

遺伝研： 恒久的なデータ・リポジトリの開設（遺伝研アドレス）

初期データ： 理研CSRSおよびかずさDNA研のデータ

（MassBase, DropMet, Komics）

投稿データ： 共同研究先を優先

プロジェクトデータ： AMED微生物叢領域との連携

かずさ： 保有するメタボロームデータの再解析

PowerGet等、ソフトウェアのアップデート 遺伝研でリリース

理研： 保有するメタボロームデータの再解析 理研でリリース

奈良先： 生体活性オントロジー構築、メタ代謝マップ構築

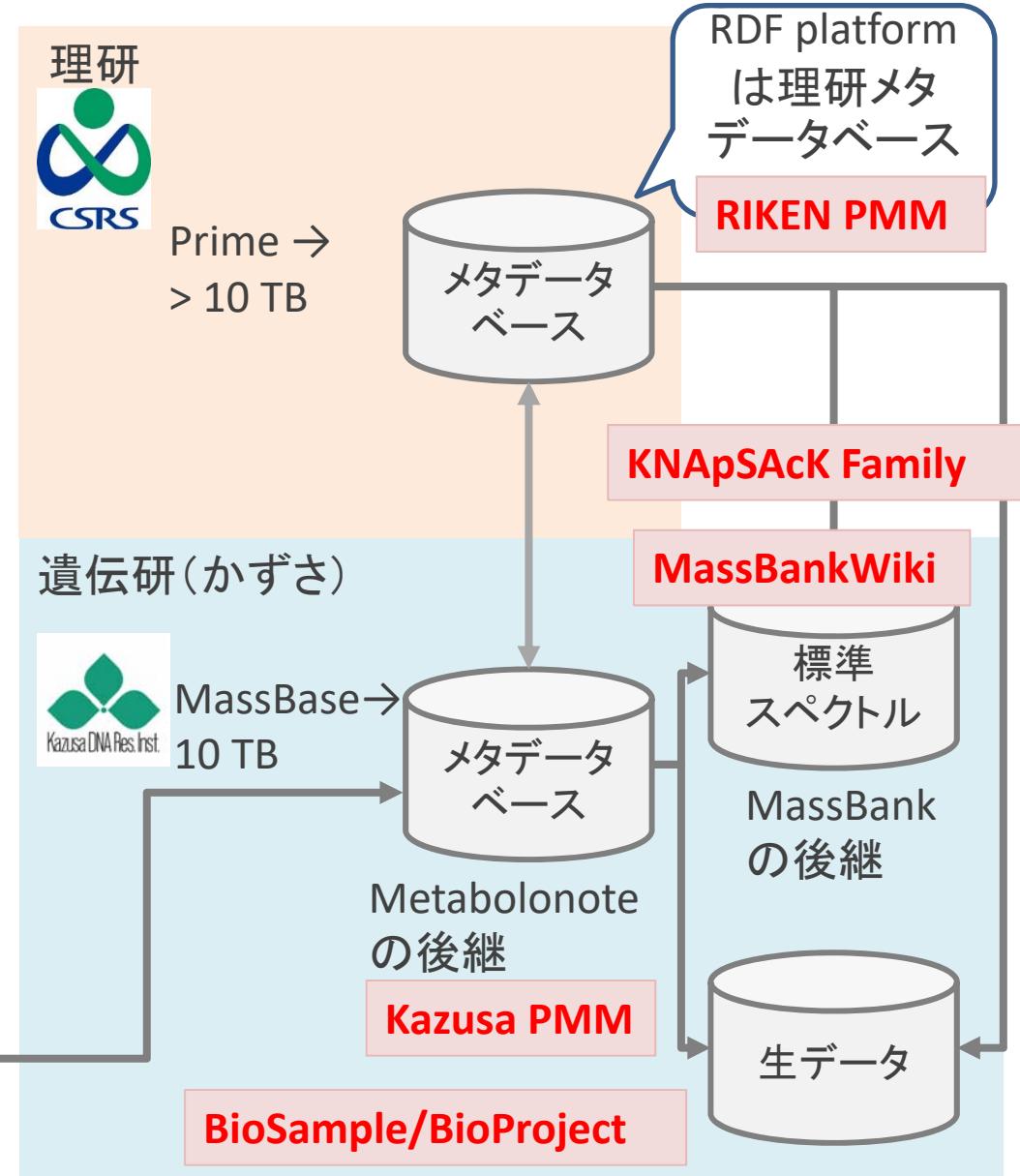
# MetaboBank 全体構想

欧州のMetaboLights, 米国のMetabolomics Workbenchに並ぶ、生データのリポジトリ。  
DDBJサービスとして開設。  
平成31～2年度には、一般からの投稿を受付。

標準スペクトルのサーバは、MassBank wikiという名前でも継続

データはOmics DI ([omicsdi.org](http://omicsdi.org))およびmetabolomeXchangeに登録。

今年度は植物のメタデータ作成と、フォーマット決定。  
来年度以降、医薬学のメタボロミクスへ拡張



# MassBank GitHub Record

MassBank.eu側がまず登録

The screenshot shows a search interface with a grey header bar containing the text "MassBank.eu" and "Search". Below the header is a search bar with the placeholder "Search" and a blue "Go" button. The main area contains a table with two columns of institution names and their counts.

Institution	Count
<a href="#">IPB_Halle</a>	(674)
<a href="#">JEOL_Ltd</a>	(44)
<a href="#">KWR</a>	(207)
<a href="#">Kazusa</a>	(273)
<a href="#">Keio_Univ</a>	(4768)
<a href="#">Kyoto_Univ</a>	(184)
<a href="#">Literature_Specs</a>	(39)
<a href="#">MPI_for_Chemical_Ecology</a>	(691)
<a href="#">MSSJ</a>	(177)
<a href="#">MetaboLights</a>	(58)
<a href="#">Metabolon</a>	(149)
<a href="#">NAIST</a>	(621)
<a href="#">NaToxAq</a>	(774)
<a href="#">Nihon_Univ</a>	(706)
<a href="#">Osaka_MCHRI</a>	(20)
<a href="#">Osaka_Univ</a>	(447)
<a href="#">Pfoss_research_group</a>	(413)
<a href="#">RIKEN</a>	(11935)
<a href="#">RIKEN_IMS</a>	(1140)
<a href="#">RIKEN_NPDepo</a>	(1956)
<a href="#">Tottori_Univ</a>	(16)
<a href="#">UFZ</a>	(1002)
<a href="#">UOEH</a>	(35)
<a href="#">UPAO</a>	(12)
<a href="#">Univ_Connecticut</a>	(510)
<a href="#">Univ_Toyama</a>	(252)
<a href="#">Washington_State_Univ</a>	(2626)
<a href="#">Waters</a>	(2992)

MSSJデータ 177 にも対応

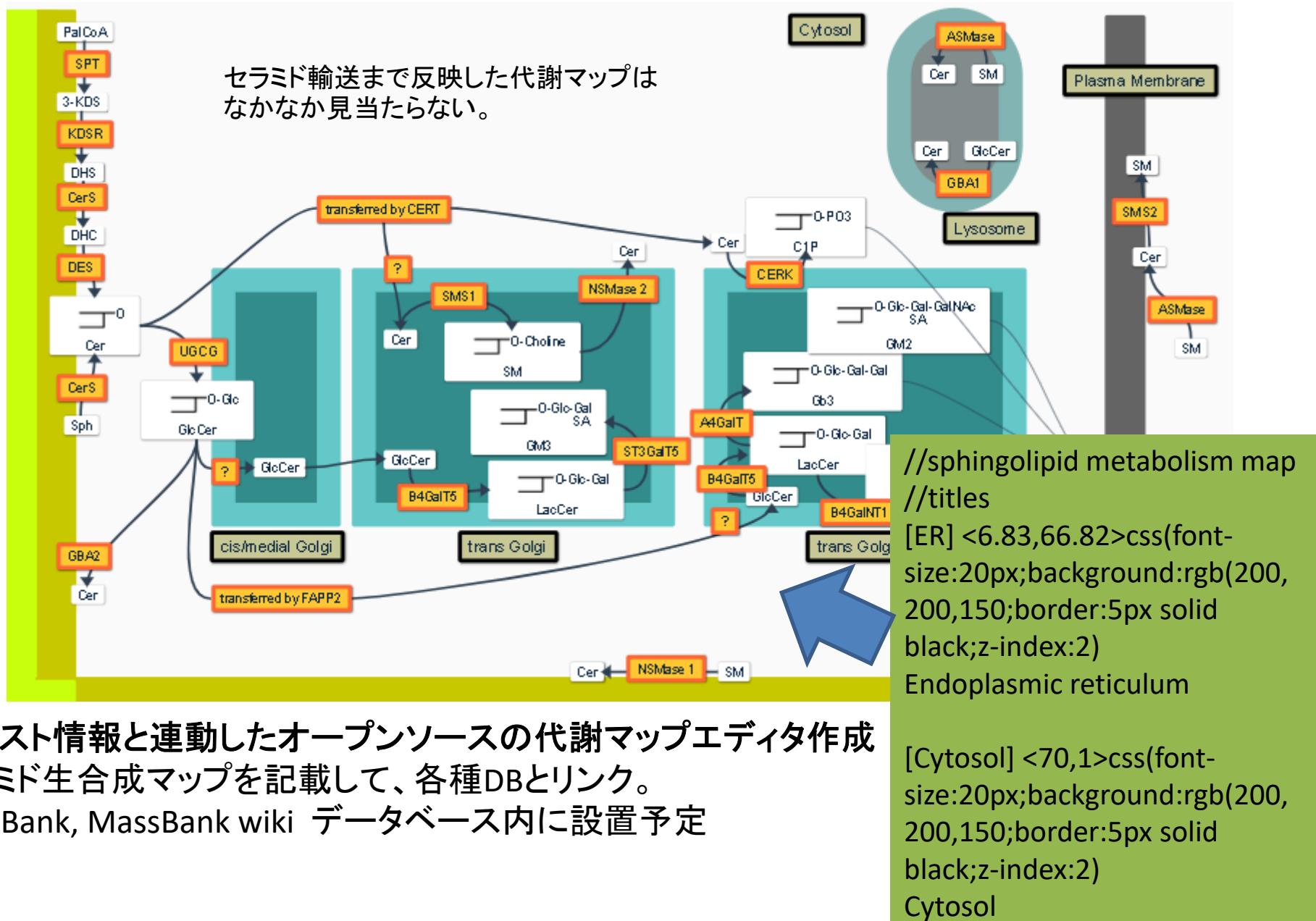
RIKEN PLASMA Data (5000程度)

RIKEN RESPECT Data (4000程度)  
→ mass calc

RIKEN IMS Lipid Data (1000程度)

NIG-Karolinska Data (800程度)

# 代謝動態データベース及びマップの構築



# 二次代謝生合成経路の悉皆的解析を目指したデータサイエンス : KNApSACk family DBの展開

Shigehiko Kanaya

Nara Institute of Science and Technology (NAIST)

Graduate School of Science and Technology

skanaya@gc.naist.jp



Last update 2019/4/1  
51,179 unique metabolites  
116,315 species-metabolite pairs  
Plant Cell Physiol.53, e1(1-12) (2012)

KNApSACk Metabolomics

3D Since 2012.11 Core System Since 2004.04 Search Engine Since 2008.12

Pocket Search for Functional Species

Food & Health	Crude Drug	Biology
YAKUZEN Lunch Box DietNavi FoodProcessor DietDish MARCHÉ	WorldMap KAMPO JAMU Tea Pot Skewered KNApSACk	Metabolite Ecology Distribution Biological Activity Natural Activity Metabolite Activity Metabolite Activity Twins 化合物類似度 データベース
YAKUZEN Lunch Box DietNavi FoodProcessor DietDish MARCHÉ	WorldMap KAMPO JAMU Tea Pot Skewered KNApSACk	Metabolite Ecology Distribution Biological Activity Natural Activity Metabolite Activity Metabolite Activity Twins 化合物類似度 データベース

Picnic Gene Annotation

Strap Correlation Coefficient

Pickaxe Metalloprotein Database

代謝データベース

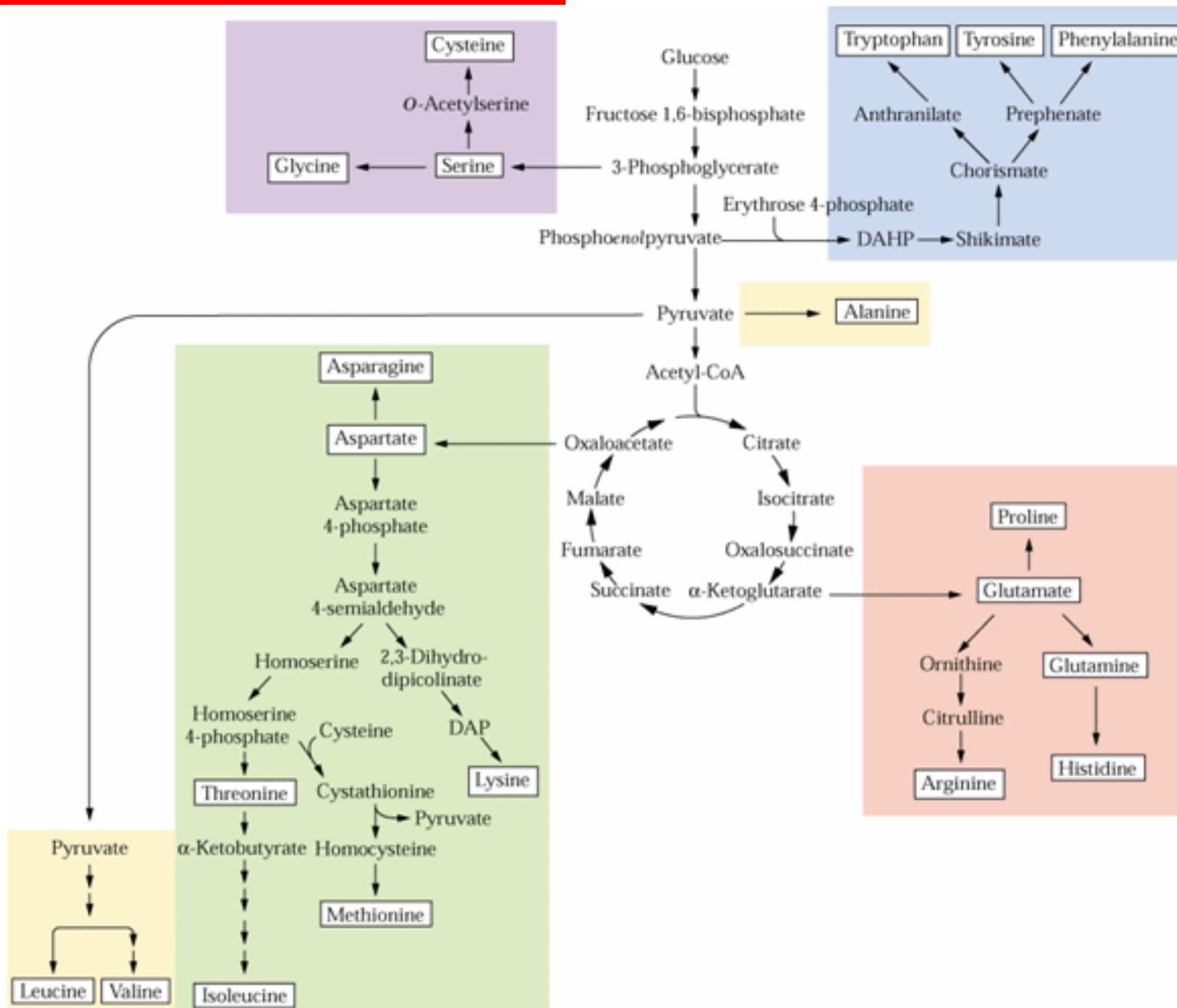
Motor Metabolic Path

Bicycle Algae Metabolic Pathway

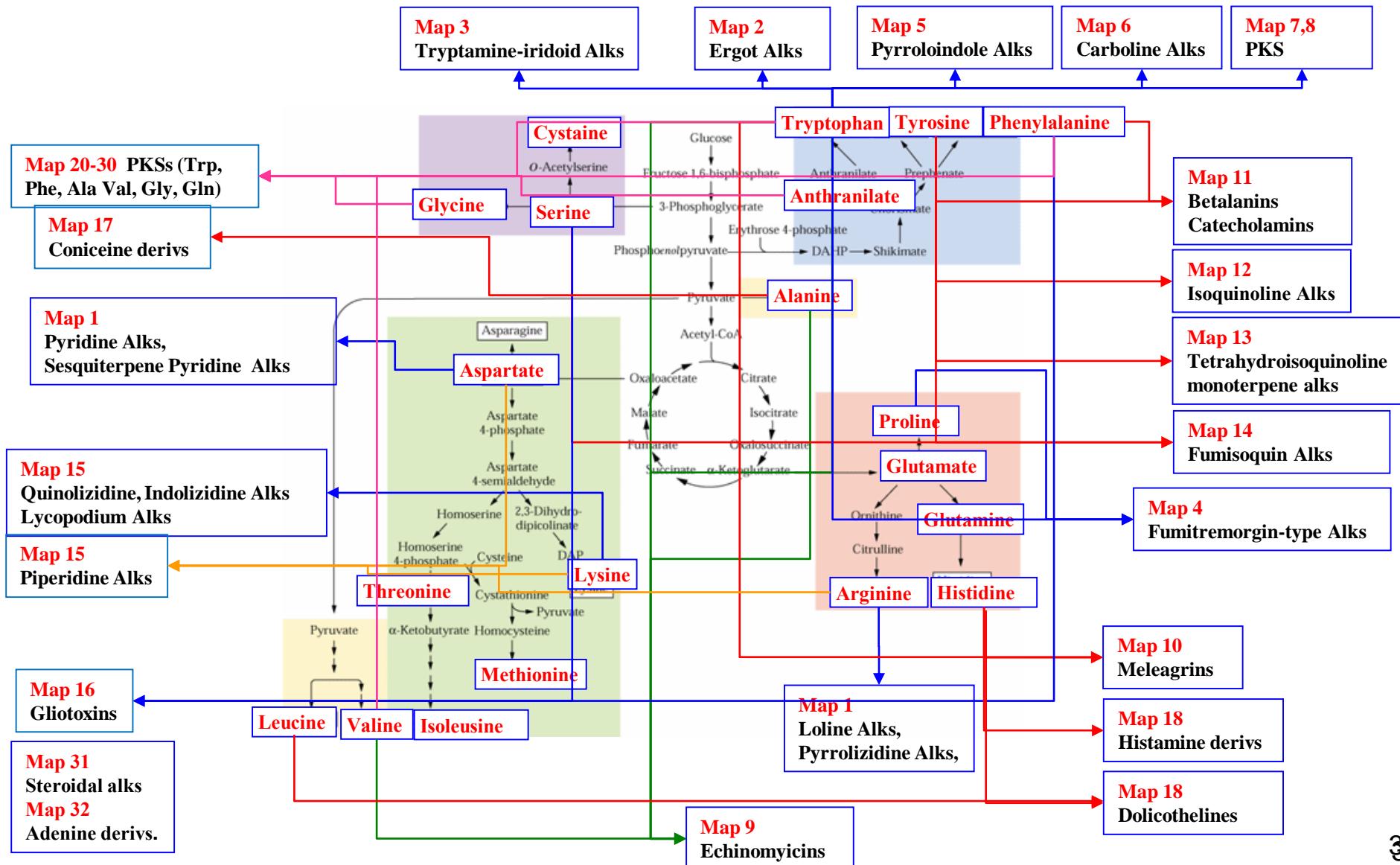
CobWeb Alkaloid Pathway

Instruction Manual (Japanese) | Instruction Manual (English) | Terms of service (Japanese) | 3305 days left!

# Primary Metabolic pathway: directly involved in normal growth, development, and reproduction



**Secondary Metabolism:** are not absolutely required for the survival of the organism.  
**Alkaloids** are biosynthesized by **primary metabolites** such as amino acids nucleotide, steroids, and **secondary metabolites** such as terpenoids..



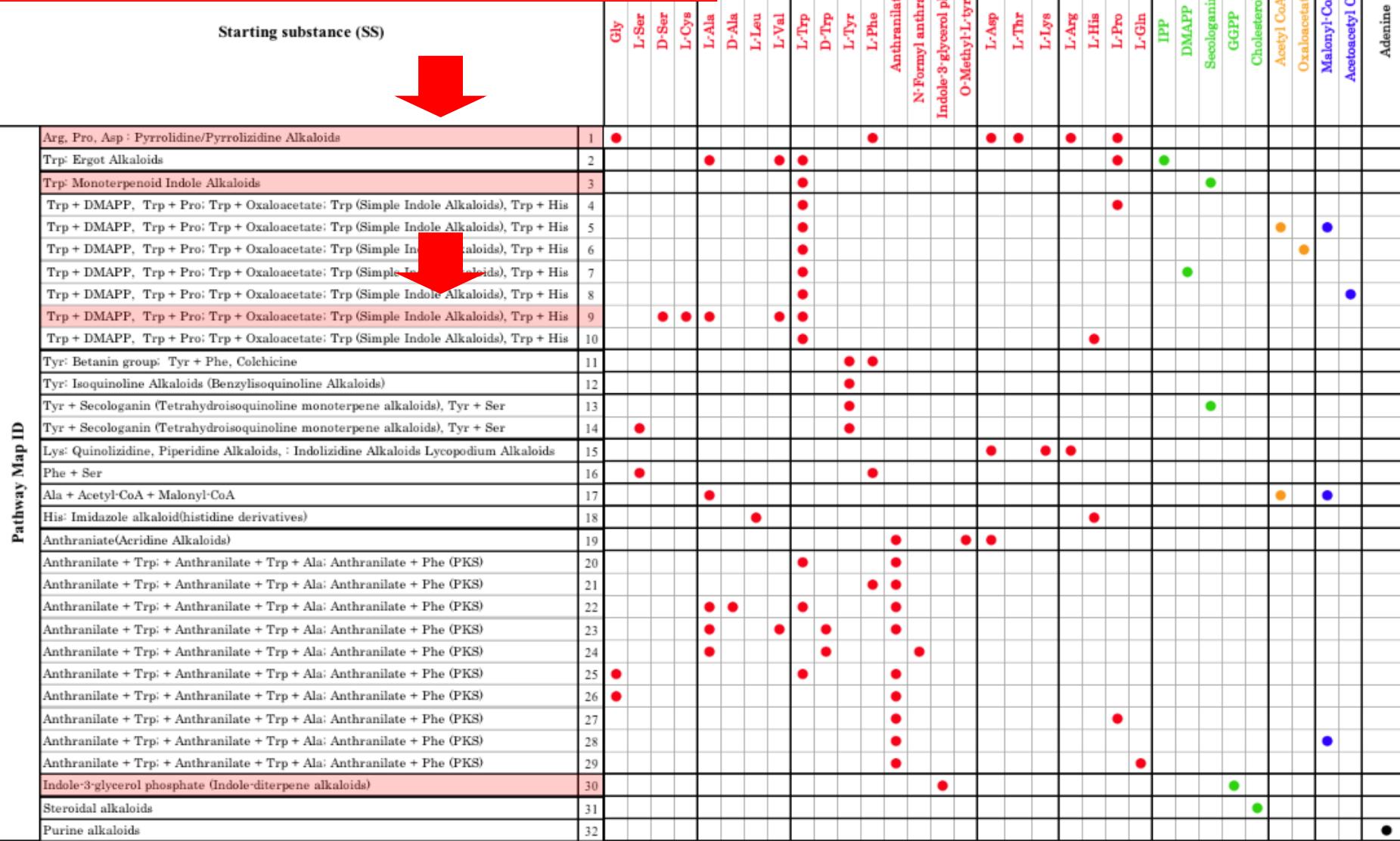


# “CobWeb”

Please input MetaboliteName. Multiple searches are possible for MetaboliteName. The delimiter is : (colon).

Please input SpeciesName. Multiple searches are impossible for SpeciesName.

Starting substance (SS)





# “CobWeb”

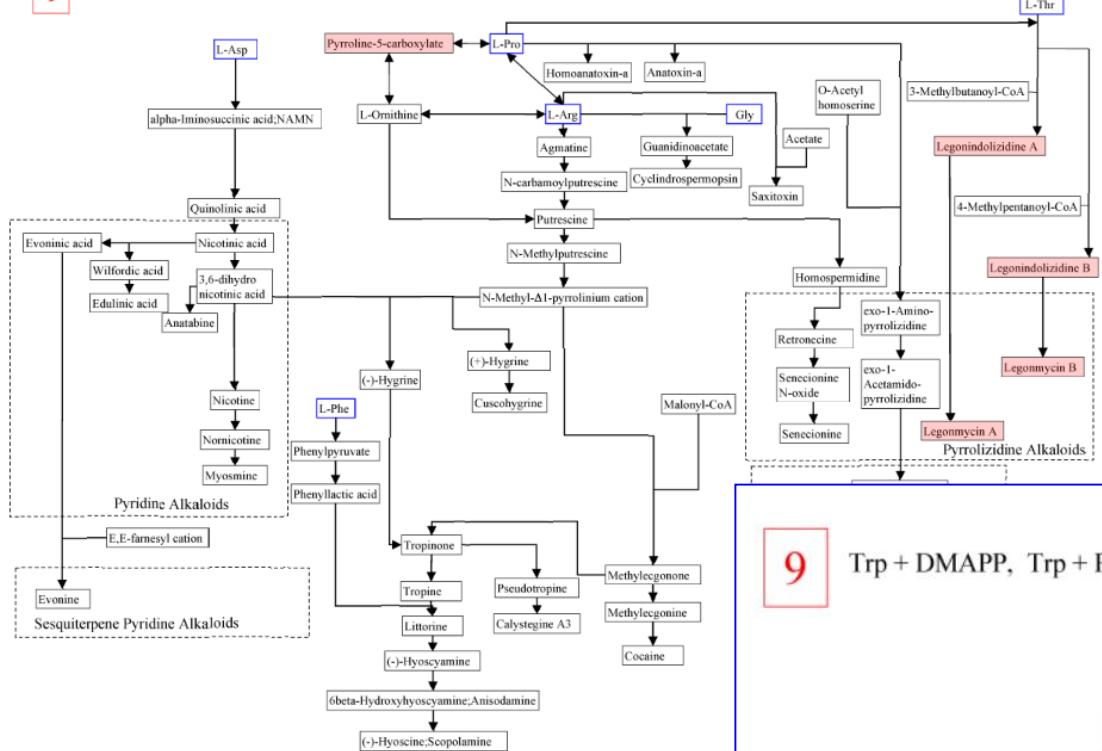
Please input MetaboliteName. Multiple searches are possible for MetaboliteName. The delimiter is : (colon).

Please input SpeciesName. Multiple searches are impossible for SpeciesName.

**Streptomyces**

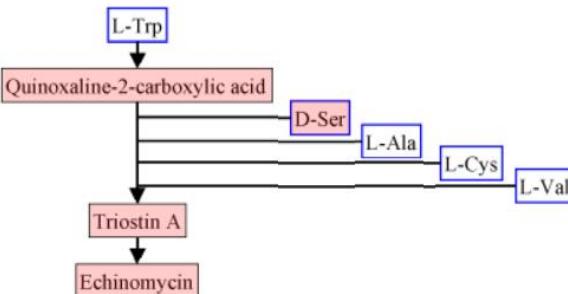
Pyruvate	Phosphoenol pyruvate	Oxaloacetate	alpha-Ketoglutarate	Terpenes	TCA cycle	Fatty acid	Nucleic acids
L-Phe Anthraniolate N-Formyl anthranilate Indole-3-glycerol phosphate	L-Pro L-Arg L-Thr L-Lys	Gly Guanidinoacetate Putrescine N-Methylputrescine N-Methyl-Δ1-pyrrolinium cation (-)-Hygrine (+)-Hygrine Cuscohygrine Malonyl-CoA	O-Acetyl homoserine Acetate Saxitoxin Homospermidine Retronecine Senecionine N-oxide Senecionine exo-1-Amino-pyrrolizidine exo-1-Acetamido-pyrrolizidine	Legonindolizidine A Legonindolizidine B Legomycin A Legomycin B Pyrolizidine Alkaloids	L-Asp Tr-Asp L-His L-Pro L-Gln IPP DMAPP Sicologanin Cholesterol Acetyl CoA Oxaloacetate Malonyl-CoA Acetoacetyl CoA	L-Arg L-His L-Pro L-Gln	Adenine
O-Methyl-L-Tyrosine							

1 Arg, Pro, Asp : Pyrrolidine/Pyrrolizidine Alkaloids



- Anthraniolate + Trp; + Anthraniolate + Trp + Ala; Anthraniolate + Phe (PKS)
- Anthraniolate + Trp; + Anthraniolate + Trp + Ala; Anthraniolate + Phe (PKS)
- Anthraniolate + Trp; + Anthraniolate + Trp + Ala; Anthraniolate + Phe (PKS)
- Anthraniolate + Trp; + Anthraniolate + Trp + Ala; Anthraniolate + Phe (PKS)
- Anthraniolate + Trp; + Anthraniolate + Trp + Ala; Anthraniolate + Phe (PKS)
- Anthraniolate + Trp; + Anthraniolate + Trp + Ala; Anthraniolate + Phe (PKS)
- Anthraniolate + Trp; + Anthraniolate + Trp + Ala; Anthraniolate + Phe (PKS)
- Indole-3-glycerol phosphate (Indole-diterpene alkaloids)
- Steroidal alkaloids
- Purine alkaloids

9 Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate; Trp (Simple Indole Alkaloids), Trp + His





# “CobWeb”

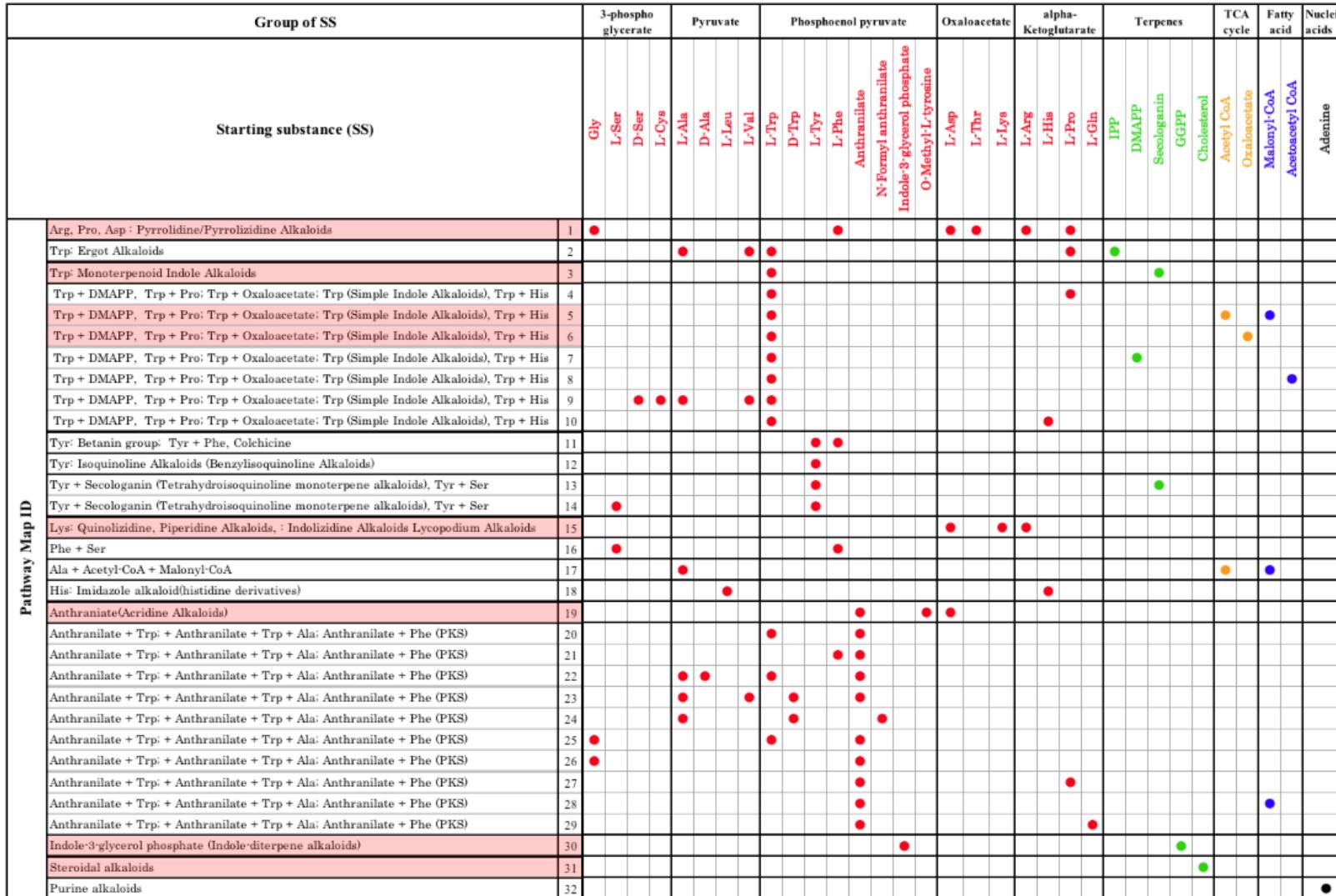
from KNApSack

Please input MetaboliteName. Multiple searches are possible for MetaboliteName. The delimiter is : (colon).

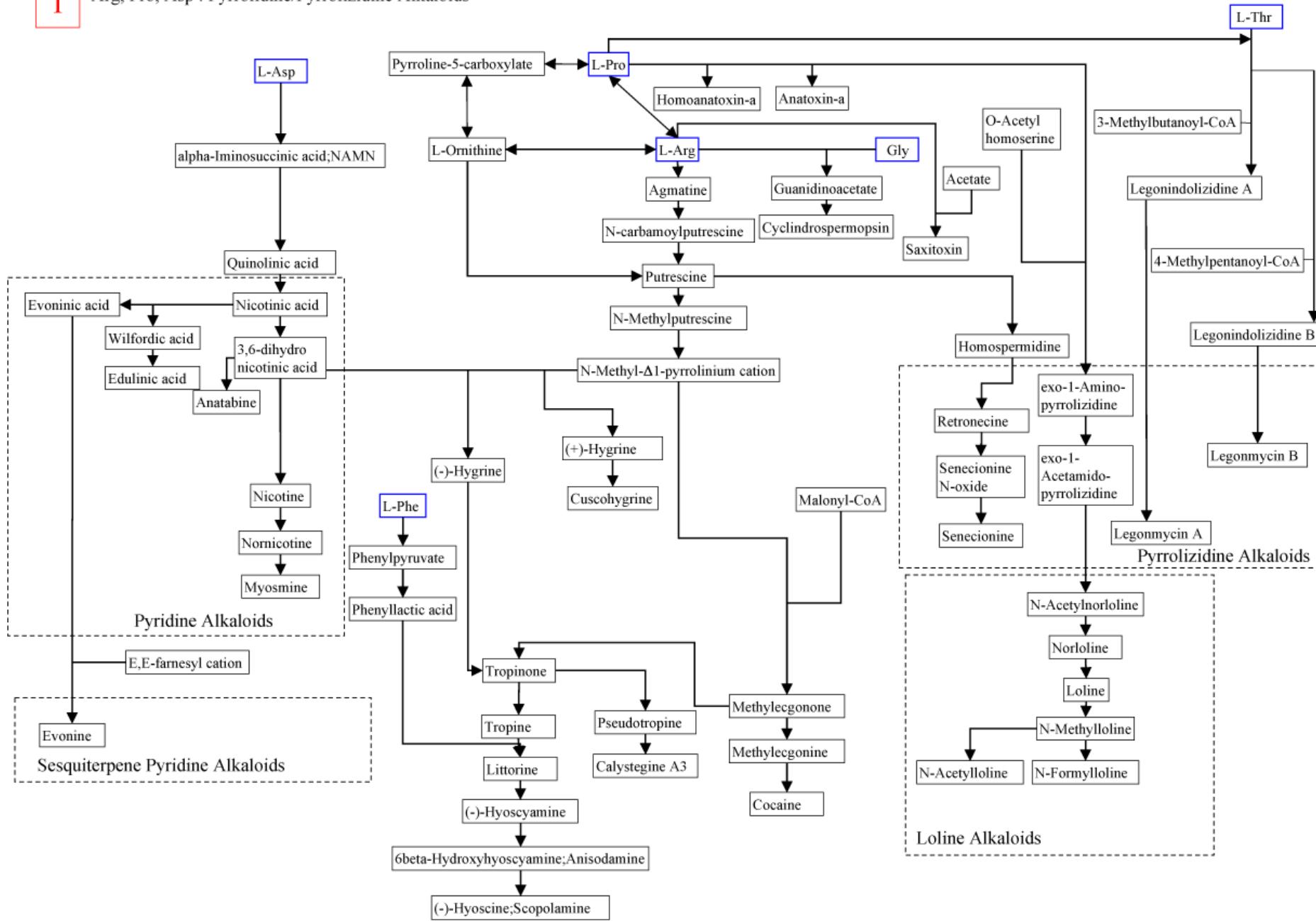
show Clear Page Clear

Please input SpeciesName. Multiple searches are impossible for SpeciesName.

Nicotiana  show Clear Page Clear



Pathway Map ID





# “CobWeb”

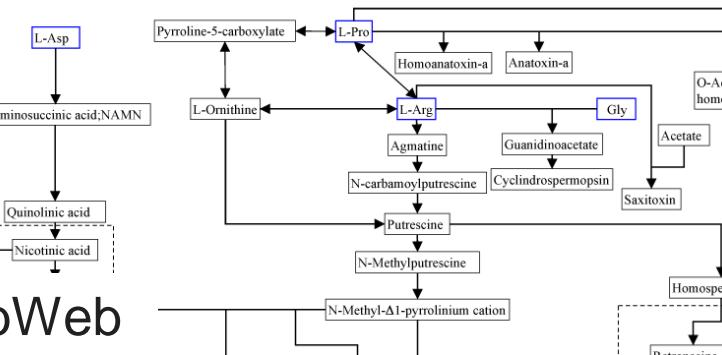
from KNAPSAck

Please input MetaboliteName.  
 show Clear Page Clear

Please input SpeciesName.  
 show Clear Page Clear

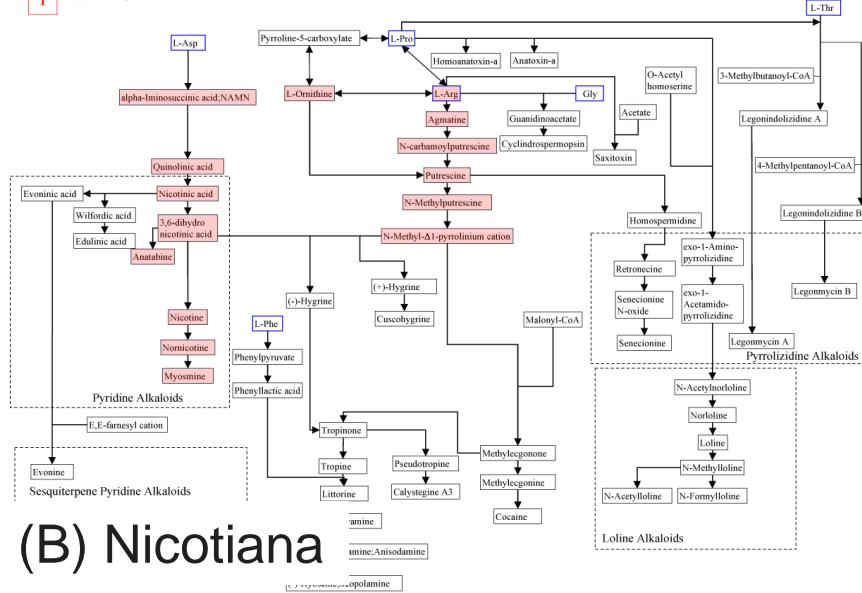
Input data :

1 Arg, Pro, Asp : Pyrrolidine/Pyrrolizidine Alkaloids



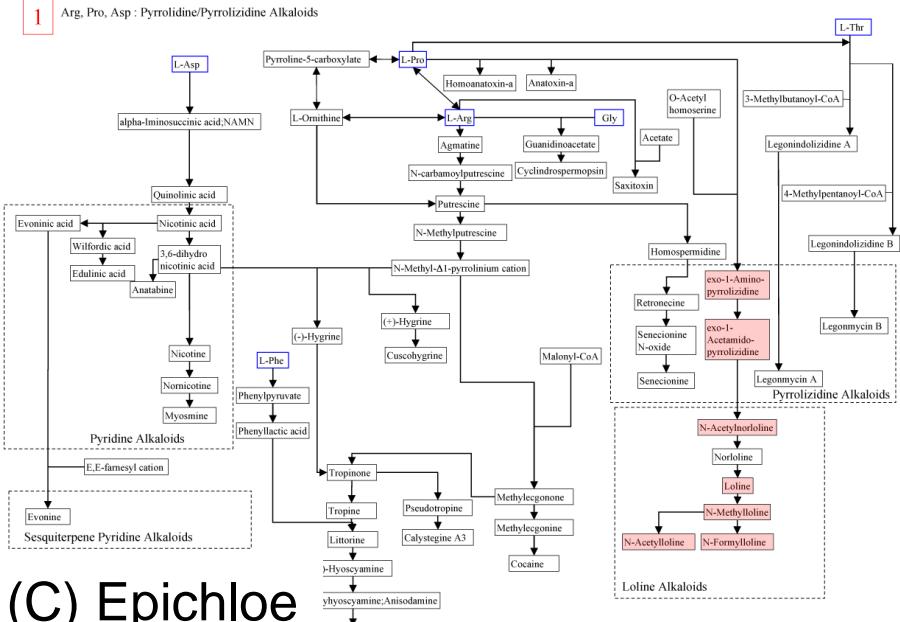
(A) CobWeb

1 Arg, Pro, Asp : Pyrrolidine/Pyrrolizidine Alkaloids



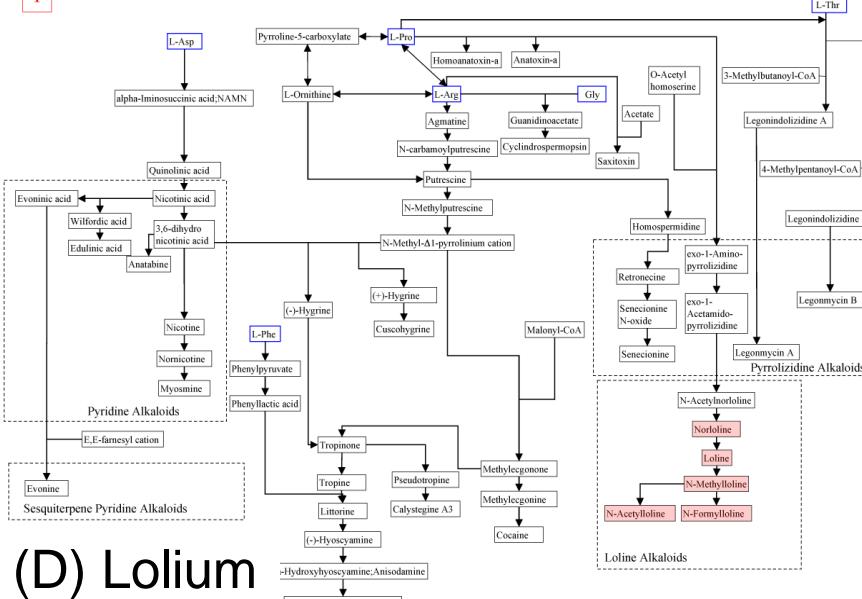
(B) Nicotiana

1 Arg, Pro, Asp : Pyrrolidine/Pyrrolizidine Alkaloids



(C) Epichloe

1 Arg, Pro, Asp : Pyrrolidine/Pyrrolizidine Alkaloids



(D) Lolium

Metabolite Information				Structural formula
Name	N-Acetylloline NAL			
Formula	C10H16N2O2			
Mw	196.12117777			
CAS RN	4914-36-7			
C_ID	C00023611  , 			
InChIKey	YIZSKLHCDNIMHK-YEXJWTGLNA-N			
InChICode	InChI=1S/C10H16N2O2/c1-6(13)11(2)9-8-5-12-4-3-7(14-8)10(9)12/h7-10H,3-5H2,1-2H3/t7-,8+,9-,10+/m0/s1			
SMILES	O1[C@H]2[C@H]3N(C[C@H]1[C@H]3N(C(C)=O)C)CC2			
Start Substs in Alk. Biosynthesis (Prediction)				
Organism	Kingdom	Family	Species	Reference
	Fungi	Clavicipitaceae	Epichloe coenophiala e19	<a href="#">Ref.</a>
	Fungi	Clavicipitaceae	Epichloe coenophiala e4163	<a href="#">Ref.</a>
	Fungi	Clavicipitaceae	Epichloe festucae E2368	<a href="#">Ref.</a>
	Fungi	Clavicipitaceae	Epichloe uncinata e167	<a href="#">Ref.</a>
	Fungi	Hypocreaceae	Acremonium coenophialum	<a href="#">Ref.</a>
	Plantae	Poaceae	Festuca arundinacea 	<a href="#">Ref.</a>
	Plantae	Poaceae	Lolium arundinaceum	<a href="#">Ref.</a>
	Plantae	Poaceae	Lolium cuneatum	<a href="#">Ref.</a>
	Plantae	Poaceae	Lolium pratense	<a href="#">Ref.</a>
	-	-	Neotyphodium pratense	<a href="#">Ref.</a>
	-	-	Neotyphodium siegelii	<a href="#">Ref.</a>
	-	-	Neotyphodium uncinatum	<a href="#">Ref.</a>
	-	-	Neotyphodium coenophialum	<a href="#">Ref.</a>
	-	-	Neotyphodium occultans	<a href="#">Ref.</a>

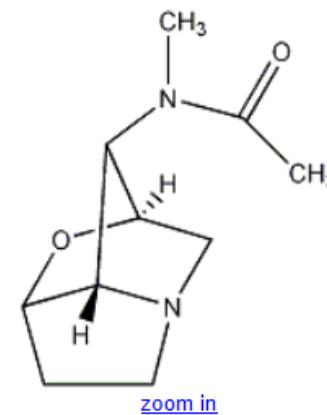
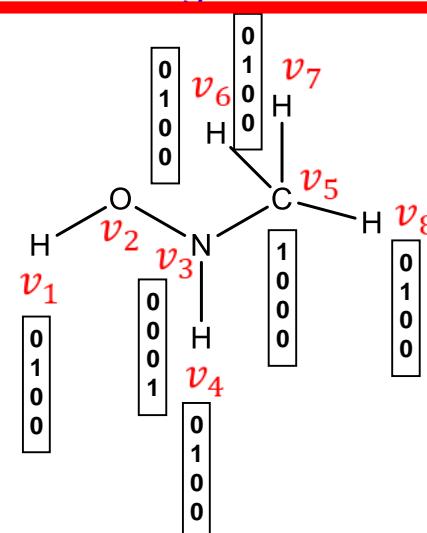


Fig.6

578 alkaloids with known starters, but 11,000 structure-known alkaloids with unknown starters.

## Prediction of starting substances of alkaloids based on Graph Convolutional Neural Networks (GCNN)

(a)



BMC Bioinf., (2019) in press

Classification of alkaloids according to the starting substances of their biosynthetic pathways using graph convolutional neural networks

Ryohei Eguchi<sup>1,2</sup>, Naoaki Ono<sup>1,2\*</sup>, Aki Hirai (Morita)<sup>1</sup>, Tetsuo Katsuragi<sup>3</sup>, Satoshi Nakamura<sup>1,2</sup>, Ming Huang<sup>1</sup>, Md. Altaf-Ul-Amin<sup>4</sup> and Shigehiko Kanaya<sup>1,2</sup>

(b)

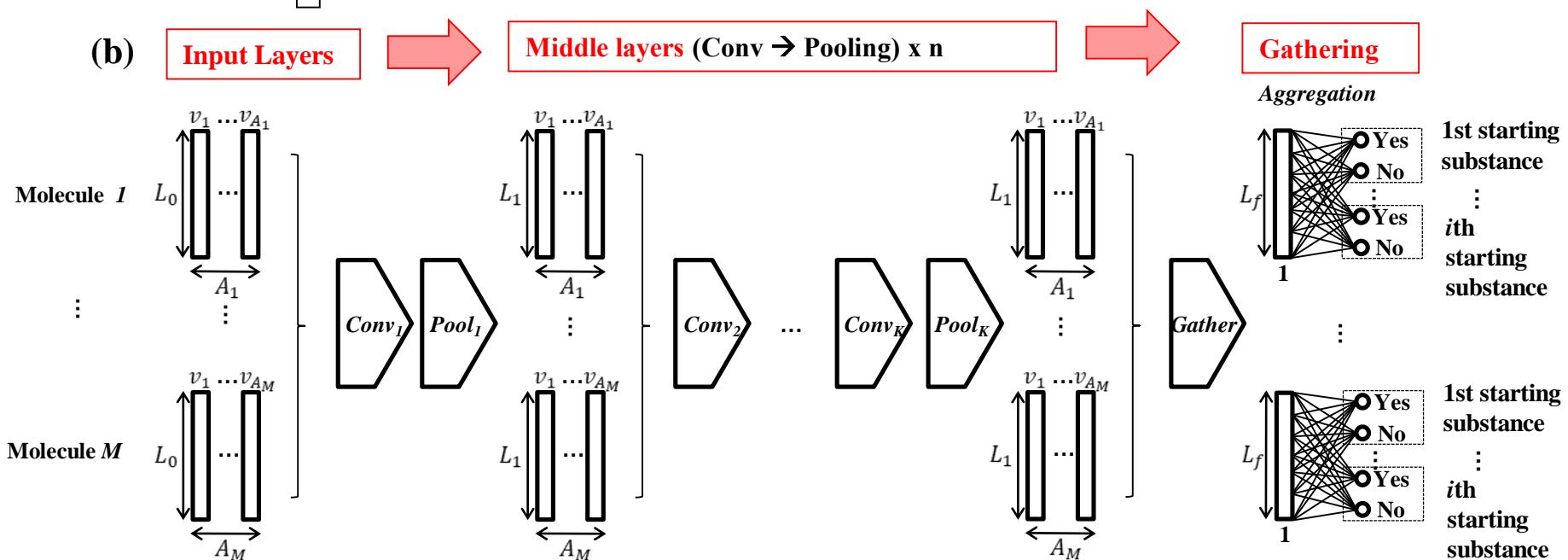
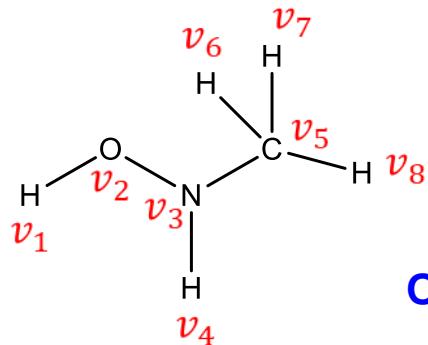
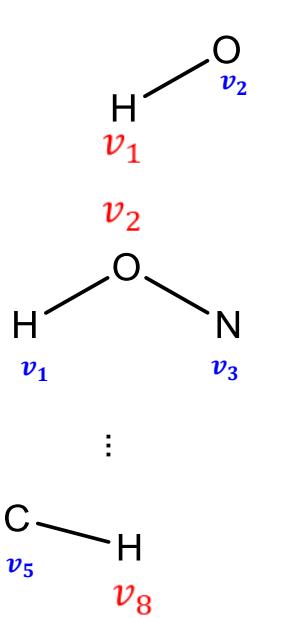


Figure 1 schematic diagram of MGNN. (a) Explanation of one hot vectors for a molecule, (b) Schematic diagram of MGNN (details are explained in the text). In the one hot vectors for  $i$ th molecule visualized in (a),  $A_i$  in (b) is set as 8.

# Graph Convolutional Neural Networks



## Optimization of Weight W



$$\begin{array}{c} M_{c+1} \\ \downarrow \\ M_c \end{array} \xrightarrow{W_c} \begin{array}{c} v_1 v_2 \\ = \\ v_1 v_2 \end{array} \Rightarrow \begin{array}{c} v_1 v_2 \\ = \\ v_1 v_2 \end{array}$$
  

$$\begin{array}{c} M_{c+1} \\ \downarrow \\ M_c \end{array} \xrightarrow{W_c} \begin{array}{c} v_1 v_2 v_3 \\ = \\ v_1 v_2 v_3 \end{array} \Rightarrow \begin{array}{c} v_1 v_2 v_3 \\ = \\ v_1 v_2 v_3 \end{array}$$
  

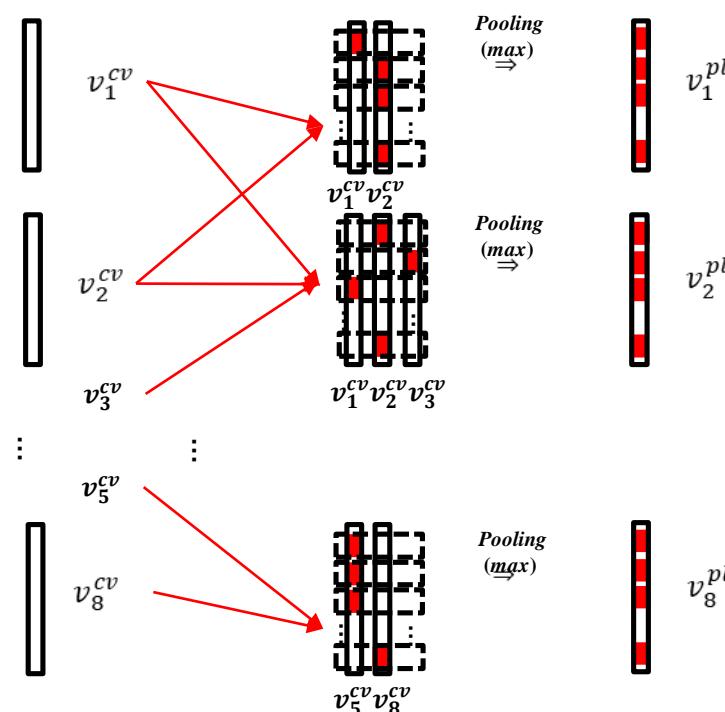
$$\vdots$$
  

$$\begin{array}{c} M_{c+1} \\ \downarrow \\ M_c \end{array} \xrightarrow{W_c} \begin{array}{c} v_5 v_8 \\ = \\ v_5 v_8 \end{array} \Rightarrow \begin{array}{c} v_5 v_8 \\ = \\ v_5 v_8 \end{array}$$

## Optimization of matrix $W_c$

### (a) Convolution

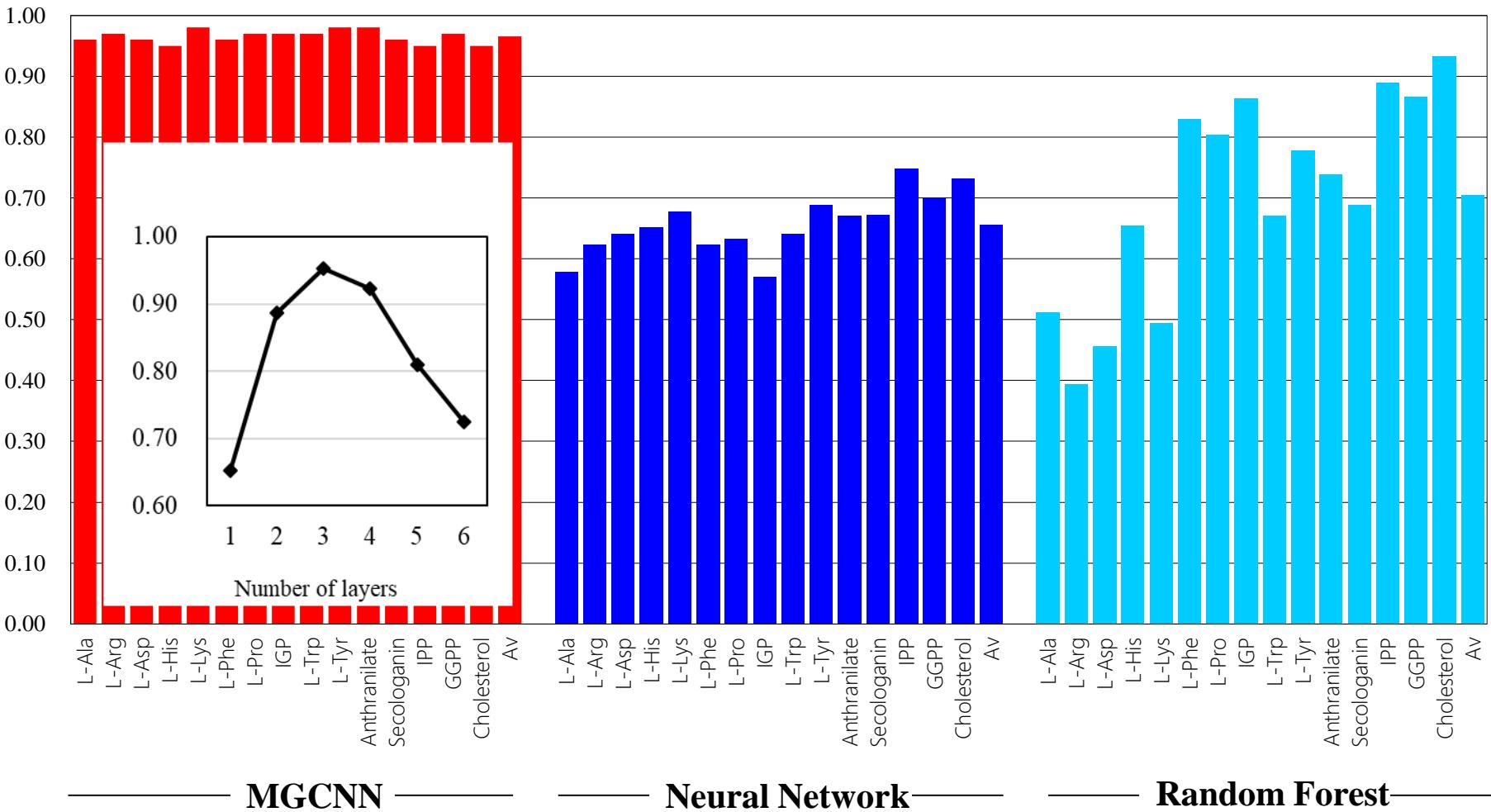
## Selection of targeted and neighboring



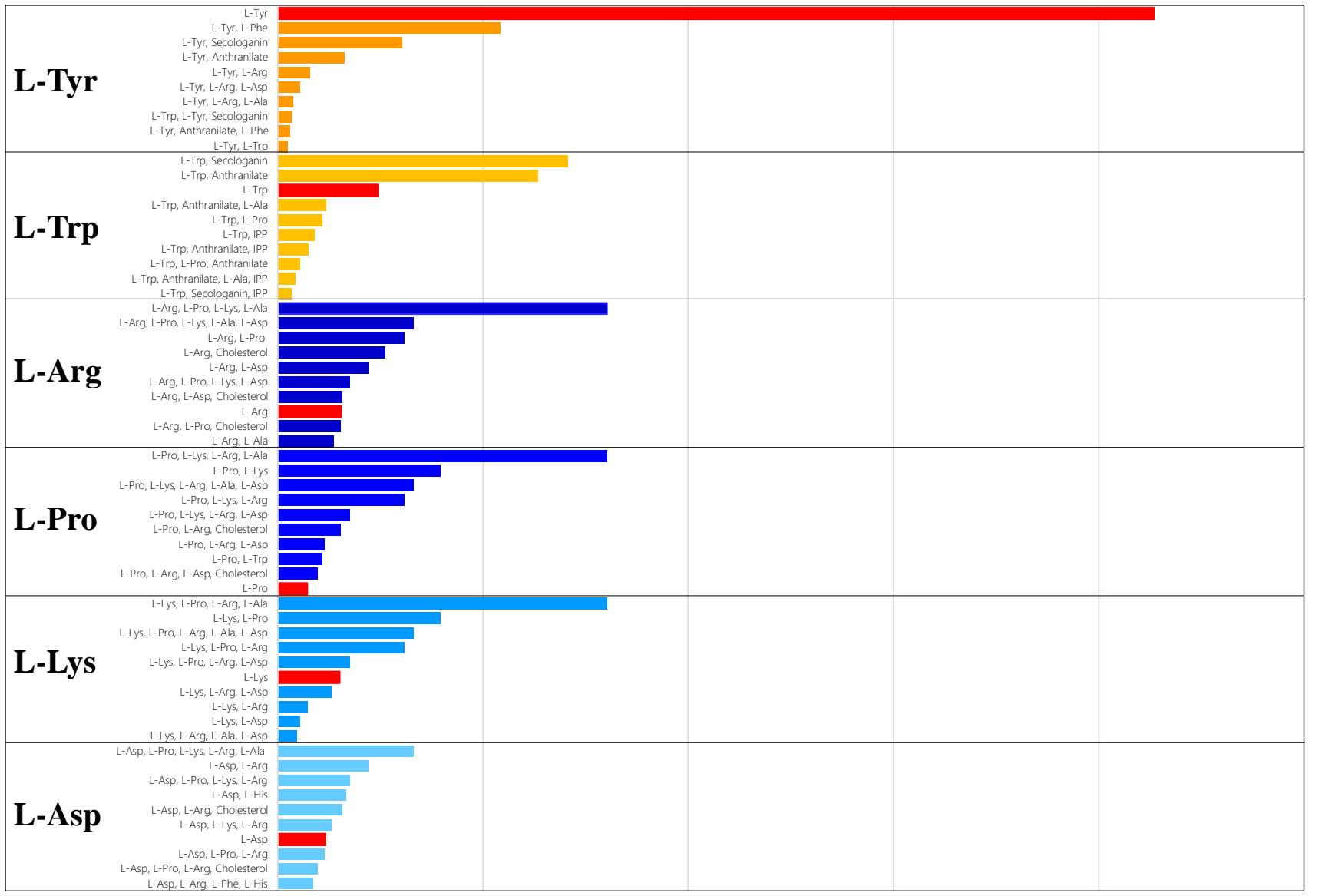
### (b) Pooling

Figure2 (a) Convolution and (b) pooling layers

# Accuracy for GCNN, Neural Network, and Random Forest



# The best 10 combination for six major starting substances in KNApSACk alkaloids



0

500

1000

1500

2000

2500

The number of alkaloids

# Biopathways involved in L-Tyrosine

Group of SS	3-phospho glycerate	Pyruvate	Phosphoenol pyruvate	Oxaloacetate	alpha-Ketoglutarate	Terpenes	TCA cycle	Fatty acid	Nucleic acids
Starting substance (SS)									
Arg, Pro, Asp : Pyrrolidine/Pyrrolizidine Alkaloids	1 ●			●	● ● ● ●				
Trp: Ergot Alkaloids	2	●	● ●	●	● ●	●	●	●	●
Trp: Monoterpene Indole Alkaloids	3		●					●	●
Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate: Trp (Simple Indole Alkaloids), Trp + His	4		●						
Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate: Trp (Simple Indole Alkaloids), Trp + His	5		●					●	
Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate: Trp (Simple Indole Alkaloids), Trp + His	6		●						
Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate: Trp (Simple Indole Alkaloids), Trp + His	7		●						
Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate: Trp (Simple Indole Alkaloids), Trp + His	8		●						
Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate: Trp (Simple Indole Alkaloids), Trp + His	9	● ● ●	● ● ●						
Trp + DMAPP, Trp + Pro; Trp + Oxaloacetate: Trp (Simple Indole Alkaloids), Trp + His	10		●			●			
Tyr: Betanin group: Tyr + Phe, Colchicine	11			● ●					
Tyr: Isoquinoline Alkaloids (Benzylisoquinoline Alkaloids)	12			● ●					
Tyr + Secologanin (Tetrahydroisoquinoline monoterpene alkaloids), Tyr + Ser	13			● ●					
Tyr + Secologanin (Tetrahydroisoquinoline monoterpene alkaloids), Tyr + Ser	14	●		● ●					
Lys: Quinolizidine, Piperidine Alkaloids, Indolizidine Alkaloids Lycopodium Alkaloids	15				● ● ●				
Phe + Ser	16	●		●					
Ala + Acetyl-CoA + Malonyl-CoA	17		●					●	
His: Imidazole alkaloid(histidine derivatives)	18		●				●		
Anthraniate(Acridine Alkaloids)	19				● ●				
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	20								
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	21								
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	22		● ●						
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	23		● ●						
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	24		● ●						
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	25	●							
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	26	●							
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	27								
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	28								
Anthraniate + Trp: Anthraniate + Trp + Ala: Anthraniate + Phe (PKS)	29								
Indole-3-glycerol phosphate (Indole-diterpene alkaloids)	30			●					
Steroidal alkaloids	31								
Purine alkaloids	32							●	

# Predicted Starting substances are reflected in KNApSAcK Core



input word = C00026421

Metabolite Information				Structural formula																				
Name	Dutadrupine																							
Formula	C17H15NO3																							
Mw	281.10519335																							
CAS RN	80151-78-6																							
C_ID	C00026421  , 			 <a href="#">zoom in</a>																				
InChIKey	BNUBFEHGKQEPIQ-UHFFFAOYSA-N																							
InChICode	InChI=1S/C17H15NO3/c1-17(2)8-6-10-13(21-17)5-4-11-14(10)18-16-12(7-9-20-16)15(11)19-3/h4-9H,1-3H3																							
SMILES	c12c(ccc3c1nc1c(c3OC)cco1)OC(C=C2)(C)C																							
Start Substs in Alk. Biosynthesis (Prediction)	<a href="#">Anthranilate</a>			<b>Anthranilate</b>																				
Organism	<table border="1"> <thead> <tr> <th>Kingdom</th><th>Family</th><th>Species</th><th>Reference</th></tr> </thead> <tbody> <tr> <td>Plantae</td><td>Rutaceae</td><td>Dutaillyea drupacea</td><td><a href="#">Ref.</a></td></tr> <tr> <td>Plantae</td><td>Rutaceae</td><td>Melicope erromangensis</td><td><a href="#">Ref.</a></td></tr> <tr> <td>Plantae</td><td>Rutaceae</td><td>Melicope semecarpifolia </td><td><a href="#">Ref.</a></td></tr> <tr> <td>-</td><td>-</td><td>Almeidia coerulea</td><td><a href="#">Ref.</a></td></tr> </tbody> </table>	Kingdom	Family	Species	Reference	Plantae	Rutaceae	Dutaillyea drupacea	<a href="#">Ref.</a>	Plantae	Rutaceae	Melicope erromangensis	<a href="#">Ref.</a>	Plantae	Rutaceae	Melicope semecarpifolia	<a href="#">Ref.</a>	-	-	Almeidia coerulea	<a href="#">Ref.</a>			
Kingdom	Family	Species	Reference																					
Plantae	Rutaceae	Dutaillyea drupacea	<a href="#">Ref.</a>																					
Plantae	Rutaceae	Melicope erromangensis	<a href="#">Ref.</a>																					
Plantae	Rutaceae	Melicope semecarpifolia	<a href="#">Ref.</a>																					
-	-	Almeidia coerulea	<a href="#">Ref.</a>																					



## KNApSACK Metabolomics



### Core System

Since 2004.04



### Search Engine

Since 2008.12



### BiClusO

Since 2019.06

#### Pocket Search for Functional Species

##### Food & Health



**YAKUZEN**  
薬膳データベース  
Since 2015.09



**Lunch Box**  
食用データベース  
Since 2008.07



**DietNavi**  
病気予防データベース  
Since 2012.11



**FoodProcessor**  
加工食品データベース  
Since 2012.11



**DietDish**  
食べ合わせデータベース  
Since 2012.11



**MARCHE**  
旬データベース  
Since 2014.04

##### Crude Drug



**WorldMap**  
世界の薬用植物  
データベース  
Since 2009.06



**KAMPO**  
漢方薬、生薬  
データベース  
Since 2008.08



**JAMU**  
IndonesiaHerb  
データベース  
Since 2009.11



**Tea Pot**  
ハーブ  
データベース  
Since 2011.08

##### Biology



**Metabolite  
Ecology**  
Distribution  
Since 2015.02



**Biological  
Activity**  
Natural Activity  
Since 2011.08



**Biological  
Activity**  
Metabolite Activity  
Since 2013.01



**Twins**  
化合物類似度  
データベース  
Since 2018.02

##### Picnic Gene Annotation



**Arabidopsis**  
Since 2008.04



**Bacillus**  
Since 2008.05



**Human**  
Since 2009.03

##### Strap Correlation Coefficient



**Arabidopsis**  
Since 2009.08



**Bacillus**

##### Pickaxe Metalloprotein Database



**MetalMine**  
Since 2009.08

##### 代謝データベース



**Motorcycle**  
Metabolic Pathway  
Since 2011.08



**Bicycle**  
Algae Metabolic Pathway  
Since 2013.09



**CobWeb**  
Alkaloid Pathway  
Since 2018.07

##### Skewered KNApSACK 串刺し検索

Since 2010.10

3185 days left!



Instruction Manual(Japanese)

Instruction Manual(English)

Terms of service(Japanese)

# MetaboBankの開発 かずさDNA研のデータ整備

2019年7月12日 統合DBプロジェクト 2019年度サイトビット（遺伝研）



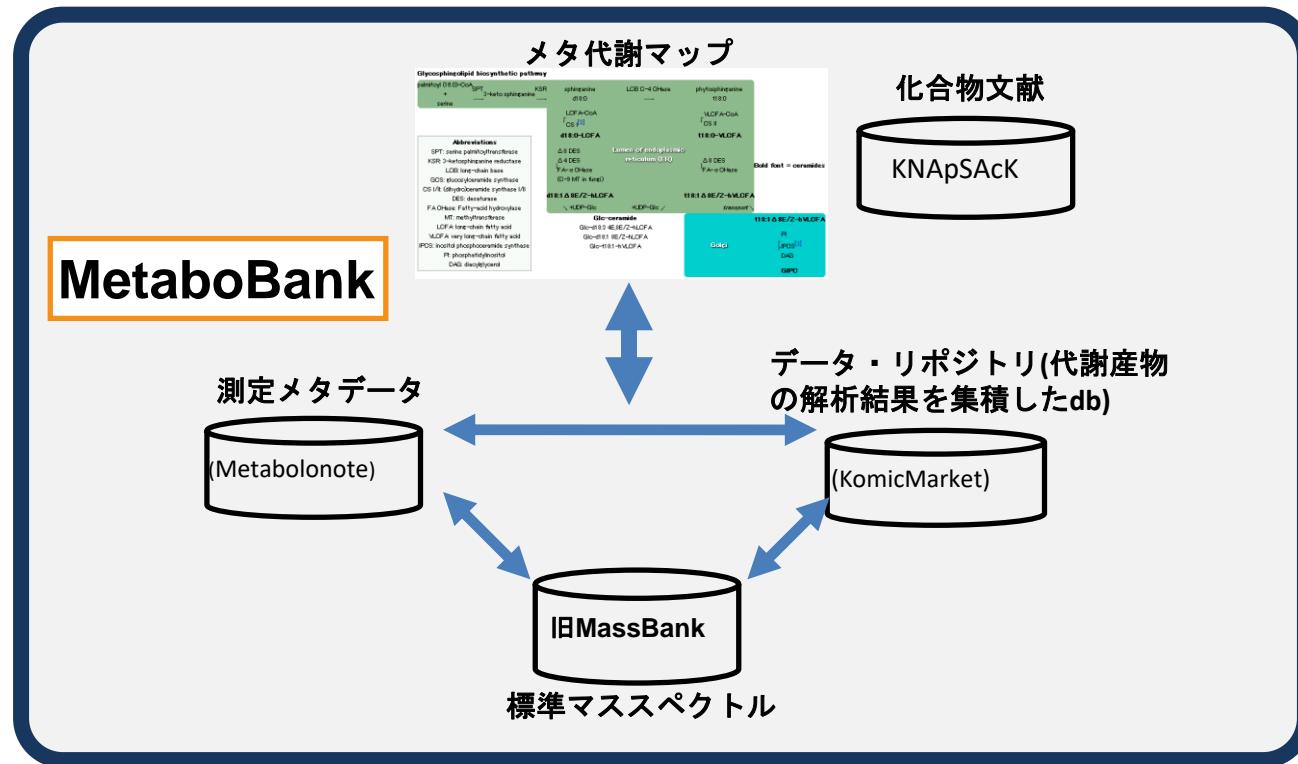
櫻井 望

国立遺伝学研究所

長崎英樹、平川英樹

かずさDNA研究所

# 開発の概要



- 実験データ
- メタデータ

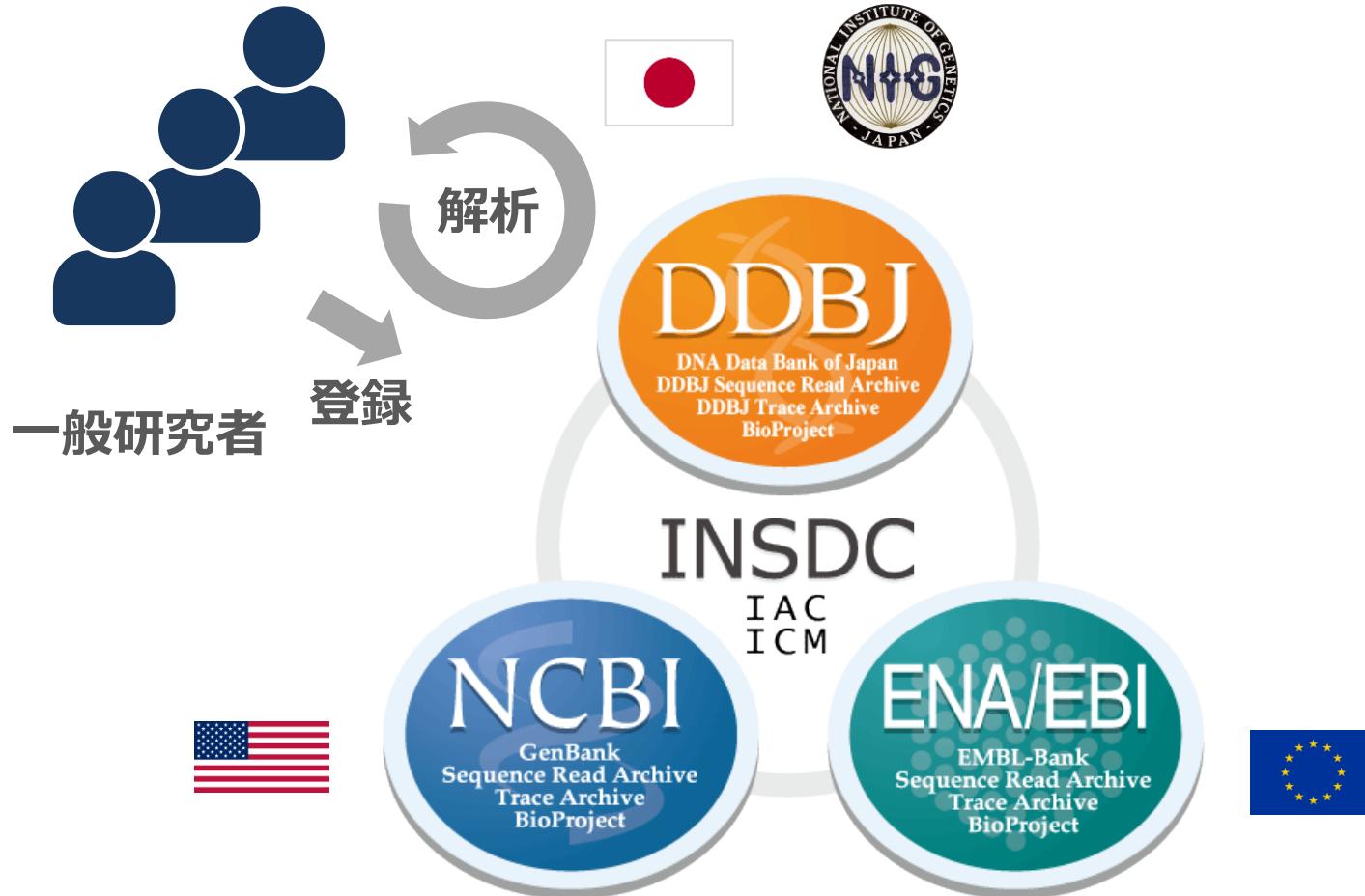


かずさDNA研究所



理化学研究所

# DDBJ : DNA配列登録の拠点の一つ



INSDC: International Sequence Database Collaboration

# メタボロームデータ登録のアジア拠点の構築



登録  
解析



MetaboBank

MetabolomeXchange



かずさDNA研究所

METABOLO  
NOTE



KOMIC  
Market

理化学研究所

PRIME

MassBank





1651 datasets available



# MetabolomeXchange

An international data aggregation and notification service for metabolomics.



food

137 datasets found matching food



## Untargeted metabolome analysis of foods using LC-MS (16) / LC-MSによる食品のノンターゲットメタボローム解析(16)

Untargeted metabolome analyses of foods were performed using LC-MS. Food items were selected from the Standard Tables of Food Composition in Japan-2005 (Seventh Revised Version) published by the Ministry of Education, Culture, Sports, Science and Technology, Japan (hereafter, referred to as the standard table). Metabolites in foods were extracted with methanol, separated by reversed-phase LC and detected using a high-resolution mass spectrometry (LTQ-FT, ThermoFisher Scientific). Two different conditions (Method 1, and Method 5) were applied to ESI positive mode detections, and Method 1 was applied to ESI negative mode detection. A series of different parameter settings for PowerGetBatch software were used for peak detection, and valid peaks were selected by an alignment of the resulted peaks with those detected from several mock samples. The detailed peak information including the results from compound database search and prediction of flavonoid aglycones using FlavonoidSearch were published from the Food Metabolome Repository (<http://metabolites.in/foods>). The IDs mean as follows: Sample ID: Prefix "S" + Food Item No. in the standard table + blanch No. & lowast; The basic blanch No. is 1, and larger numbers were assigned for different types of the food items, and so on. M01: Method 1 for accurate mass measurement in ESI positive mode M11: Method 1 for accurate mass measurement in ESI negative mode M05: Method 5 for acquire MS3 spectrum in ESI positive mode M90, M91: Pseudo analysis representing a set of analyses used in the alignment for each positive and negative analysis, respectively [Japanese] LC-MSを用いて食品をノンターゲットメタボローム分析した。食品は、日本食品標準成分表2015年版(七訂)（文部科学省）（以後「食品標準成分表」）から代表的なものを選んだ。メタノール抽出後、相HPLCで分離し、高分解能フーリエ変換型質量分析（LTQ-FT, ThermoFisher Scientific）による分析を行った。



SE172 View

Log in / create account

## SE172:/

Browse metadata SE172

lower page S070121

### Sample Set Information

ID SE172

Title Untargeted metabolome analysis of foods using LC-MS (16) / LC-MSによる食品のノンターゲットメタボローム解析(16)

Description Untargeted metabolome analyses of foods were performed using LC-MS. Food items were selected from the Standard Tables of Food Composition in Japan-2005 (Seventh Revised Version) published by the Ministry of Education, Culture, Sports, Science and Technology, Japan (hereafter, referred to as the standard table). Metabolites in foods were extracted with methanol, separated by reversed-phase LC and detected using a high-resolution mass spectrometry (LTQ-FT, ThermoFisher Scientific). Two different conditions (Method 1, and Method 5) were applied to ESI positive mode detections, and Method 1 was applied to ESI negative mode detection. A series of different parameter settings for PowerGetBatch software were used for peak detection, and valid peaks were selected by an alignment of the resulted peaks with those detected from several mock samples. The detailed peak information including the results from compound database search and prediction of flavonoid aglycones using FlavonoidSearch were published from the Food Metabolome Repository (<http://metabolites.in/foods>).

The IDs mean as follows:

Sample ID: Prefix "S" + Food Item No. in the standard table + blanch No.

\* The basic blanch No. is 1, and larger numbers were assigned for different types of the food items, and so on.

M01: Method 1 for accurate mass measurement in ESI positive mode

M11: Method 1 for accurate mass measurement in ESI negative mode

M05: Method 5 for acquire MS3 spectrum in ESI positive mode

M90, M91: Pseudo analysis representing a set of analyses used in the alignment for each positive and negative analysis, respectively

[Japanese]

LC-MSを用いて食品をノンターゲットメタボローム分析した。食品は、日本食品標準成分表2015年版(七訂)（文部科学省）（以後「食品標準成分表」）から代表的なものを選んだ。メタノール抽出後、相HPLCで分離し、高分解能フーリエ変換型質量分析（LTQ-FT, ThermoFisher Scientific）による分析を行った。

# Metabolonote

H29年度 調査研究予算で連携達成

# 2019年5月14日

## EBIミーティング



A screenshot of the MetaboLights website. The header includes the EMBL-EBI logo, navigation links for About us, Training, Research, Services, and a search bar. Below the header is a banner with the MetaboLights logo and a search field. The main content area features a section titled "MetaboLights" with a brief description of the database, a "Quick tour" link, and a "Tweets" sidebar from the @Metabolights Twitter account.

# 2019年6月26日

## EBI, NCBIミーティング@Metabolomics2019

- MetaboBank  
インターフェースの開発
- かずさDNA研究所の  
データ整備

# 進捗： RDF検索（実装テスト）



Home Repository Database APIs Statistics Help▼

Simple Search Advanced Search

シンプルな全文サーチ

Search  AND  OR

## Search Results

Show 10 entries

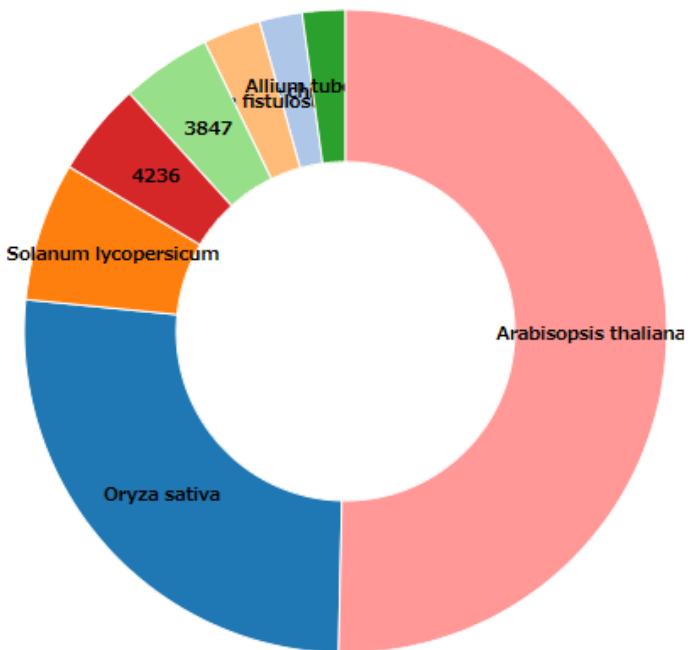
Project ID	Sample	Experiments	#Analysis
RPMM0006	Metabolomic correlation-network modules in Arabidopsis based on a graph-clustering approach	[GC-TOF/MS analysis]	54
RPMM0019	Physiological Roles of the beta-Substituted Alanine Synthase Gene Family in Arabidopsis	[Metabolites analyzed by GC-TOF/MS, Metabolites analyzed by CE-TOF/MS]	36

Showing 1 to 2 of 2 entries

## RDFによる柔軟な絞り込み検索

## Statistics

Statistical information about this website.



# 進捗： データ入力UIの検討 (Mockup)

The screenshot shows a web browser window with the title "New Project" and the URL "localhost:8080/rp/submit/project/new". The page is titled "New/Edit Project" and features a navigation bar with tabs: Project (selected), Experiments, Samples, Methods, People, and Publications.

The main content area contains the following fields:

- ID**: A field containing a dash (-).
- Title**: An input field.
- Description**: A text area.
- Publications**: A list box containing "Sakurai et al., (2018) Bioinformatics \*\*" and "Ogata et al., (2018) PLOS ONE \*\*". An "add" button is located below the list.
- Contact Person**: A list box containing "Sakurai N" with a remove button (X).
- Editable User**: A list box containing "Tokimatsu T" and "Arita M" with remove buttons (X). An "add" button is located below the list.
- Readable User**: A list box containing "reviewer\_pnas" with a remove button (X). An "add" button is located below the list.
- Status**: Radio buttons for "Unpublic" and "Public".
- AttachmentFiles**: A file input field.

A green "My page" button is located in the top right corner of the page area.

# データ登録の課題

## ● Metabolomics Standards Initiative (MSI) 準拠

2007~

The screenshot shows the MSI homepage with a yellow header containing the logo and the text "The Metabolomics Standards Initiative (MSI)". Below the header, there's a message about redirection to a new site. The main content area has sections for "MSI WGs", "DOCUMENTATION", "MEETINGS", and "LINKS". A sidebar on the right contains a list of working groups and their descriptions, followed by a paragraph about the workflow model and a numbered list of standards being developed.

**MSI WGs**

- Biological context metadata WG
- Chemical analysis WG
- Data processing WG
- Ontology WG
- Exchange format WG

**DOCUMENTATION**

- Minutes
- Presentations

**MEETINGS**

- MSI@PSI

**LINKS**

- Metabolomics Society
- Conferences

The Metabolomics Standards Initiative (MSI)

From November 2015: You will be redirected to the new MSI site <http://metabolomics-msi.org/>

The [Metabolomics Society](#) has appointed an Oversight Committee to monitor, coordinate and review the efforts of working groups (WG) in specialist areas that will examine standardization and make recommendations. The overall chair of this committee is [Oliver Fiehn](#). The five MSI WGs, some of which are divided into further subgroups, are listed here:

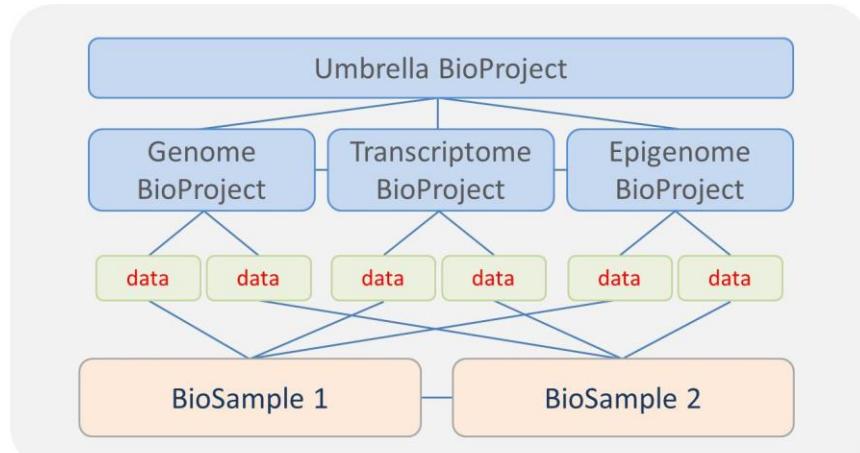
- Biological context metadata WG
- Chemical analysis WG
- Data processing WG
- Ontology WG
- Exchange format WG

The structure of the WGs thus follows the general "workflow" model in metabolomics: from a description of the study design to sample workup, data acquisition, processing and export, bound together by controlled vocabularies and relationships between the terms used.

The MSI WGs are working towards developing the following standards:

1. **Core information for Metabolomics Reporting (CIMR):** This document will specify the minimal guidelines reporting metabolomics work. It will do so in a textual form and will seek in the long term to cover all application areas and analysis technologies. This document will be developed by the [biological context metadata WG](#), the [chemical analysis WG](#), the [data processing WG](#), the [exchange format WG](#) and the [ontology WG](#).
2. **Ontology:** The CV and ontology will be developed iteratively by the [ontology WG](#) on the basis of CIMR. This will be done in collaboration with: [HUPO-PSI](#), [MGED](#) and [FuGO](#).
3. **Exchange format:** A data model and exchange format will be developed by the [exchange format WG](#) on the basis of CIMR, making use of the MSI ontology. This will be done in a collaborative and iterative approach, including [HUPO-PSI](#) and [FUGE](#).

## ● BioSample/BioProject



## ● RDF化

# Comparison of the repositories (1)

	MetaboLights	Metabolomics Workbench	MetaboBank (our plan)
BioSample support	no	no	yes
MSI's reporting guideline	recommended	recommended	compliant
Metadata Input Interfaces	File-based (ISA creator), Web GUI (beta)	Web GUI, File-based (flat-file)	File-based (excel), Web GUI
File transfer	FTP client, Aspera client (project-based management)	FTP client, Aspera client	SFTP, Aspera, HTTPS, (Under investigation)
Group editing	no	no	yes
Private access for peer review	yes (on another server)	(probably no)	yes
Embargo	yes (up to 60 month)	yes (no limitation)	yes
Local installation	yes	(probably no)	yes (for duplication of the system)

# MSIに準拠したBioSampleパッケージの作成

The screenshot shows a Google Sheets spreadsheet with the following structure:

- Row 1:** Headers for columns A through AP.
- Row 2:** Headers for columns A through AP, including "MSI's guideline" and "Modification".
- Row 3:** Headers for columns A through AP, including "BioSample" and "MSI".
- Row 4:** Headers for columns A through AP, including "MSI's guideline" and "Modification".
- Rows 5-38:** A large section labeled "General" containing numerous entries such as "Project Name", "Sample Name", "Sample Title", "Description", "Geographical loc.", "Geographical Co.", "Collection Date", "environmental\_Mg", "environmental\_Mg", "environment (i\_Mg)", "Microbial/eukaryotic", "Breed name - chif", "Cultivar name - cult", "Ecotype", "Genotype", "Phenotype", "Identification", "A label for sample", "name and address", "unique identifier", "type of perturbation", "list of chemical com.", "Processing applied", "Method or device", "Amount or size of", "total count of", "oxy\_stan\_samp", "temperature of", "duration for which", "sample\_store\_dur", "location at which", "temperature at whi", "any other measure", each with a status column (M, O, D, S, etc.) and a "Modification" column.
- Rows 39-54:** A section for "Sample/Subject" with rows for "Species/Strain", "Common Name", "Taxonomic information", "Source, Supplier, Provider", "Generation of mixed strain", "Gender", "Age", "Weight", "Medical History", and "BMI". Each row has a status column (M, O, D, S, etc.) and a "Modification" column.

**BioSample :**  
必須項目は少ないが、  
400近い項目が定義され、ゲノム情報等に使われている

**MSI:**  
細かい必須項目が存在

# Comparison of the repositories (1)

	MetaboLights	Metabolomics Workbench	MetaboBank (our plan)
BioSample support	no	no	yes
MSI's reporting guideline	recommended	recommended	compliant
Metadata Input Interfaces	File-based (ISA creator), Web GUI (beta)	Web GUI, File-based (flat-file)	File-based (excel), Web GUI
File transfer	FTP client, Aspera client (project-based management)	FTP client, Aspera client	SFTP, Aspera, HTTPS, (Under investigation)
Group editing	no	no	yes
Private access for peer review	yes (on another server)	(probably no)	yes
Embargo	yes (up to 60 month)	yes (no limitation)	yes
Local installation	yes	(probably no)	yes (for duplication of the system)

# Comparison of the repositories (2)

	MetaboLights	Metabolomics Workbench	MetaboBank (our plan)
API	(probably no)	yes (RESTful)	yes (RESTful)
Search study by	technology (MS/NMR), organism, organism part, status (public/submitted)	subject type, species, project/study title, institution, submitted year, analysis type (MS/NMR), free text	RDF-based. technology, organism, organism part, free text,
Search compound by	organism, organism part, <b>pathways</b> , reactions, technology (MS/NMR), status (public/submitted), compound name	species, compound name, MS ion mode (positive/negative), <b>compound substructure</b> , compound class, m/z (adduct, compound classes), retention time,	organism, organism part, compound name, m/z, retention time, <b>MSn</b>
Data analysis	no	<b>normalization, clustering, univariate- and multivariate analysis, mapping to the pathways...</b>	yes (not planned in detail yet)

# Comparison of the repositories (3)

	MetaboLights	Metabolomics Workbench	MetaboBank (our plan)
Data Structure	Study Assay	Project Study (per Analysis) Sample	Project Experiment Sample Measurement
Definition of the experimental run	Sample name + protocols	Sample x Analysis	Sample x Measurement
multilingual	no	no	yes
ID	MTBLS (study) MTBLC (compound)	PR (project), ST (study), SA (sample), TR (treatment), SP (sample prep), CO (collection), AN (analysis), NM (NMR), MS (MS)	Measurement, BioSample, BioProject

# Metadata input framework

Definition of the metadata attributes

JSON Schema

```
1 {  
2   "title": "Sample",  
3   "type": "object",  
4   "required": [  
5     "sample_name",  
6     "sample_title",  
7     "collection_date",  
8     "organism"  
9   ],  
10  "properties": {  
11    "sample_name": {  
12      "title": "Sample Name",  
13      "type": "string",  
14      "minLength": 1,  
15    },  
16    "sample_title": {  
17      "title": "Sample Title",  
18      "type": "string",  
19      "minLength": 1,  
20    },  
21    "collection_date": {  
22      "title": "Collection Date",  
23      "type": "string",  
24      "format": "date",  
25    },  
26    "organism": {  
27      "title": "Organism",  
28      "type": "string",  
29    }  
30  }  
31}
```

Definition of required attributes

JSON

```
1 {  
2   "condition": {  
3     "property_name": "sample_type",  
4     "property_value": "Plant",  
5     "condition": "equals"  
6   },  
7   "required": [  
8     "temperature",  
9     "sample_name",  
10    "taxonomy_id",  
11    "collection_date",  
12    "organism"  
13  ],  
14  "required-either-of": [  
15    "organism",  
16    "sample_name"  
17  ],  
18  "optional": [  
19    "sample_title",  
20    "collection_date",  
21    "organism"  
22  ]  
23}
```

Automatic generation of data input forms

JavaScript

Sample [Collapse] Edit JSON Object Properties

Sample Name  
yes  
Leaf MeJA rep1 Short name of the sample. It should be unique in the same project.

Sampling Date  
yes  
年 /月/日 The date of sampling

Organism  
yes  
Arabidopsis thaliana (L.) HeyThe most descriptive organism name for this sample (to the species, if relevant).

NCBI Taxonomy ID  
yes  
3702 NCB Taxonomy identifier. If it is not in the database, enter a tentative ID (e.g., 1). The DDBJ staff applies the tentative ID replaced by an assigned TaxID.

temperature  
yes  
24

Sample Type  
yes  
Plant

{  
 "sample\_name": "Leaf MeJA rep1",  
 "collection\_date": "2014-07-10",  
 "organism": "Arabidopsis thaliana (L.) Hey",  
 "taxonomy\_id": "3702",  
 "temperature": 24,  
 "sample\_type": "Plant"  
}



AJV Json Schema Validator

Metadata

JSON

```
1 [ {  
2   "condition": {  
3     "property_name": "sample_type",  
4     "property_value": "Plant",  
5     "condition": "equals"  
6   },  
7   "required": [  
8     "temperature",  
9     "sample_name",  
10    "taxonomy_id",  
11    "collection_date",  
12    "organism"  
13  ],  
14  "required-either-of": [  
15  ],  
16  "optional": [  
17  ]  
18 } ]
```

Definition of ontologies  
JSON or OWL

RDF

XML

NoSQL db



DDBJ  
BioSample/  
BioProject registration  
system

# 今後の予定

## 9月

MetaboBank (データレポジトリ部) プロトタイプ  
(内部テスト可) の完成

## 12月

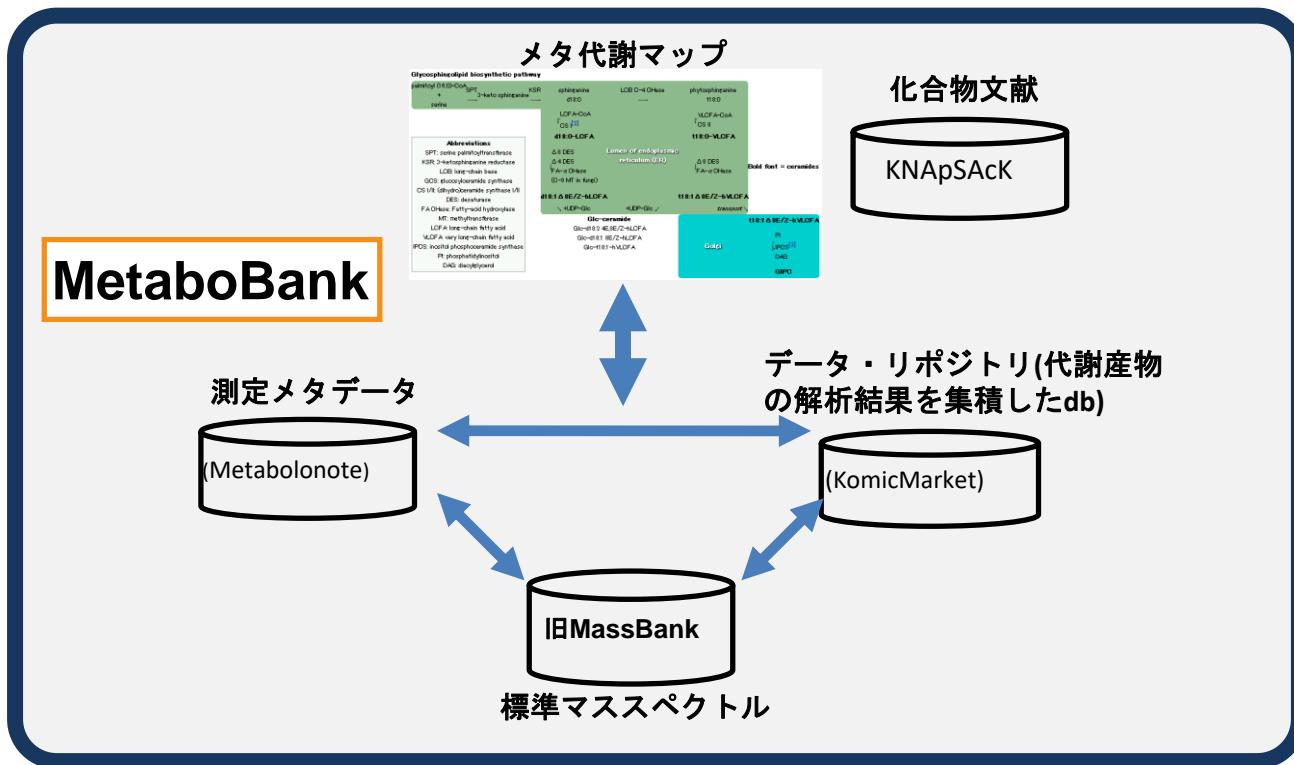
かずさ・理研データの搭載、負荷試験、セキュリティーテスト

## 3月

一般データの受け入れ開始

- MetaboBank  
インターフェースの開発
- かずさDNA研究所の  
データ整備

# かずさ研の役割



- 実験データ
- メタデータ



かずさDNA研究所



# Metabolonoteメタデータ移植作業

The screenshot shows the homepage of the Metabolonote website at [metabolonote.kazusa.or.jp](http://metabolonote.kazusa.or.jp). The page features a red header with the Metabolonote logo and navigation links for Main page, Public Pages, Metadata Search, Community portal, Current events, Help, Active User Ranking, Toolbox, Special pages, Printable version, and Browse properties. The main content area has a red background with white text. It includes a large "Welcome to Metabolonote" heading, a brief introduction, and several sections with icons and descriptions: "メタデータを閲覧する" (View metadata), "API" (Metabolonote's data can be obtained from external programs), "データの登録と編集" (Register and edit data), "システム管理者向け" (For system administrators), and "ヘルプ" (Help). Below this is a "最新情報／お知らせ" (Latest news/information) section with a timestamp of April 24, 2013, and a note about a version 1.2.2 update. At the bottom, there's a diagram showing the flow from "Sample preparation" through "Instrument analysis" and "Data analysis" to "METABOLONOTE", with "META DATA" boxes at each step. A note at the bottom states: "データベースから利用でき、2) データを公開しやすいという利点があります。 詳しくは" (Available from the database, 2) The advantage is that it is easy to publish data. For details, see).

- 植物、微生物等 : 82件
- 食品 : 16件
- その他 : 6件

計104件

- 理研関連 : 82件 (理研で整備)

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

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

# RDFへの変換

## ● RDF (Resource Description Framework)



## ● Metabolonoteメタデータからのデータ構築

### Analytical Method Details Information

ID	MS01
Title	LC-FT-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	<p>Frozen powder of material was extracted with three times volumes of methanol containing 25 <math>\mu</math>M of 7-hydroxy-5-methylflavone as an internal standard. After homogenization using a Mixer Mill MM 300 (Qiagen) at 25 Hz for 2 min twice, homogenates were centrifuged (17,400 g, 5 min, 4 °C). The supernatant was filtered through 0.2 <math>\mu</math>m PTFE membrane (Millipore). Hydrophobic compounds in the filtrate were removed by absorbing to C18 silica column (MonoSpin C18, GL Science, Tokyo, Japan).</p> <p>Mock sample was prepared with the same procedure without adding the plant material.</p> <p>20 <math>\mu</math>l of methanol solution was applied to a TSK-gel column ODS-100V (4.6 x 250 mm, 5 <math>\mu</math>m; TOSOH). Water (HPLC grade; solvent A) and acetonitrile (HPLC grade; solvent B) were used as the mobile phase with 0.1% v/v formic acid added to both solvents. The gradient program was as follows: 3% B (0 min), 97% B (90 min), 97% B (100 min), 3% B (100.1 min), and 3% B (107 min). The flow rate was set to 0.25 ml/min (0-100 min) and 0.5 ml/min (100.1-107 min). The flow rate was set to 0.5 ml/min, and the column oven temperature was set at 40 °C. Compounds are detected in ESI-positive mode in the range m/z 100-1500. Multistage MSn analyses were carried out using collision-induced dissociation (CID) in a linear ion trap detector at a normalized collision energy of 35.0% and an isolation width of 4.0 (m/z), and monitored by both ion trap detector and FT-ICR detector at 25,000 (at m/z 400) mass resolution. The ESI setting was as follows: spray voltage 4.0 kV and capillary temperature 300 °C. Nitrogen sheath gas and auxiliary gas were set at 40 and 15 arbitrary units, respectively. To monitor HPLC elution, a photodiode array detector was used in the wavelength range 200–650 nm.</p> <p>Mass scan events were set as follows (referred as method 1): full mass scan with FT-ICR in resolution 100,000 and MS2 scans for the most intense 5 ions of full mass scan with ion trap (IT). A dynamic exclusion setting was applied as follows: repeat count, 3; repeat duration, 30 sec; exclusion list size, 500; margin, 10 ppm, and exclusion duration, 20.</p> <p>The data were acquired by Xcalibur software version 2.07 (Thermo Fisher Scientific).</p>
サンプル調整	
HPLC	
マス	

共通の項目(クラス)でリンクされた50シートのエクセルファイルからRDF化合わせて正誤確認(記述ミス、著者所属確認等)

Chromatography クロマトグラフィ	comment コメント	description 説明	temperature gradient 温度勾配	column type カラムの種類	column temperature カラム温度	column pressure カラム圧力	column name カラム名
pm_chromato:SE112_MS01	"Water (HPLC grade; solvent A) and acetonitrile (HPLC grade; solvent B) were used as the mobile phase with 0.1% [temp:40C]						"TSK-gel column ODS-100V (4.
pm_chromato:SE112_MS05	"Water (HPLC grade; solvent A) and acetonitrile (HPLC grade; solvent B) were used as the mobile phase with 0.1% [temp:40C]						"TSK-gel column ODS-100V (4.
pm_chromato:SE112_MS11	"Water (HPLC grade; solvent A) and acetonitrile (HPLC grade; solvent B) were used as the mobile phase with 0.1% [temp:40C]						"TSK-gel column ODS-100V (4.

# MetabolonoteのXMLファイルからの変換処理

```
<Page ID="" Title="/S01/M01">
<Template Name="M"><Field Name="M_ID">M01</Field><Field Name="M_Title">LC-FTICR-MS, ESI Positive analysis</Field><Field Name="M_Method Set ID">MS1</Field><Field Name="M_Sample Amount">6.7 mg</Field><Field Name="M_Comment">[MassBase ID] MDLC1_05711</Field></Template>
<Free_Text id="1">{{LinkTo_MassBase|MDLC1_05711}}</Free_Text>
</Page>
<Page ID="" Title="/S01/M01/D01">
<Template Name="D"><Field Name="D_ID">D01</Field><Field Name="D_Title">PowerGet data analysis for Bio-MassBank</Field><Field Name="D_Data Analysis Set ID">DS1</Field><Field Name="D_Recommended decimal places of m/z">6|ITMS 2</Field></Template>
<Free_Text id="1">{{LinkTo_BioMassBank|keyword=SE1_S01_M01_D01}}
&lt;!-- {{LinkTo_KomicMarketTmp}} --&gt;
{{LinkTo_KM2|SE1}}</Free_Text>
</Page>
<Page ID="" Title="/S01/M01/D02">
<Template Name="D"><Field Name="D_ID">D02</Field><Field Name="D_Title">PowerGet data analysis for KomicMarket2</Field><Field Name="D_Data Analysis Set ID">DS2</Field><Field Name="D_Recommended decimal places of m/z">6|ITMS 2</Field></Template>
<Free_Text id="1">&lt;!-- {{LinkTo_KomicMarketTmp}} --&gt;
{{LinkTo_KM2|SE1}}</Free_Text>
```



```
#S01/M01
M_ID: M01
M_Title: LC-FTICR-MS, ESI Positive analysis
M_Method Set ID: MS1
M_Sample Amount: 6.7 mg
M_Comment: [MassBase ID] MDLC1_05711
```

PGDBjやMassBaseなど外部へのリンクもまとめて出力

```
#S01/M01/D01
D_ID: D01
D_Title: PowerGet data analysis for Bio-MassBank
D_Data Analysis Set ID: DS1
D_Recommended decimal places of m/z: 6|ITMS 2
```

```
#S01/M01/D02
D_ID: D02
D_Title: PowerGet data analysis for KomicMarket2
D_Data Analysis Set ID: DS2
D_Recommended decimal places of m/z: 6|ITMS 2
```

でも基本データ手入力

# 進捗：

- 104件完了
- RIKENメタデータデータベースに登録

→ 検索テスト・入力フォーマット(6/3版)の調整

The screenshot shows the 'plantMetabolomics' database interface. At the top, there's a navigation bar with a 'Database' icon, the title 'plantMetabolomics', and links for 'Download', 'SPARQL', 'History', and 'Inquiry'. Below the title, a section titled 'Classes belonging to this database' lists several categories with their sub-components:

- Project:
  - person / organization
    - person
    - Organisation
- samples:
  - Sample
    - Compound
    - Biological Sample
  - Light Condition
    - Treatment
- MS analyze:
  - Sample Preparation
    - Sample Extraction Protocol
    - Sample Amount
  - Chromatography
    - Measurement Condition
    - Mass Spectrometry
- raw data:
  - Measurement
    - Raw Dataset
    - File
- analyzed data:
  - Annotation Method Details
    - Data Analysis
    - Pre-processing
  - Statistical Data Analysis
    - Normalisation
    - Multivariate Analysis
  - Univariate Analysis
    - Data Transformation
    - Scaling

At the bottom, there's a footer with the RIKEN MetaDatabase logo, language links ('en ja'), and copyright information: 'metadbdev.riken.jp/sandbox/db/plantMetabolomics/http://metadb.riken.jp/ontology/plantMetabolomics/0.1/Project' and 'Creative Commons Attribution-ShareAlike 4.0 International License, CC 2010-2018, RIKEN'.

RPMM移植試験で使用したメタデータ

SE1 基本的

SE6 LTQ-Orbitrap使用

SE18 Mouse liver 解析書き込み多い

SE40 サンプル多い

SE61 サンプル、解析種類多い

SE62 サンプル、解析書き込み多い

SE88 Micro-Tom 解析種類多い

SE97 Lotus japonicus 解析種類多い

SE112 食品レポジトリ

SE114 食品レポジトリ

SE116 食品レポジトリ

SE189 解析書き込み多い

# 今後の予定

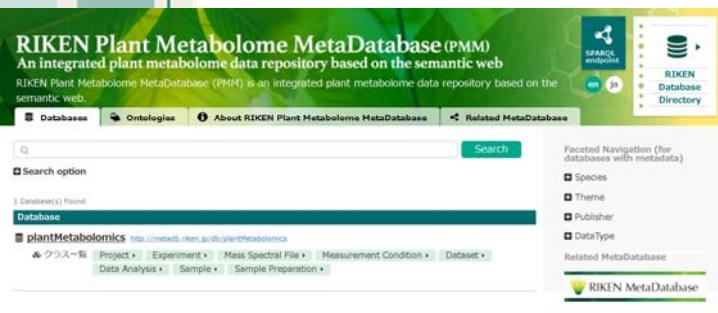
- 実験データの再解析

化合物データベースの充実化によって検索でヒットする化合物  
が増えることが見込まれる。

- 植物ゲノム統合DBへのメタボロームデータの統合

ゲノム関連データとの横断的な検索を可能にする。

# RIKEN Plant Metabolome MetaDatabase: an integrated plant metabolome data repository based on the semantic web



<http://metabobank.riken.jp/>

Atsushi Fukushima  
RIKEN Center for Sustainable Resource Science

2019 07 12  
Togo DB Site Visit



# Aim of this project

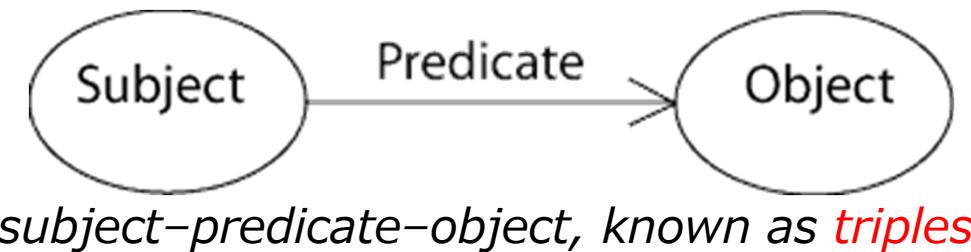
- Development of the RIKEN Plant Metabolome MetaDatabase (RIKEN PMM)
  - Semantic web based
  - Construction of the Plant Metabolomics Ontology
- Reanalysis/reannotation of GC-MS data in plants
  - Versioning of GC-MS spectra
  - Improvement of metabolite annotation

# RDF/SPARQL

A standard model for data interchange on the Web

## RDF (Resource Description Framework)

- has features that facilitate data merging even if the underlying schemas differ
- specifically supports the evolution of schemas over time without requiring all the data consumers to be changed



## SPARQL - SPARQL Protocol and RDF Query Language

- an RDF query language
- can retrieve and manipulate data stored in RDF format

[https://ja.wikipedia.org/wiki/Resource\\_Description\\_Framework](https://ja.wikipedia.org/wiki/Resource_Description_Framework)

## Methods & Implementation

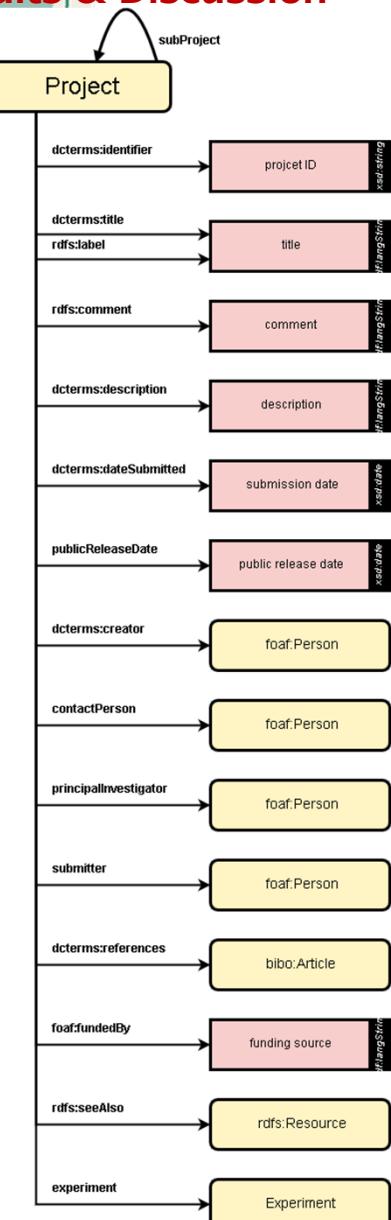
# RIKEN MetaDatabase

The screenshot shows the RIKEN MetaDatabase homepage. At the top, there's a green header with the title "RIKEN MetaDatabase". Below it, a sub-header explains the service: "理研メタデータベースは、理研の成果をより広く研究者の方に活用していただくことを目的に、理研の研究者が公開するデータベースのメタデータを体系的に整理して公開するサービスです。" There are four navigation links: "データベース", "オントロジー語彙", "理研メタデータベースについて", and "関連メタデータベース". On the right side, there are icons for "SPARQL endpoint" (with "en" and "ja" language options) and a "理研 データベース 電話帳" link. A search bar is at the top left, followed by a "検索" button. Below the search bar, there's a "検索オプション" section with a "検索込み (メタデータのあるDBのみ)" checkbox. To the right, there are filters for "生物種", "対象", "センター", and "データ種". The main content area displays search results for "112件のデータベースが見つかりました". The results are paginated from 1 to 6, with "All" selected. Each result entry includes a thumbnail, the database name, and its URL. For example, "理研データベース電話帳" has a URL of <http://metadb.riken.jp/db/DBcatalog>. Other entries include "FANTOM5 SSTAR" ([http://metadb.riken.jp/db/sciNets\\_ria365](http://metadb.riken.jp/db/sciNets_ria365)) and "FANTOM4" ([http://metadb.riken.jp/db/sciNets\\_ria187](http://metadb.riken.jp/db/sciNets_ria187)). Each entry shows a list of categories or classes. At the bottom, there's a footer with the RIKEN logo, a CC-BY-SA license notice, and links for "ホーム", "別ウィンドウで開く", "サイトポリシー", "管理画面", and language selection ("en", "ja").

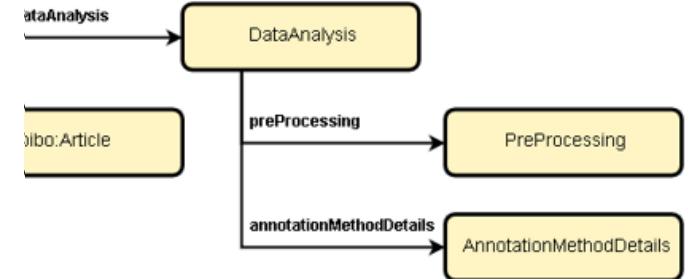
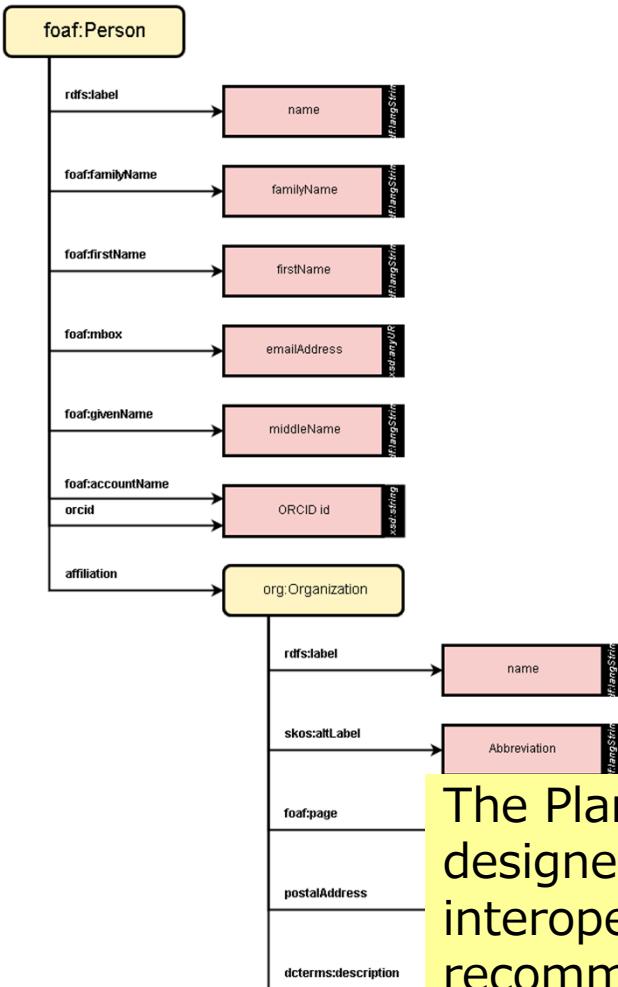
RIKEN MetaDatabase  
<http://metadb.riken.jp/>

Kobayashi, N et al. LNCS 10055, 99-115 (2016)

## Results & Discussion



# RDF Schema in RPMM



The Plant Metabolomics Ontology is designed to realise metadata interoperability according to the recommendation of the Metabolomics Standards Initiative (MSI, <http://www.metabolomics-msi.org/>)

## Results & Discussion

# Metadata registration

Using a spreadsheet (e.g., MS Excel)

A	B	C	D	E	F	G	H	I	J	K
EnglishAttribute	日本語属性	biological sample name	生物サンプル ID	comment	comment	description	taxonomy	cultivar	genotype	plant organ
PropertyURI		dcterms:identifier	rdfs:label	rdfs:comment	rdf:langString	dcterms:descr	分類群	cultivar	genotype	植物器官
Datatype	BiologicalSample	xsd:string	rdf:langString	rdf:langString	rdf:langString	obo:R0_0002162	owl:Class	xsd:string	xsd:string	owl:Class
1	EnglishAttributor	Biological Sample	生物サンプル							
2	日本語属性									
3	PropertyURI									
4	Datatype									
5	pm_biosamp: test699_F_037_1	"F_037_1"	"F_037_1"	"This is test data"@en		http://purl.bioontology.org/ontology/NCBITAXON/"Reijo"				http://purl.obo.org/
6	pm_biosamp: test699_F_037_old_1	"F_037_old_1"	"F_037_old_1"	"This is test data"@en		http://purl.bioontology.org/ontology/NCBITAXON/"Reijo"				http://purl.obo.org/
7	pm_biosamp: test699_F_040_1	"F_040_1"	"F_040_1"	"This is test data"@en		http://purl.bioontology.org/ontology/NCBITAXON/"Reijo"				http://purl.obo.org/
8	pm_biosamp: test699_F_040_old_1	"F_040_old_1"	"F_040_old_1"	"This is test data"@en		http://purl.bioontology.org/ontology/NCBITAXON/"Reijo"				http://purl.obo.org/
9	pm_biosamp: test699_F_043_1	"F_043_1"	"F_043_1"	"This is test data"@en		http://purl.bioontology.org/ontology/NCBITAXON/"Reijo"				http://purl.obo.org/

An example of spreadsheet describing RIKEN PMM's "BiologicalSample" class.

G	H	I	J
description	taxonomy	cultivar	genotype
説明	分類群	品種	遺伝型
dcterms:descr	obo:R0_0002162	cultivar	genotype
rdf:langString	owl:Class	xsd:string	xsd:string
	http://purl.bioontology.org/ontology/NCBITAXON/"Reijo"	http://purl.obo.org/	

Generation of PMM's RDF

The screenshot shows the RIKEN Plant Metabolome MetaDatabase (PMM) interface. At the top, there is a navigation bar with links for SPARQL endpoint, RIKEN Database Directory, and language selection (en, ja). Below the navigation, there is a search bar and a search button. The main content area displays a search result for 'plantMetabolomics'. The result includes a summary table with columns for Database, Project, Experiment, Mass Spectral File, Measurement Condition, Dataset, Data Analysis, Sample, and Sample Preparation. On the right side, there is a sidebar titled 'Faceted Navigation (for databases with metadata)' with options for Species, Publisher, and DataType. At the bottom, there is a footer with the RIKEN logo and copyright information.



## Results & Discussion

# RIKEN PMM's graphical user interface (development ver)

## Summary page

Database

plantMetabolomics

plantMetabolomics

Classes belonging to this database

Project Experiment Mass Spectral File  
Measurement Condition Dataset Data Analysis  
Sample Sample Preparation Other classes

Data statistics

Data statistics	
number of triples	186012
number of classes (concepts)	64
number of properties	138
number of instances associated with classes	29
Project	55
Experiment	1377
Mass Spectral File	44
Measurement Condition	0
Dataset	27
Data Analysis	4992
Sample	88
Sample Preparation	17
Organisation	

Download SPARQL History Inquiry

Plant Metabolome MetaDatabase

Home Open in new window Site Policy

Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 International License. © 2016-2019, RIKEN

## Project page

Instance

plantMetabolomics Project

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

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(Project)

Show URI

Download SPARQL History Inquiry

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

Project

- RPMM0003

project ID

title

- 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

comment

description

- Tomato (*Solanum lycopersicum*) is a model crop for studying development regulation and ripening in flesh fruits and vegetables. Supplementary light to maintain the optimal light environment can lead to the stable growth of tomatoes in greenhouses and areas without sufficient daily light integral. Technological advances in genome-wide molecular phenotyping have dramatically enhanced our understanding of metabolic shifts in the plant metabolism across tomato fruit development. However, comprehensive metabolic and transcriptional behaviors along the developmental process under supplementary light provided by light-emitting diodes (LED) remain to be fully elucidated. We present integrative omic

Home Open in new window Site Policy

Except where otherwise noted, this work is subject to a Creative Commons Attribution 4.0 International License. © 2016-2019, RIKEN

## Results & Discussion

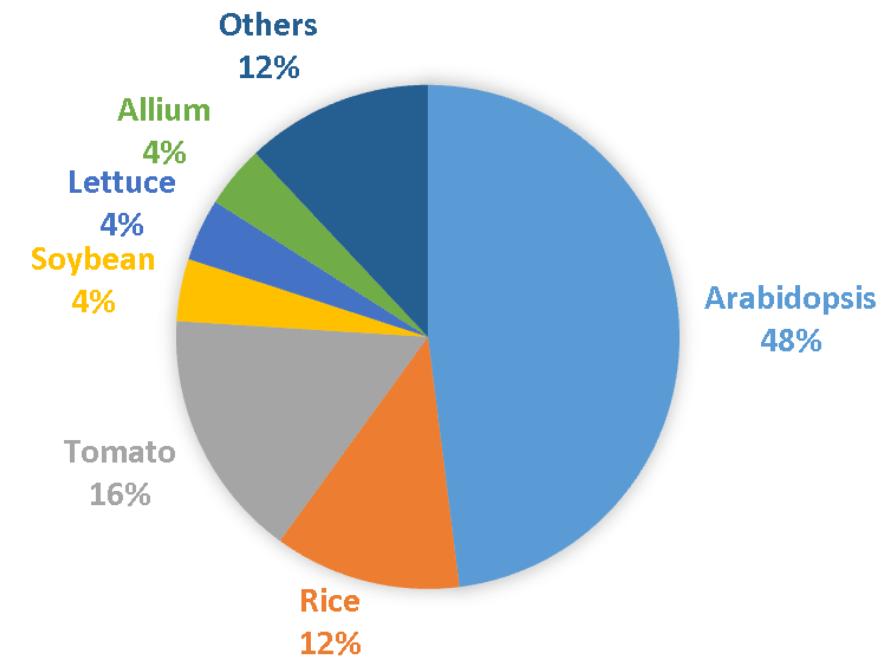
# The current status of RPMM (development ver)

Summary page

A screenshot of the RPMM summary page. At the top, there's a navigation bar with tabs for 'Project' (selected) and 'Instance'. Below the navigation is a search bar and a download button. A callout box highlights the following statistics: '29 projects', '> 4K biological samples', and '> 4K raw data files'. The main content is a table with 29 rows, each representing a project. The columns include 'Project ID', 'title', 'comment', 'description', 'submission date', and 'public release date'. The first few rows show projects like RPMM0001, RPMM0002, and RPMM0003, each with a detailed description of its purpose.

Project	Project ID	title	comment	description	submission date	public release date
• Unbiased characterization of genotype-dependent metabolic regulations by	RPMM0001	• Unbiased characterization of genotype-dependent metabolic regulations by		• Metabolites are not only the catalytic products of enzymatic reactions but also the active regulators or the		
• Metabolomic Characterization of Knockout Mutants in Arabidopsis: Development	RPMM0002	• Metabolomic Characterization of Knockout Mutants in Arabidopsis: Development		• Despite recent intensive research efforts in functional genomics, the functions of only a limited number of		
• A systems analysis with "simplified source-sink model"	RPMM0003	• A systems analysis with "simplified source-sink model"		• Tomato ( <i>Solanum lycopersicum</i> ) is a model crop for studying		

Species stat



Arabidopsis  
48%

# Summary

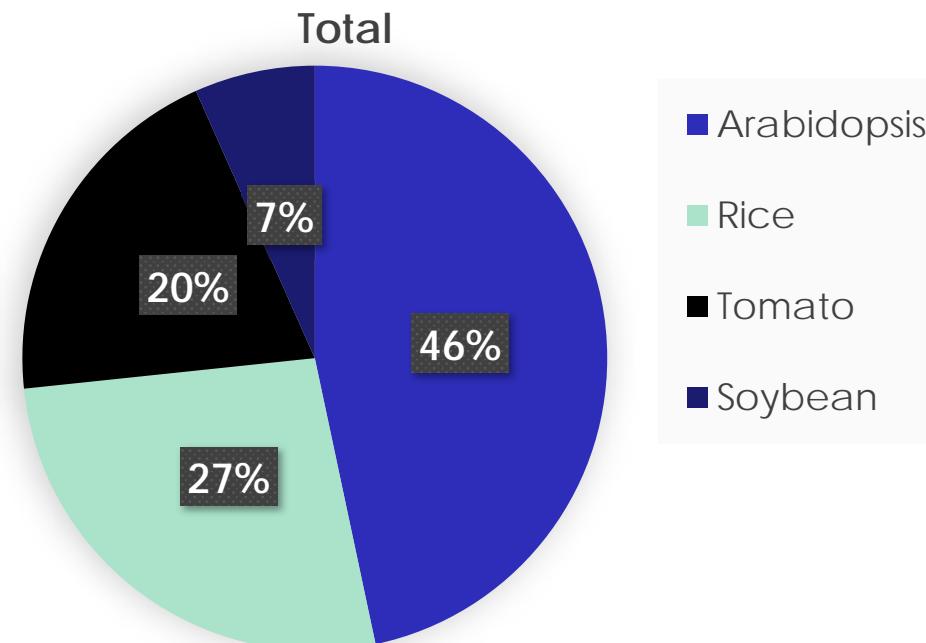
- developed a semantic web-based repository database known as the RIKEN PMM using native RDF technologies
- 29 projects with high-quality metadata
- efficient data sharing, spreading and retrieving in plant metabolomics



<http://metabobank.riken.jp/>

# **Re-analysis of GC/MS data**

# 15 Re-analyzed datasets



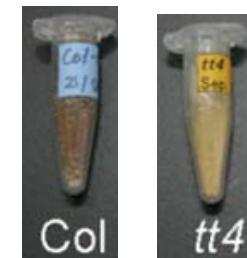
# An example

## Metabolomic correlation-network modules in *Arabidopsis* based on a graph-clustering approach

Fukushima et al. (2011); Kusano et al. (2007)

**RPMM0006, MTBLS45, SE47**

- *Arabidopsis thaliana*
- 3 genotypes (Col-0, *mto1*, and *tt4*)
- Root samples (biological replicates,  $n \geq 15/\text{genotype}$ )  
53 samples
- Related to **RPMM0001** (MTBLS40 and SE46) (aerial part)
- GC-MS



# Re-analysis method

- MS-DIAL (Tsugawa et al.)
- Comparison with original annotation (Kusano et al. )

## Results

	Extracted peaks	Annotated peaks	Known peaks
Original annotation	518	166	84
Re-analyzed annotation	444	253	204
	RI (error) < 10		58

Match rate: 58/84 = **69%**

M000018\_A152001-101-xxx\_NA\_1515.87\_TRUE\_VAR5\_ALK\_Methionine (2TMS) RT

