

## メタボローム・データベースの開発

金谷 重彦・西岡孝明

奈良先端科学技術大学院大学 (NAIST)

情報科学研究科・情報生命科学専攻

・計算システムズ生物学講座

櫻井 望

(財)かずさDNA研究所・

産業基盤開発研究部

有田 正規

(独)理化学研究所

植物科学研究センター

平成24年2月24日

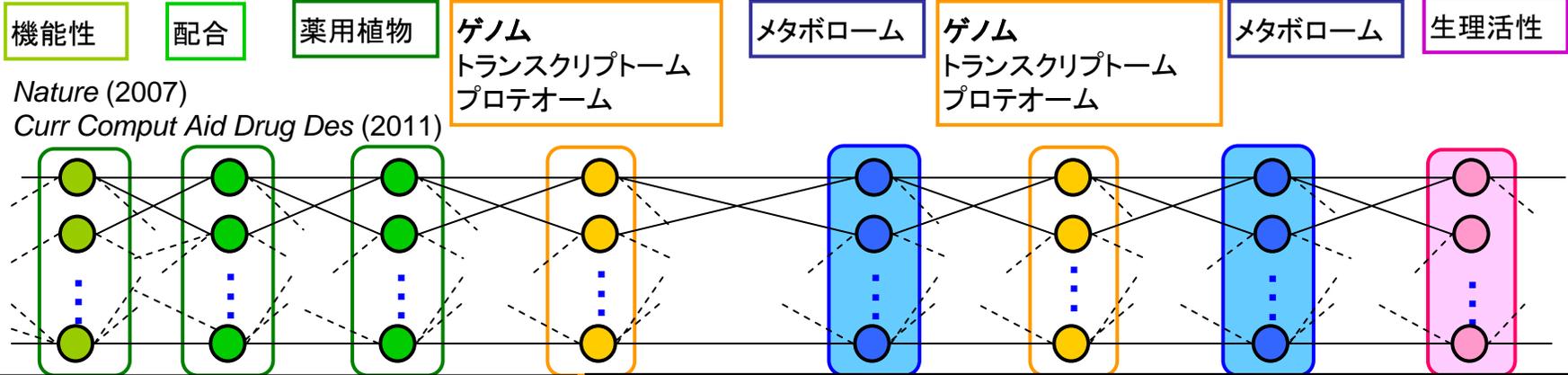
# メタボローム・データベースの意義

## 日本は二次代謝物研究で世界をリード

薬/食用知識

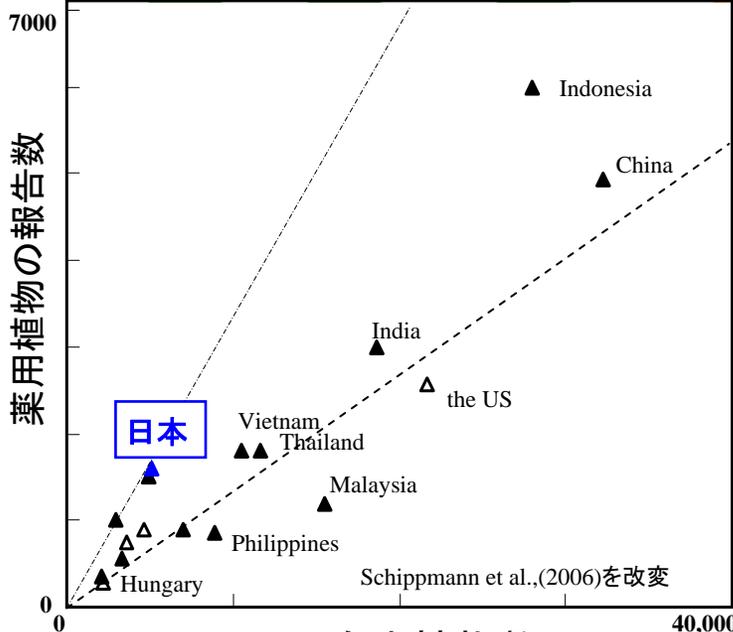
有用生物

ヒト



Nature (2007)

Curr Comput Aid Drug Des (2011)



メタボローム研究の課題

検出できるが、同定率が低い。

(1)代謝物MSデータを集約・共有するDBが必要

日本は微生物・植物・海洋生物の資源大国

2次代謝物は生理活性の宝庫

(2)メタボロームと生理活性のリンクが必要

(3)2次代謝DBとゲノム情報のリンクが必要

# メタボローム・データベース構想

質量スペクトルから生理活性情報の統合化をめざす。

日本メタボローム・データベース  
パブリック・リポジトリ

メタボロミクス研究者、一般ユーザ  
デポジット・閲覧・アノテーション・キュレーション

## [1] 質量スペクトルDB (MS DB)

▪ 学会MSDB  
MassBank, LipidBank  
▪ 研究グループMSDB  
PRiMe

▪ 個別研究における代謝物の同定  
文献情報  
▪ 大規模メタボロームMSDB  
MassBase, MS2T, KomicMarket

[1-1] 化合物MS  
DB

[1-2] メタボローム  
MS DB

## [2] 代謝物情報DB

▪ 文献情報  
KNAPSAcK DB

▪ 質量スペクトルのアノテーション  
MassBankフラグメンテーション・ライブラリ

[2-1] 代謝物-  
{生物種, 生理活  
性}関係DB

[2-2] MS-化合物構  
造の関係知識DB

## [3] メタボローム統合DB

[3-1] ウィキDBによるメタボローム  
データの統合管理

[3-2] メタボローム・アノテーシ  
ョン・システム

[3-3] ゲノム情報とのリンク

基礎研究: 持続可能社会に向けた生物資源の有効利用、生物種、分野、目的に応じた基礎研究  
産業応用: 新規有用代謝物質の探索、ゲノム育種による有用/新規代謝物質の生産 など

# 研究開発項目

	H23	H24	H25
<p>1.質量スペクトルDB(化合物MSDBとメタボロームMSDB)の拡充</p> <p>1.1化合物MS DBの開発</p> <ul style="list-style-type: none"> <li>・プロトタイプシステムの開発</li> <li>・データのDBへの蓄積ならびに公開</li> </ul>			
	公開 MassBank		
<p>1.2メタボロームMS DBの開発</p> <ul style="list-style-type: none"> <li>・DBプロトタイプシステムの開発</li> <li>・データのDBへの蓄積ならびに公開</li> </ul>			
	公開 BioMassBank		
<p>2.代謝物質情報DBの構築</p> <p>2.1 代謝物質と生物活性の関係データベース</p> <ul style="list-style-type: none"> <li>・データベース構築</li> <li>・データ蓄積ならびに公開</li> </ul>			
	公開 Metabolite Activity		
<p>2.2 MS データと化学構造の関係知識DBと化学構造式推定ツールの開発・実装</p> <ul style="list-style-type: none"> <li>・システム開発</li> <li>・データの充実ならびに公開</li> </ul>			
	公開 MassBank		
<p>3.メタボローム統合DB</p> <ul style="list-style-type: none"> <li>・wikiを中心としたデータ統合技術開発</li> <li>・wikiを中心とした統合データの公開</li> </ul>			
	MassBank Wiki		

# メタボローム・データベース構想

質量スペクトルから生理活性情報の統合化をめざす。

日本メタボローム・データベース  
パブリック・リポジトリ

メタボロミクス研究者、一般ユーザ  
デポジット・閲覧・アノテーション・キュレーション

## [1] 質量スペクトルDB (MS DB)

▪ 学会MSDB  
MassBank, LipidBank  
▪ 研究グループMSDB  
PRIME

▪ 個別研究における代謝物の同定  
文献情報  
▪ 大規模メタボロームMSDB  
MassBase, MS2T, KomicMarket

[1-1] 化合物MS  
DB

[1-2] メタボローム  
MS DB

## [2] 代謝物情報DB

▪ 文献情報  
KNAPSAcK DB

▪ 質量スペクトルのアノテーション  
MassBankフラグメンテーション・ライブラリ

[2-1] 代謝物-  
{生物種, 生理活  
性} 関係DB

[2-2] MS-化合物構  
造の関係知識DB

## [3] メタボローム統合DB

[3-1] ウィキDBによるメタボローム  
データの統合管理

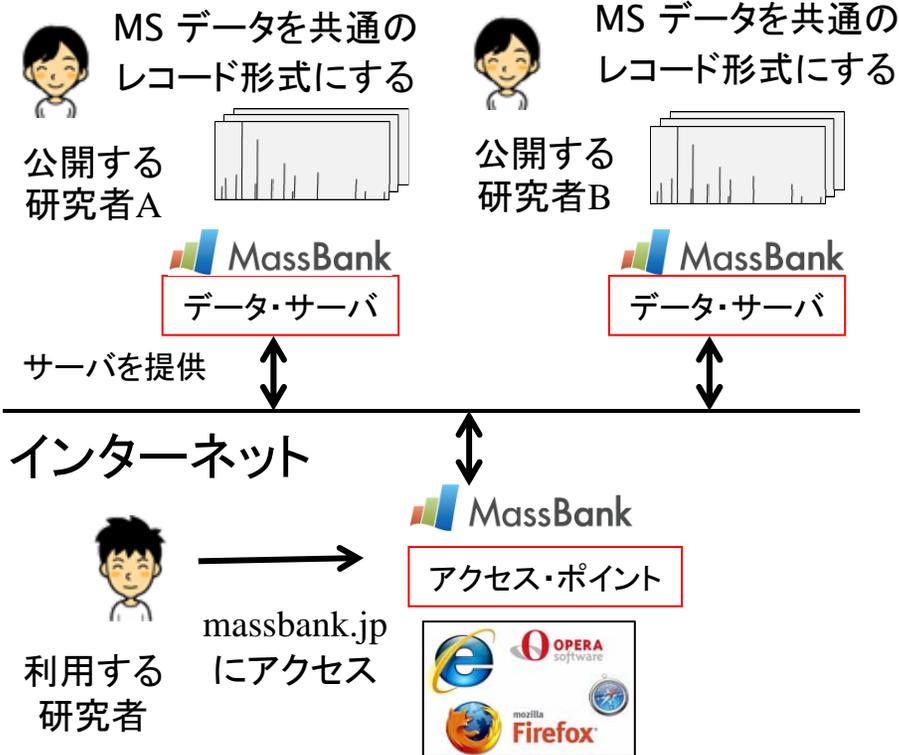
[3-2] メタボローム・アノテーシ  
ョン・システム

[3-3] ゲノム情報とのリンク

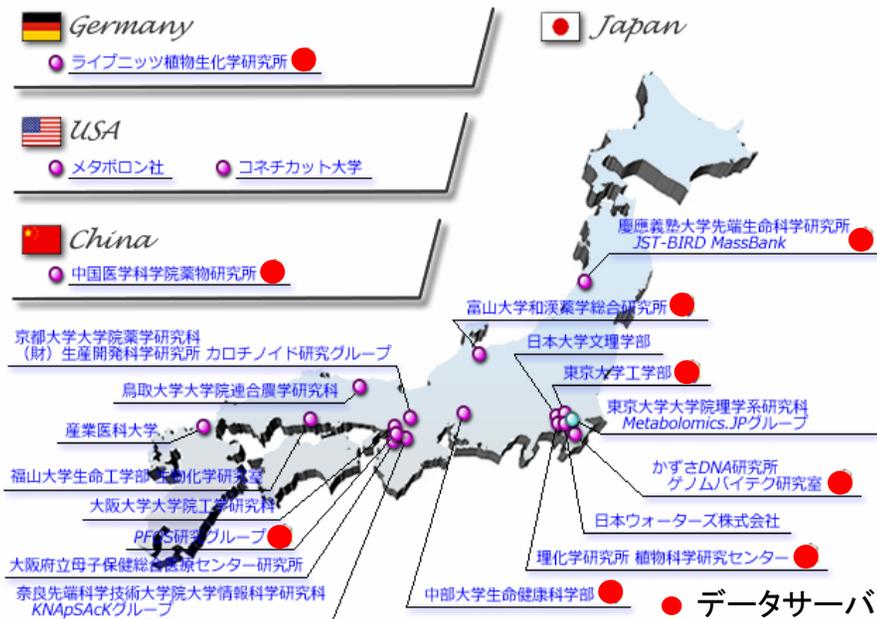
基礎研究: 持続可能社会に向けた生物資源の有効利用、生物種、分野、目的に応じた基礎研究  
産業応用: 新規有用代謝物質の探索、ゲノム育種による有用/新規代謝物質の生産 など

# [1-1] 化合物スペクトルDB (MS DB) 分散型DB(MassBank)による世界貢献

## 化合物MS DB (BIRD II期)



○分散型学術データベース  
提供者に権限を一任できる。  
**メタボロームMS DBの構築へ拡張**  
構造未知の代謝物のMS DB化  
(ゲノムにおける機能不明の遺伝子と対応)



20 研究グループが13,463化合物について  
測定した30,574 MS データを公開

訪問者数(1ヶ月間のユニークIP数) = 10,063

→ Spectral Browser → Batch Service → Browse Page → Record Index

○豊富なツールを開発  
・MSから既存代謝物の同定(バッチ検索も可能)  
・分子式、部分化学構造による既存代謝物検索  
を用いた既存代謝物の同定

# [1.1]化合物スペクトルDB (MS DB)(理研、奈良先端大)

(23年度)

(a) プロトタイプシステムの開発

(b) 学会と連携した活動

(c) Mass++との連携。大量クエリのための batch 検索機能を開発

(24年度)

(a) ミラーサーバの設置

安定なサービス提供

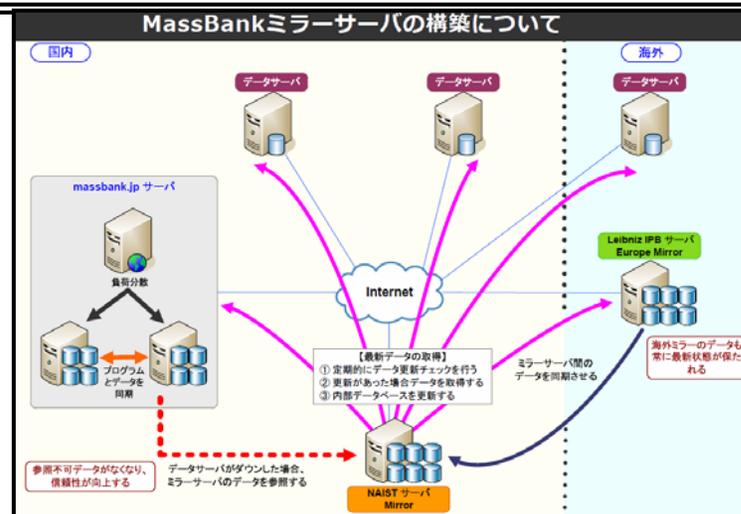
大量のLC-MS2データをクエリーとするバッチ検索の効率化

(b) 国際学会投稿システムの開発

日本質量分析学会、アジア、オセアニア地域の質量分析学会：“Mass Spectrometry”を刊行予定。投稿規定、「**著者はマススペクトルデータをMassBankで公開することがのぞましい**」

(c) 専門家によるデータ監修会の開催

(d) Mass++との連携。大量クエリのための batch 検索機能を開発



# [1-2] メタボロームDB (かずさDNA研究所・理研)

## 日本のメタボローム情報は世界一

多種多様な測定

メタボロームデータの定式化  
(代謝物ピークの抽出)

代謝物の同定  
アノテーション



ガスクロマトグラフィー  
質量分析装置  
(GC-MS)



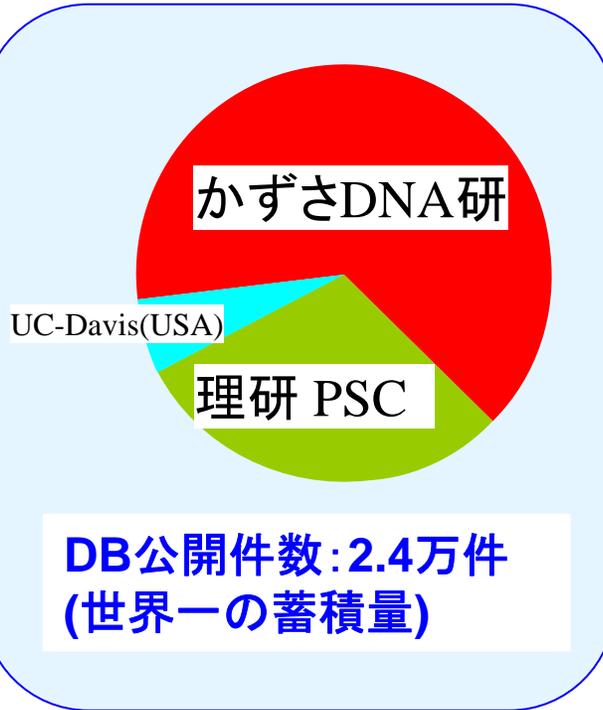
キャピラリー電気泳動  
質量分析装置  
(CE-MS)



液体クロマトグラフィー  
質量分析装置  
(LC-MS)



液体クロマトグラ  
フィー・フーリエ変換型  
質量分析装置  
(LC-FT-MS)



高機能代謝データベース

KNApSAcK (BIRD II期)  
生物種-代謝物関係DB  
(世界最大)  
代謝物(5万)、生物種-代  
謝物関係(10万対)  
(NAIST)

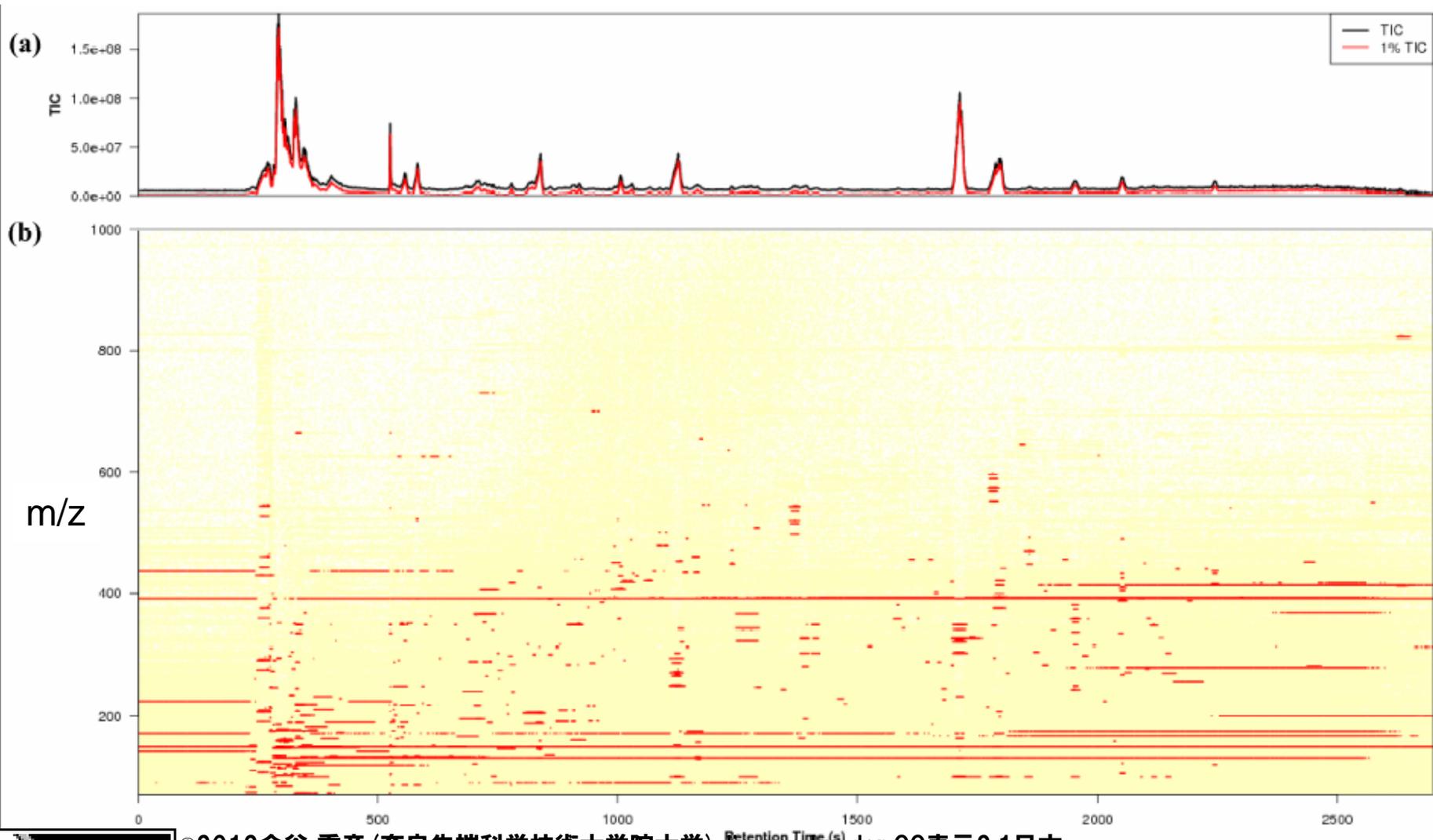
分析総数: 9.7万件

メタボローム解析のプロトタイ  
プ構築に成功

- MassBank、Mass++プロジェクトと資産を活用し、生物の総体としてのメタボロームDBの統合により世界をリード。
- 構造未知の質量スペクトルも含めデータベース化
- 目標数3.6万件を目指す。

# LC-MS

Hiroki Takahashi et al, *BMC Bioinformatics*,  
12.259(2011) AMDORAP: Non-targeted metabolic  
profiling based on high-resolution LC-MS



# メタボロームデータフォーマットの設計(かずさDNA研究所、理研)

(23年度)

メタデータ(実験条件など)、高次ピーク情報の格納方式  
KNApSAcK、MassBankとスムーズに連携

(24年度)

データの拡充

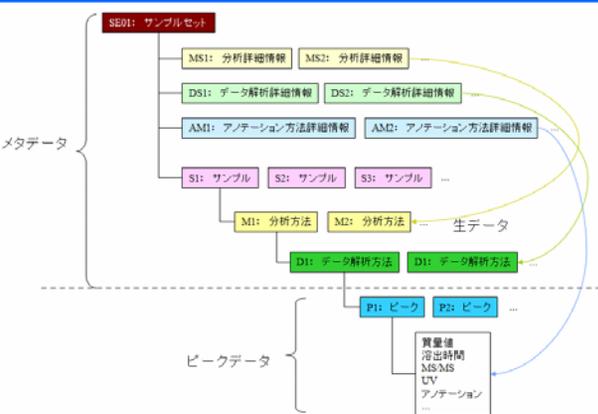
メタボロームデータの登録と、データ処理パイプラインの更なる効率化  
アノテーション用語のオントロジー構築の検討

- ①メタデータ階層構造の検討
- ②記載項目・書式の検討

- ③フォーマットに従ったサンプルデータの提供

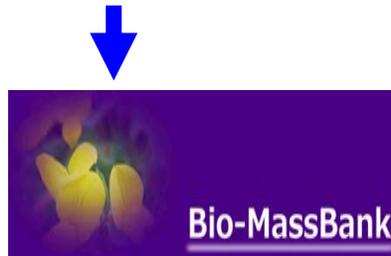
- ④フォーマットの公開

## メタボロームデータのデータ構造



## Record Index

Contributor : Kazusa (1,300)



<http://webs2.kazusa.or.jp/togodb>

# メタボロームデータの整理

メタボロームデータ変換するためのパイプラインの効率化を行った(LC-FT-MS, LC-Orbitrap-MS)

分析結果ファイル

ピーク抽出、MS/MSデータ抽出、  
MS/MSスペクトルの品質チェック、  
化合物・化合物グループ名のアノテーション

処理スキームの効率化  
(パラメータの最適化とその  
適用処理)

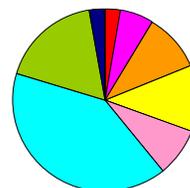
ピーク情報データ

フォーマット変換

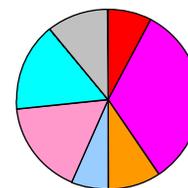
半自動処理システムの構築

共通フォーマット

アノテーションされた化合物群内訳



*A. thaliana*



*L. japonicus*

- Aminocarboxylic acids
- Sugars
- Nucleotide
- Fatty acid derivatives
- Organic acids
- Phenolics
- Flavonoids
- Glucosinolates
- Iridoids
- Others

## アノテーション用語の整理

化合物カテゴリー名 : Sugar, Aminocarboxylic acid, Nucleotide, Fatty acid derivative, Carotenoid, Coumarin, Glucosinolate, Flavonoid, Iridoid, Lignin, Organic acid, Phenolic, Porphyrin, Steroid, Terpenoid, dimer, trimer, tetramer, phosphate, sulfate, +Hex,

アノテーション状態名 : putative, hypothetical candidate,

エビデンス名 :

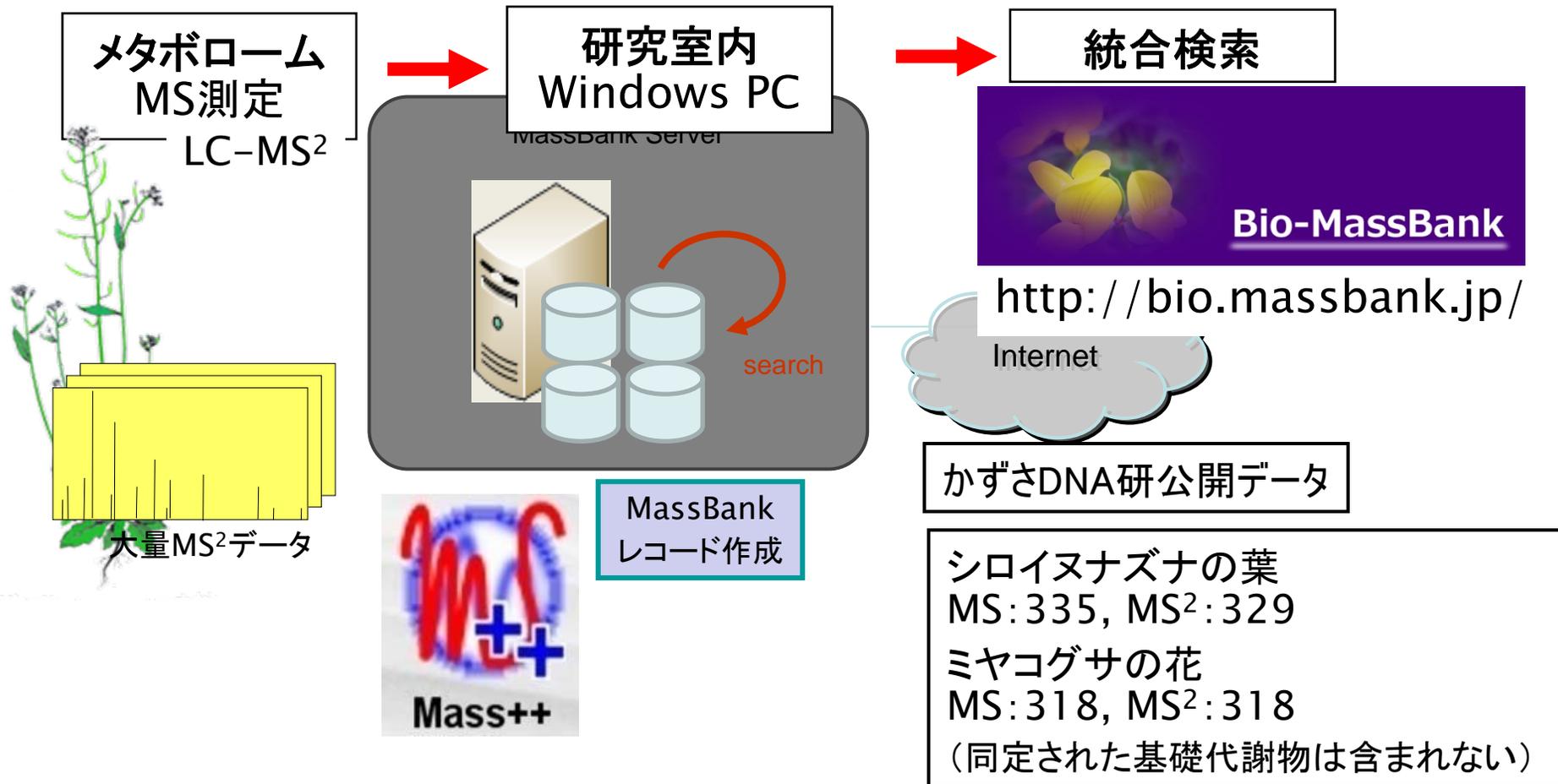
MS2 spectra suggested substructure of X(-COOH, carboxyl, ...) group

This species have this compound/related compounds/pathway/...

PDA spectra suggested substructure of X group

# [1-2]メタボロームDB:Bio-MassBank の開発 (理研、奈良先端大)

## 生物サンプル間の比較



# [1-2] メタボロームDB

Bio-MassBank の開発(理研、奈良先端大、かずさDNA研究所)

## MS2 データの類似性を利用して検索

シロイヌナズナ葉(赤地) と ミヤコグサ花(緑地) に同じ未同定代謝物質の探索

Precursor 213.02	spectra	<a href="#">KZSLJ00042</a> Prec: 213.01836 RTIME: 5.284 [M02]	<a href="#">KZSLJ00250</a> Prec: 213.01846 RTIME: 5.145 [M03]
Precursor 222.03	spectra	<a href="#">KZSAT00010</a> Prec: 222.02932 RTIME: 5.293 [M01]	<a href="#">KZSAT00188</a> Prec: 222.02932 RTIME: 5.296 [M02]
Precursor 225.08	spectra	<a href="#">KZSAT00172</a> Prec: 225.07584 RTIME: 32.652 [M01]	

ミヤコグサに  
ある代謝物

シロイヌナズナに  
ある代謝物

[KZSLJ00250](#) ----- データ ID  
Prec: **213.01846** ----- 前駆イオン *m/z*  
RTIME: **5.145** ----- 保持時間  
[M03] ----- メタデータ ID

ミヤコグサ  
シロイヌナズナの  
両方にある代謝物  
前駆イオン  
*m/z* 360.15

→ 共通代謝経路の探索

Precursor 360.15	spectra	<a href="#">KZSAT00040</a> Prec: 360.14997 RTIME: 6.021 [M01]	<a href="#">KZSAT00214</a> Prec: 360.14993 RTIME: 6.017 [M02]	<a href="#">KZSLJ00018</a> Prec: 360.15006 RTIME: 6.707 [M01]	<a href="#">KZSLJ00060</a> Prec: 360.15046 RTIME: 6.796 [M02]
---------------------	---------	--	--	--	--

# メタボローム・データベース構想

質量スペクトルから生理活性情報の統合化をめざす。

日本メタボローム・データベース  
パブリック・リポジトリ

メタボロミクス研究者、一般ユーザ  
デポジット・閲覧・アノテーション・キュレーション

## [1] 質量スペクトルDB (MS DB)

▪ 学会MSDB  
MassBank, LipidBank  
▪ 研究グループMSDB  
PRiMe

▪ 個別研究における代謝物の同定  
文献情報  
▪ 大規模メタボロームMSDB  
MassBase, MS2T, KomicMarket

[1-1] 化合物MS  
DB

[1-2] メタボローム  
MS DB

## [2] 代謝物情報DB

▪ 文献情報  
KNAPSAcK DB

▪ 質量スペクトルのアノテーション  
MassBankフラグメンテーション・ライブラリ

[2-1] 代謝物-  
{生物種, 生理活  
性}関係DB

[2-2] MS-化合物構  
造の関係知識DB

## [3] メタボローム統合DB

[3-1] ウィキDBによるメタボローム  
データの統合管理

[3-2] メタボローム・アノテーシ  
ョン・システム

[3-3] ゲノム情報とのリンク

基礎研究: 持続可能社会に向けた生物資源の有効利用、生物種、分野、目的に応じた基礎研究  
産業応用: 新規有用代謝物質の探索、ゲノム育種による有用/新規代謝物質の生産 など

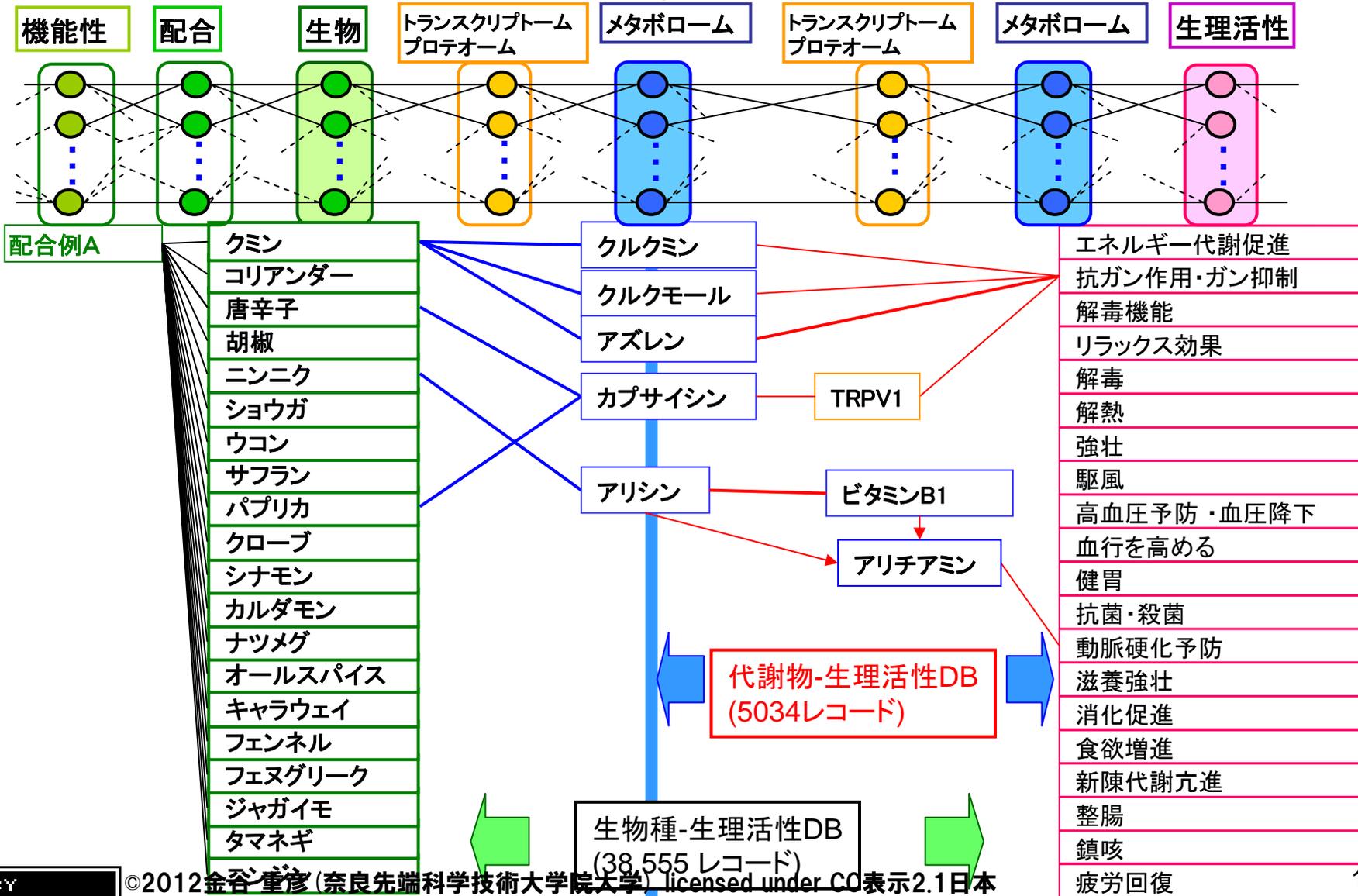
# [2-1] 代謝物情報DB: 代謝物-生理活性DB

生理活性からの有用2次代謝物探索が可能になる。

薬/食用知識

有用生物

ヒト



# [2-1] 代謝物情報DB: 代謝物-生理活性DB 生理活性からの有用2次代謝物探索が可能になる。



**“KNApSack” Family**  
Since 2008.07

**NAIST** (Institute of Science and Technology) and **PERTALIA** (Institut Pertanian Bogor) logos are present.

<p><b>Pocket</b> Search for Functional Species</p> <ul style="list-style-type: none"> <li><b>WorldMap</b> 世界の薬用植物データベース Since 2009.06</li> <li><b>Biological Activity</b> Natural Activity Since 2011.08</li> <li><b>Biological Activity</b> Metabolite Activity Since 2012.02</li> </ul>	<p><b>KNApSack Metabolomics</b></p> <p><b>KNApSack</b> Metabolomics Search Engine Since 2008.12</p>	<p><b>Picnic</b> Gene Annotation</p> <ul style="list-style-type: none"> <li><b>Arabidopsis</b> Since 2008.04</li> <li><b>Bacillus</b> Since 2008.05</li> <li><b>Human</b></li> </ul>	<p><b>Strap</b> Correlation Coefficient</p> <ul style="list-style-type: none"> <li><b>Arabidopsis</b> Since 2009.08</li> <li><b>Bacillus</b></li> </ul>
<ul style="list-style-type: none"> <li><b>KAMPO</b> 漢方薬、生薬データベース Since 2008.08</li> <li><b>JAMU</b> IndonesiaHerbデータベース Since 2009.11</li> <li><b>Lunch Box</b> 食用データベース Since 2008.07</li> <li><b>Tea Pot</b> ハーブデータベース Since 2011.08</li> </ul>	<p><b>KN</b> Core System Since 2004.04</p>	<p><b>Motorcycle</b> Metabolic Pathway 代謝データベース</p>	<p><b>PCP</b> PLANT &amp; CELL PHYSIOLOGY</p> <p><b>KNApSack Family Databases: Integrated Metabolite-Plant Species Databases for Multifaceted Plant Research</b></p> <p>Farit Mochamad Afendi<sup>1,2,7</sup>, Takeeto Okada<sup>3,7</sup>, Mami Yamazaki<sup>4,7</sup>, Aki Hirai-Morita<sup>1</sup>, Yukiko Nakamura<sup>1</sup>, Kensuke Nakamura<sup>1</sup>, Shun Ikeda<sup>1</sup>, Hiroki Takahashi<sup>1</sup>, Md. Alatu-Ul-Amin<sup>1</sup>, Latifah K. Darusman<sup>5</sup>, Kazuki Saito<sup>4,6</sup> and Shigehiko Kanaya<sup>1,6,*</sup></p> <p>Special Issue</p>

**2012.2.22**  
**NBDCのアーカイブシステムに格納完了**  
**(松矢さん 調査担当)**



# Metabolite Activity



代謝物情報  
生物活性  
生物種  
から検索できる

Select by ...

Metabolite  C\_ID  CAS\_ID  Biological Activity (Function)  Targeted Species  Molecular formula

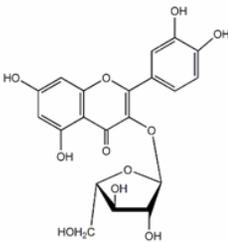
List Clear

## Arabidopsis

Number of matched data : DB match=8

C_ID	Biological Activity (Function)	Target Species	Reference
<a href="#">C00000176</a>	induce peduncle elongation	<i>Arabidopsis thaliana</i>	Clouse,Brassinosteroids-Chemistry,Bioactivity and Applications,(1991),122-140
<a href="#">C00004631</a>	Allelopathy	<i>Arabidopsis thaliana</i>	Parvez,3rd World Congress on Allelopathy, (2002),p.247
<a href="#">C00004632</a>	Allelopathy	<i>Arabidopsis thaliana</i>	Parvez,3rd World Congress on Allelopathy, (2002),p.247
<a href="#">C00005367</a>	Allelopathy	<i>Arabidopsis thaliana</i>	Parvez,3rd World Congress on Allelopathy, (2002),p.247
			Parvez,3rd World Congress on Allelopathy, (2002),p.247
			Parvez,3rd World Congress on Allelopathy, (2002),p.247
			Parvez,3rd World Congress on Allelopathy, (2002),p.247
			Parvez,3rd World Congress on Allelopathy, (2002),p.247

input word = C00005367

Metabolite information		Structural formula		
Name	Avicularin Avicularine Avicularoside Quercetin 3-alpha-L-arabinofuranoside Quercetin 3-O-alpha-L-arabinofuranoside	 <p>zoom in</p>		
Formula	C20H18O11			
Mw	434.08491142			
CAS RN	572-30-5			
C_ID	C00005367			
Organism	Kingdom	Family	Species	Reference
	Plantae	Ericaceae	Arctostaphylos uva-ursi	<a href="#">Ref.</a>
	Plantae	Ericaceae	Ledum palustre	<a href="#">Ref.</a>
	Plantae	Ericaceae	Richea angustifolia	<a href="#">Ref.</a>
	Plantae	Ericaceae	Richea scoparia	<a href="#">Ref.</a>
	Plantae	Ericaceae	Vaccinium myrtillus	<a href="#">Ref.</a>
	Plantae	Hypericaceae	Hypericum scabrum	<a href="#">Ref.</a>
	Plantae	Juglandaceae	Juglans regia	<a href="#">Ref.</a>
	Plantae	Myrtaceae	Pimenta dioica	<a href="#">Ref.</a>
	Plantae	Myrtaceae	Psidium guajava	<a href="#">Ref.</a>
	Plantae	Polygonaceae	Polygonum aviculare	<a href="#">Ref.</a>
	Plantae	Rosaceae	Chaenomeles sinensis KOEHNE	<a href="#">Ref.</a>
	Plantae	Rosaceae	Cowania mexicana	<a href="#">Ref.</a>
	Plantae	Rosaceae	Dryas octopetala	<a href="#">Ref.</a>
Plantae	Solanaceae	Solanum glaucophyllum	<a href="#">Ref.</a>	
Plantae	Taxodiaceae	Taxodium distichum	<a href="#">Ref.</a>	

(23年度)  
代謝物-生理活性DB のプロトタイプ  
Metabolite Activityを開発。

(24年度)  
悉皆的データの充実を図る。

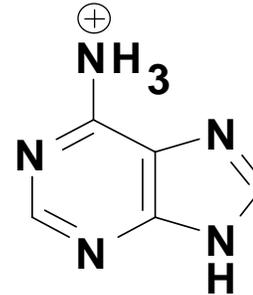
# [2-2] 代謝物情報DB: MS-化合物構造の関係知識DB

## 部分構造による同定率の向上をめざす。

イオンピークと部分化学構造式の関係ルールDB

136.057  
C<sub>5</sub>H<sub>6</sub>N<sub>5</sub><sup>+</sup>

1.0  
←→  
0.57

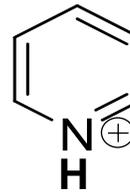


(BIRD II期)

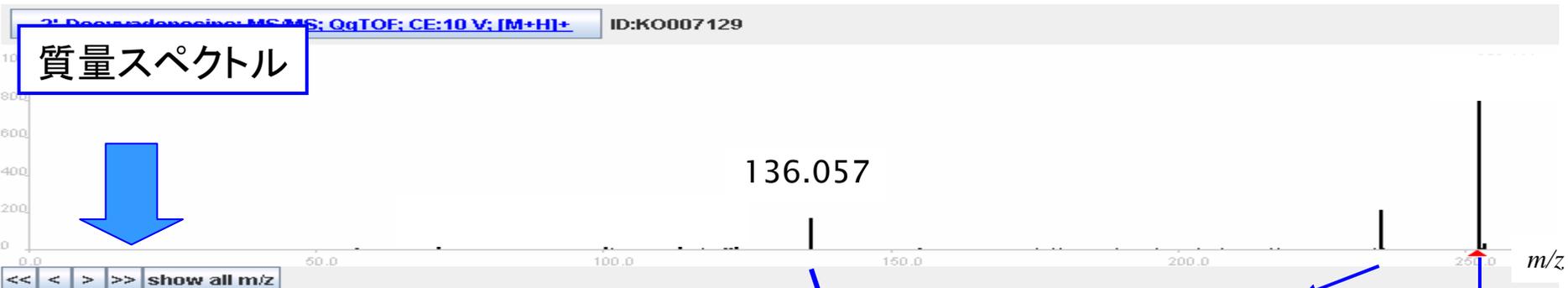
標品サンプル: 3000種  
報告代謝物数: 5-6万種

78.031  
C<sub>5</sub>H<sub>4</sub>N<sup>+</sup>

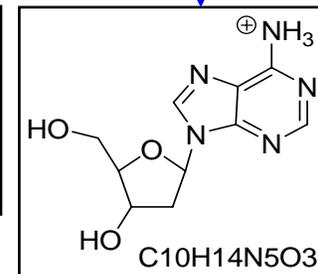
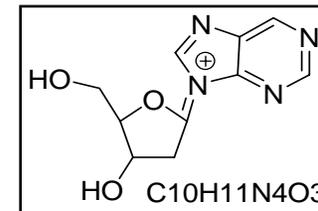
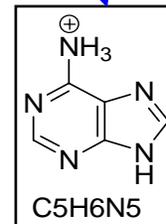
0.76  
←→  
0.74



質量スペクトル



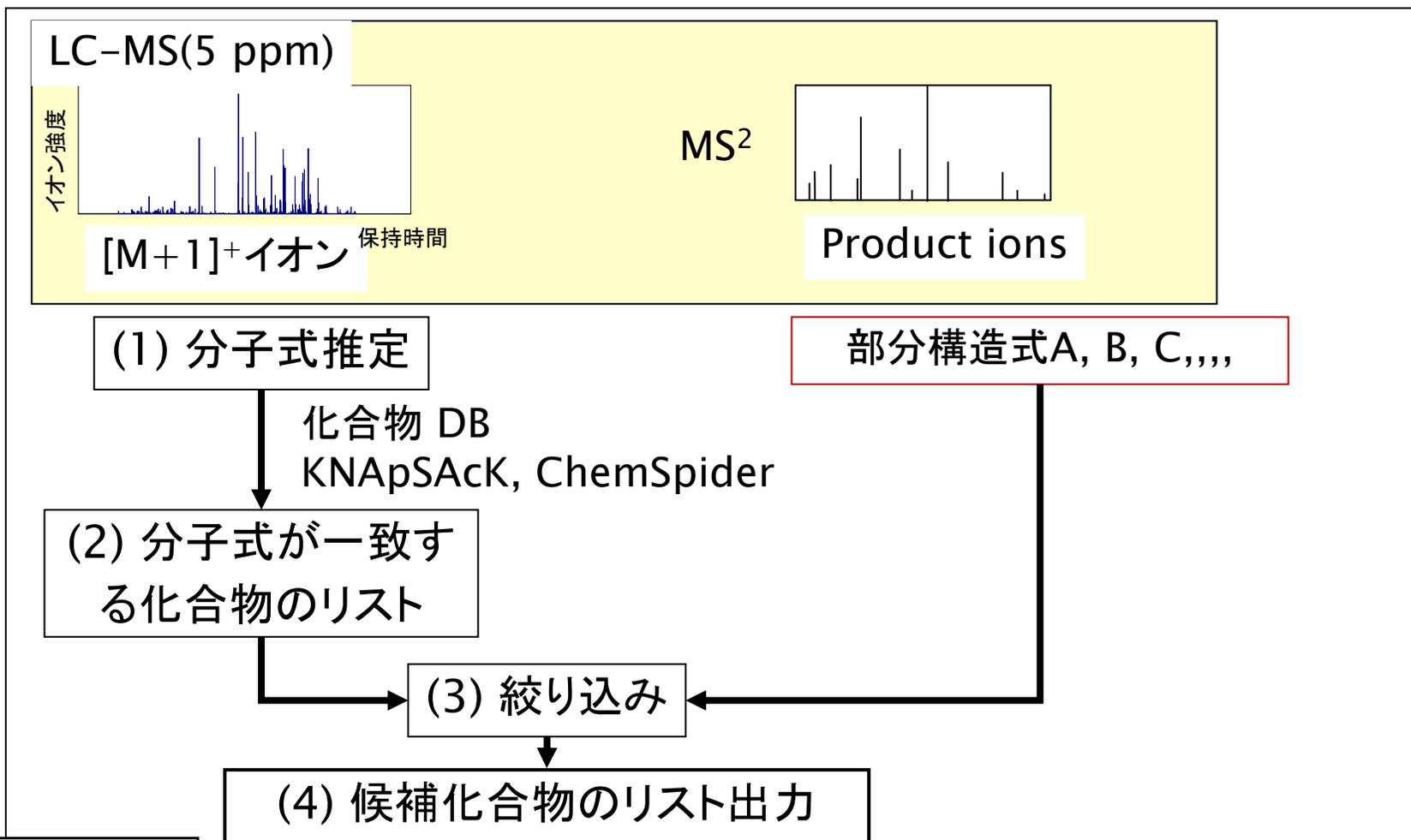
部分構造推定



構造情報に基づいた絞込み: 生物種-代謝物関係DB(KNApSAcK)など

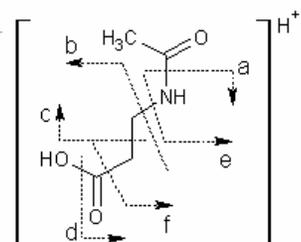
## [2-2] 化学構造式推定ツール(1) (奈良先端大)

23年度 化学構造式推定ツールの開発  
24年度 予測精度の向上(二次代謝物質)



# [2.2] 化学構造式推定ツールの開発 (2)

N-Acetyl-b-alanine A138+

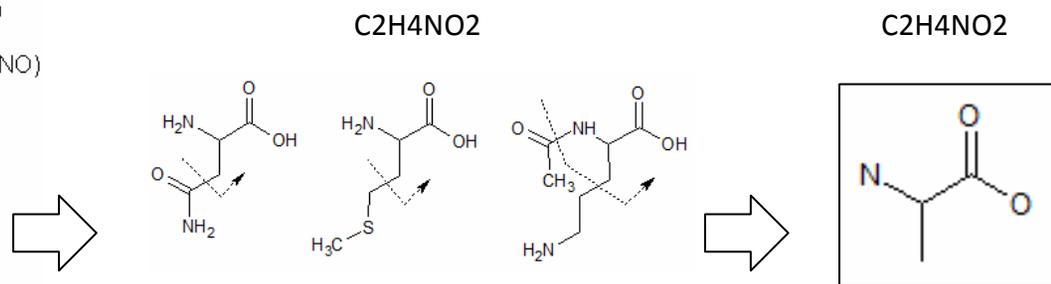


114: d (C<sub>5</sub>H<sub>8</sub>NO<sub>2</sub>)  
 90: a (C<sub>3</sub>H<sub>8</sub>NO<sub>2</sub>)  
 86: f (C<sub>4</sub>H<sub>8</sub>NO)  
 72: c[a+d] (C<sub>3</sub>H<sub>6</sub>NO)  
 69: (C<sub>5</sub>H<sub>9</sub>)  
 60: e (C<sub>2</sub>H<sub>6</sub>NO)  
 55: b+d (C<sub>3</sub>H<sub>3</sub>O)

M+1=132

M+1=132 → 114,90,72,55

## ピークと部分構造式の関係



Step 1

解裂スキームの作成

Step 2

ピークの分子式ごとに  
部分化学構造式を収集

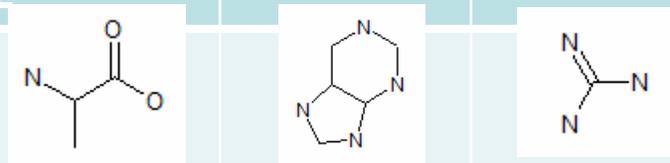
Step 3

分子式に共通な部分  
化学構造式を収集

ピークの分子式 (269) →



部分化学構造式 (94) →



ピークと部分化学構造式の関係 (630 ペア)

# [2.2] 化学構造式推定ツールの開発 (3)

クエリMS<sup>2</sup>データ  
ピークm/zと分子式

Matched Formulae : 15

m/z	39.0235	41.0396	42.0321	54.0339	55.0413	56.0497	66.0351	68.0503	81.0459	82.0547	83.0616	93.0434	95.0591	110.072	156.076
Formula	C3H3	C3H5	C2H4N	C3H4N	C3H5N	C3H6N	C4H4N	C4H6N	C4H5N2	C4H6N2	C4H7N2	C5H5N2	C5H7N2	C5H8N3	C6H10N3O2
No.	--	--	--	ion-pos-0020 ion-pos-0021	--	ion-pos-0016 ion-pos-0018 ion-pos-0019 ion-pos-0020 ion-pos-0021	ion-pos-0021	ion-pos-0016 ion-pos-0019 ion-pos-0020	ion-pos-0016 ion-pos-0019 ion-pos-0020 ion-pos-0021	ion-pos-0016 ion-pos-0020 ion-pos-0021	ion-pos-0016 ion-pos-0019 ion-pos-0020 ion-pos-0021	ion-pos-0020 ion-pos-0021	ion-pos-0016 ion-pos-0019 ion-pos-0020 ion-pos-0021	ion-pos-0020	ion-pos-0020 ion-pos-0054

Hit Relationships : 6

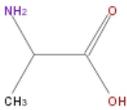
**ion-pos-0016**  
Substructure



Formula, Precision & Recall, TP

C4H6N2	0.71	0.15	10
C4H7N2	0.5	0.16	11
C5H7N2	0.7	0.21	14
C3H6N	0.45	0.16	10

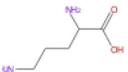
**ion-pos-0018**  
Substructure



Formula, Precision & Recall, TP

C3H6N	0.41	0.39	43
-------	------	------	----

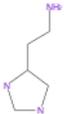
**ion-pos-0019**  
Substructure



Formula, Precision & Recall, TP

C5H7N2	0.5	0.4	10
C3H6N	0.55	0.24	6
C4H5N2	0.57	0.32	8
C4H6N	0.64	0.28	7

**ion-pos-0020**  
Substructure



Formula, Precision & Recall, TP

C4H6N2	0.5	0.31	21
C4H7	0.61	0.34	23
C5H5	0.3	0.19	13
C6H7	0.8	0.22	15

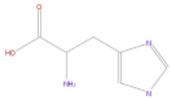
[M+1]<sup>+</sup>  
C6H10N3O2



分子式に関係づけられた部分化学構造式

Results : 2 Hit.

**Histidine**



Formula : C6H9N3O2  
Exact Mass : 155.0695

DB Links  
KNApSAck : [C00001363](#)  
KEGG : [C00135](#)  
KEGG : [C00738](#)

Hit Relationship-No.  
ion-pos-0016, ion-pos-0018, ion-pos-0019, ion-pos-0020,  
ion-pos-0021, ion-pos-0054

**beta-Pyrazol-1-ylalanine**



Formula : C6H9N3O2  
Exact Mass : 155.06948

DB Links  
KNApSAck : [C00001390](#)  
KEGG : [C01162](#)

Hit Relationship-No.  
ion-pos-0018

推定された  
化合物の候補

# メタボローム・データベース構想

質量スペクトルから生理活性情報の統合化をめざす。

日本メタボローム・データベース  
パブリック・リポジトリ

メタボロミクス研究者、一般ユーザ  
デポジット・閲覧・アノテーション・キュレーション

## [1] 質量スペクトルDB (MS DB)

▪ 学会MSDB  
MassBank, LipidBank  
▪ 研究グループMSDB  
PRIME

▪ 個別研究における代謝物の同定  
文献情報  
▪ 大規模メタボロームMSDB  
MassBase, MS2T, KomicMarket

[1-1] 化合物MS  
DB

[1-2] メタボローム  
MS DB

## [2] 代謝物情報DB

▪ 文献情報  
KNAPSAcK DB

▪ 質量スペクトルのアノテーション  
MassBankフラグメンテーション・ライブラリ

[2-1] 代謝物-  
{生物種, 生理活  
性}関係DB

[2-2] MS-化合物構  
造の関係知識DB

## [3] メタボローム統合DB

[3-1] ウィキDBによるメタボローム  
データの統合管理

[3-2] メタボローム・アノテーシ  
ョン・システム

[3-3] ゲノム情報とのリンク

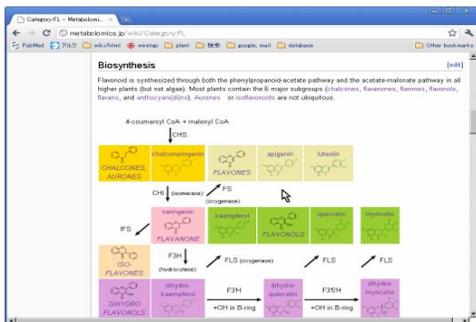
基礎研究: 持続可能社会に向けた生物資源の有効利用、生物種、分野、目的に応じた基礎研究  
産業応用: 新規有用代謝物質の探索、ゲノム育種による有用/新規代謝物質の生産 など

# [3-1] メタボローム統合DB: ウィキ・データベース

## 分散DBによるメタボローム知の共有をめざす。

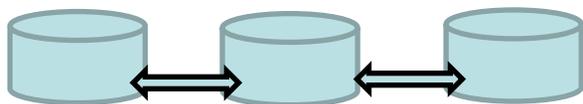
MediaWikiで情報統合: 各データを名前空間に分けて登録。検索によりページを再構成。

一般ユーザから  
情報収集

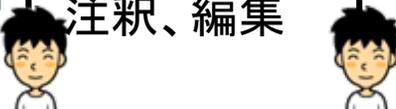


↑

専門家による  
注釈、編集



サーバ間も検索を可能にし、見かけ上は単一DBとして機能

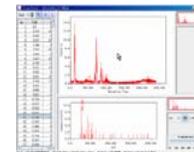


DB型ウィキ  
「フラボノイド・データベース」  
「ウィキによるフラグメント・ライブラリ」  
「イオンと部分化学構造式の関係」

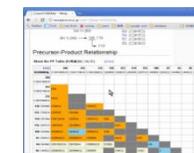
- 一貫性の維持、
  - 分散して複数のサーバを配置可能、
  - 知識の集約、
  - 持続可能なDB、
- としての有用性を実証した。



[1-1] 化合物MS DB



[1-2] メタボローム DB



[2-1] MS-化合物構造  
の関係知識 DB



[2-2] 代謝物  
- {生理活性, 生物種}  
関係DB



[3-3] ゲノム情報との  
リンク



[3-2] メタボローム・アノテーション・システム

## [3-3] 統合化、標準化に向けて(理研)

(23年度)

### CAS, InChIコード逆引き辞書の作成

基礎代謝物1800化合物についてCAS, InChIコードか構造情報を逆引きするwikiページを構築した。KNAPSAcKおよびFlavonoid DBのデータも追加する予定。

### MassBankのローカル版構築

個別のPC上で動くMassBankの検索エンジンを実装し、化合物名や計測条件など文字列検索も自由におこなえるJavaアプリケーションを作成した。

### トリ、セスキテルペンの構造分類

トリ・セスキテルペンの骨格を生合成経路に基づいて分類し、とりわけ合成酵素における骨格のリアレンジメントがわかる形でまとめたwikiページを作成した。

### メタボローム国際ワークショップの開催:標準化

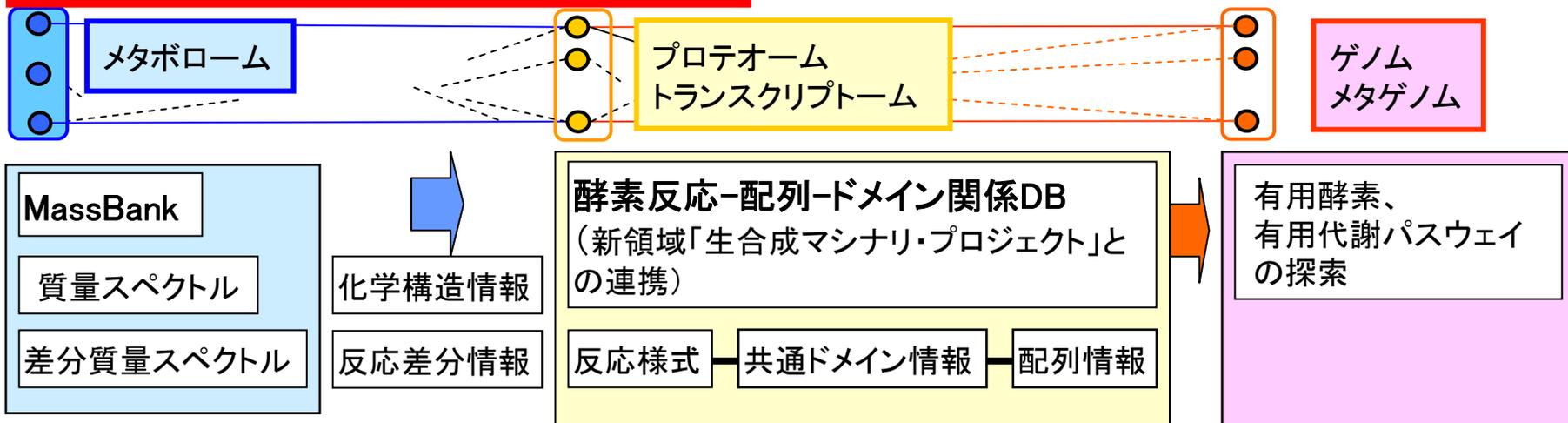
日米欧の主要PIを集め、データの公開とInChIとPubChem番号をIDとして利用することに同意した。

(24年度)

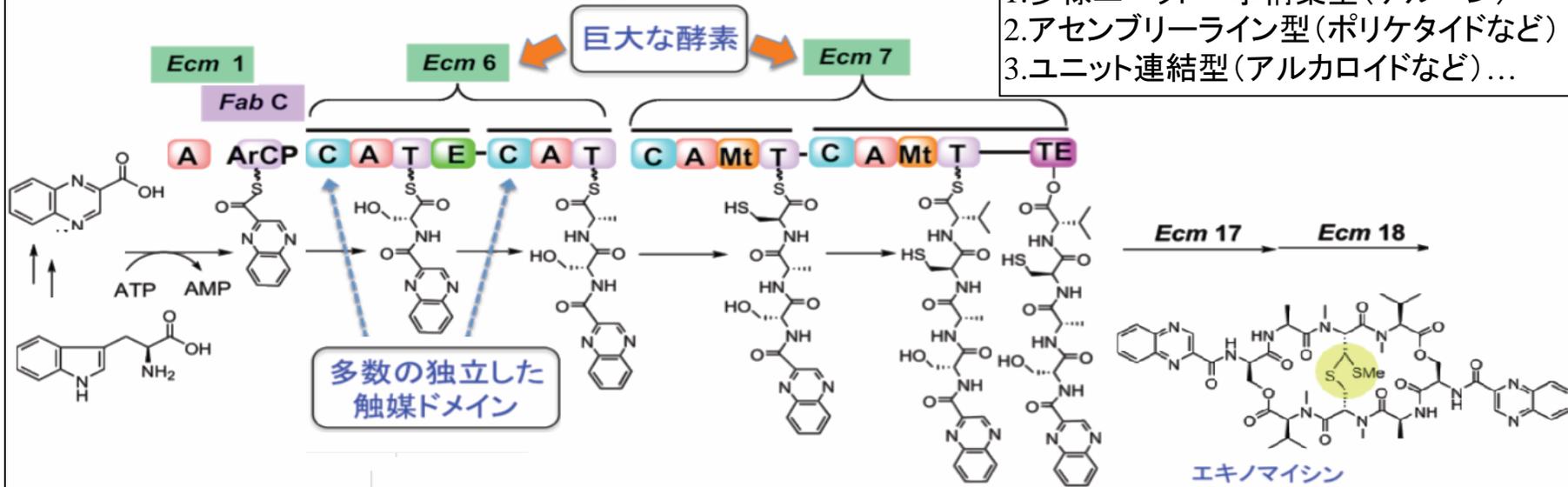
日米欧(2012.2.17-19)の合意に基づいたデータの標準化ならびに統合を図る。「低炭素社会のためのメタボロミクス」

# [3-3] ゲノム情報とのリンク(奈良先端大)

## 2次代謝物から遺伝情報とのリンク



### 非リボソーム依存性ペプチド合成酵素



将来的には、新型シーケンサから出力されるトランスクリプトームをもとに二次代謝経路の推定を行いたい。



# “KNApSack” Family

Since 2008.07

### Pocket

Search for Functional Species

- WorldMap**  
世界の薬用植物データベース  
Since 2009.06
- Biological Activity**  
Natural Activity  
Since 2011.08
- Biological Activity**  
Metabolite Activity  
Since 2012.02
- 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

### KNApSack Metabolomics

- KNApSack**  
Metabolomics Search Engine  
Since 2008.12
- KNApSack**  
Core System  
Since 2004.04

### Picnic

Gene Annotation

- Arabidopsis**  
Since 2008.04
- Bacillus**  
Since 2008.05
- Human**  
Since 2009.03
- MOUSE**

### Strap

Correlation Coefficient

- Arabidopsis**  
Since 2009.08
- Bacillus**

### Motorcycle

Metabolic Pathway

代謝データベース  
Since 2011.08

### Pickaxe

Metalloprotein Database

**Motorcycle**

Motorcycle Keyword Search

Select by ...  
 KRID  Enzyme  Species  Gene Name  
 Phomopsis amygdali [List] [Clear] [Page Clear]

Equation  
 \_\_\_\_\_ and \_\_\_\_\_  
 [List] [Clear] [Page Clear]

Motorcycle Blastp Search  
 BLASTP Search

生物種、酵素名、遺伝子名から酵素反応の検索

アミノ酸配列から酵素反応の検索

特定領域研究「バイオマシナリー」との協同作業

**Motorcycle**

アミノ酸配列データベースと比較します。

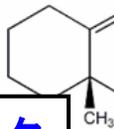
```

MTIMDIDHSLRDLAVSLIRGAAEGYSPKYGYGSMSCAAAYDTAWVSLYAKFPNGIKKWLFPQSFLLLENGENDGSGWR
QTSFYDRILNTAAPLLSLORHAREPLOLHDVCEGDLDORHRATLSLORLSDWIDIASTAHVGFIEIIVPALLNLLAAEG
LCFSFKAGDELKMKVNAARLKRFDPEVLLYGRKHTLLHLSLEAFVGIIDFDKVVHHRVGGFSMASPATAAYLMNASQWDD
EAEDYIKNVLNNGSRGHRGAVPSPAYPSNSFEYVLLSTLLHAGFTAKDIEOPELCOVAGMLNENFTEGGGIFGARSISS
DADDTAKAAFALNKLGYDVSYNEMVKEFEYKHFOTYPSERDASL.SANCNTLLALLHOKOVGAYOPQILKCVKFLTRCWW
NTDGPTRDKWNSHLVSTMLMVAQALTEFOAILDOKGLPYGLNTVEMARYSICLFGGCLRTMLQOCEDEGWSHSREQTAYA
VLT.LGOARRLSTFKHLQSGIDSAIDAAATFIRLHSDVLAQOGLPEFVTEKVSYSYSPLYTEAYCLAALKVATSLVDMPI
VGESLDLGI.PSRORIDK.YIWLHOTP.LFRSLPEVOLRASFEIHLFLPIVNEHRLDVFPK.NMDDDDYIRL.IPF.TWTAT
NINRNF.TASP.WLDMIMVSYVDYQADEFMEAVAGLTFSEDL.SMLVGLIQEVLTPYKTEPASPYTSLIDMSVYOCPRAKL
SN.LASGDIEEYR.TCLRRFASFLLDHPAVRNAQRDRATAWREYHNYLVAHVHRTODNMRLNLGEOQRWYYSRNPYFHW
  
```

Search Clear Page Clear

同源性検索BLASTP

All rights reserved. © 2011 NARA INSTITUTE OF SCIENCE and TECHNOLOGY

Metabolite Information		Structure											
Name	5-epi-Aristolochene												
Formula	C15H24												
Mw	20												
CAS RN													
C_ID	C00007635												
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>Cruciferae</td> <td>Arabidopsis thaliana</td> <td><a href="#">Ref.</a></td> </tr> <tr> <td>Plantae</td> <td>Solanaceae</td> <td>Nicotiana tabacum</td> <td><a href="#">Ref.</a></td> </tr> </tbody> </table>	Kingdom	Family	Species	Reference	Plantae	Cruciferae	Arabidopsis thaliana	<a href="#">Ref.</a>	Plantae	Solanaceae	Nicotiana tabacum	<a href="#">Ref.</a>
Kingdom	Family	Species	Reference										
Plantae	Cruciferae	Arabidopsis thaliana	<a href="#">Ref.</a>										
Plantae	Solanaceae	Nicotiana tabacum	<a href="#">Ref.</a>										

**生物種-代謝物関係データへのリンク**

Select Keyword = KRID  
input word = KR0001800

KRID	KR0001800
Enzyme	5-epi-Aristolochene synthase
KEGG ID	--
EC	--
Equation	(E,E)-Farnesyl diphosphate ->  5-epi-Aristolochene (79) +  Pyrophosphate
Rclass	Terpene
Rsubclass	Sesquiterpenoids
FinalProduct	--
Eclass	Sesquiterpene synthase
Reaction Mechanism	<a href="#">ME000035.gif</a>
Pathway	--
Curator	Shigehiko KANAYA

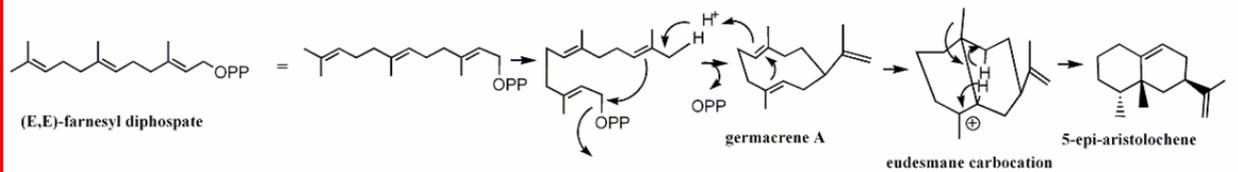
**反応情報**

代謝物-生物

**反応メカニズムへのリンク**

Sequences producing significant alignments:

Accession	Protein	Score	E Value
KR0001800	Sesquiterpene synthase Terpene Sesquiterpenoids	Nic... 1100	0.0
KR0001803	Sesquiterpene synthase Terpene Sesquiterpenoids	Nic... 1053	0.0
KR0001802	Sesquiterpene synthase Terpene Sesquiterpenoids	Nic... 1043	0.0
KR0001804	Sesquiterpene synthase Terpene Sesquiterpenoids	Nic... 1039	0.0
KR0001805	Sesquiterpene synthase Terpene Sesquiterpenoids	Nic... 972	0.0
KR0001801	Sesquiterpene synthase Terpene Sesquiterpenoids	Cap... 880	0.0
KR0001814	Sesquiterpene synthase Terpene Sesquiterpenoids	Gos... 873	0.0
KR0001835	Sesquiterpene synthase Terpene Sesquiterpenoids	Hyo... 789	0.0
KR0001777	Sesquiterpene synthase Terpene Sesquiterpenoids	Pog... 601	e-174
KR0001795	Sesquiterpene synthase Terpene Sesquiterpenoids	Vit... 540	e-155
KR0001813	Sesquiterpene synthase Terpene Sesquiterpenoids	Gos... 528	e-152
KR0001811	Sesquiterpene synthase Terpene Sesquiterpenoids	Gos... 527	e-151
KR0001784	Sesquiterpene synthase Terpene Sesquiterpenoids	Pop... 527	e-151
KR0001812	Sesquiterpene synthase Terpene Sesquiterpenoids	Gos... 523	e-150
KR0001826	Sesquiterpene synthase Terpene Sesquiterpenoids	San... 520	e-149
KR0001810	Sesquiterpene synthase Terpene Sesquiterpenoids	Gos... 520	e-149
KR0001825	Sesquiterpene synthase Terpene Sesquiterpenoids	Vit... 504	e-145
KR0001824	Sesquiterpene synthase Terpene Sesquiterpenoids	Cit... 494	e-141
KR0001848	Sesquiterpene synthase Terpene Sesquiterpenoids	Med... 489	e-140
KR0001762	Sesquiterpene synthase Terpene Sesquiterpenoids	Med... 470	e-134



# 連携体制：世界最大級の連携体制をめざす。

[1-1]化合物MS DB, [1-2]メタボロームMS DB (かずさDNA研究所、理研)

1. 日本脂質生化学会(LipidBankデータベース)
2. 日本質量分析学会(MassBankデータベース)

1. 富山大学・和漢医薬学総合研究所・資源開発研究部門(田中 謙先生)
2. 理研・植物科学センターメタボローム研究推進部門(斎藤和季先生)
3. かずさDNA研究所・産業基盤開発研究部(柴田大輔先生)
4. 中部大学生命健康科学部(田口 良先生)
5. 千葉大学・院・薬学研究科(山崎真巳先生)
6. 大阪府立大学・院・生命環境科学研究科(太田大策先生)
7. 広島大学・院・医歯薬学総合研究科(升島 努先生)など

1. ドイツ・ライプニッツ植物化学研究所(Dr. Steffen Neumann)
2. フィンランド・ヘルシンキ大学計算機科学科(Prof. Juho Rousu)
3. スウェーデン・ウメオ植物科学研究センター(Prof. Thomas Moritz)
4. ドイツ・マックスプランク植物分子生理学研究科 (Dr. Dirk Walther)

[2-1] MS-化合物構造の  
関係知識 DB (NAIST)

[2-2]代謝物-{生理活性, 生物  
種}関係DB (NAIST)

酵素反応-ペプチド配列-塩基  
配列関係DB(新領域「生合成  
マシナリープロジェクト」)

[3]メタボローム統合DB: ウィキ・データベース(理研)

# データの統合に向けて

Traditional & Modern Knowledge of Medicinal Plants

Plant Omics

Human Omics

Therapeutic Usage

Prescription

Medicinal Herb.

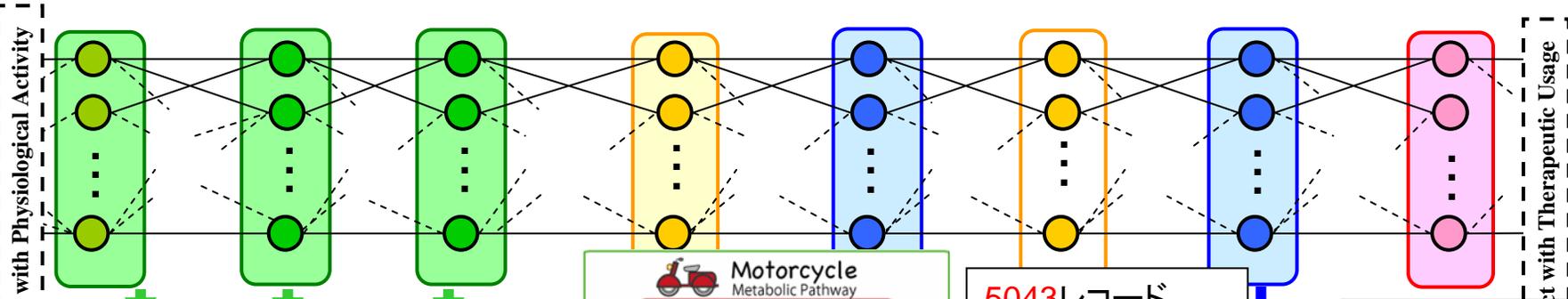
Proteome Interactome Transcriptome

Metabolomics

Proteome Interactome Transcriptome

Metabolomics

Physiological Activity



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

336種の漢方処方  
278種生薬

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

5310の配合処方  
1133種生薬

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

709種の食用生物

**Motorcycle**  
Metabolic Pathway  
代謝データベース  
Since 2011.08

163カ国:  
41,548対の薬用植物と使用国の関係

5043レコード  
**Biological Activity**  
Metabolite Activity  
Since 2012.02

50048種の代謝物  
101500対の生物種-代謝物の関係

**KNApSack**  
Core System  
Since 2004.04

38,555レコード  
**Biological Activity**  
Natural Activity  
Since 2011.08

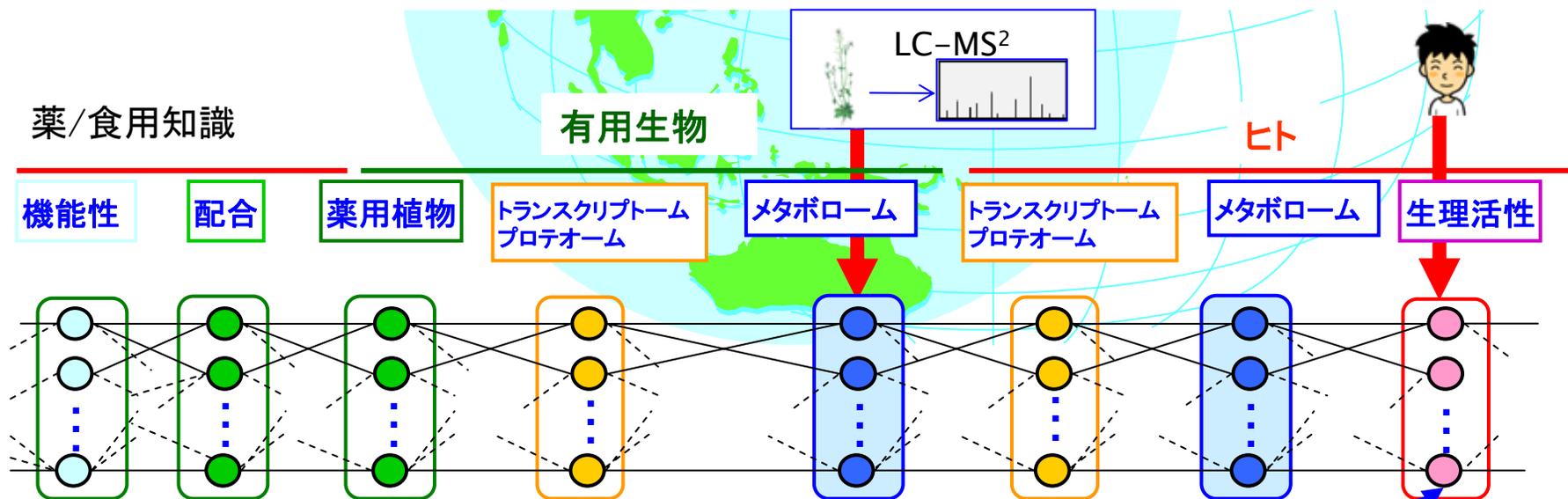


MassBank

Bio-MassBank

# 科学的・社会的なインパクト

- メタボローム研究の世界の拠点(パブリック・リポジトリ)
- 低コスト運営、持続可能なメタボロームを中心とした知の集約
- 生物種、分野、目的に応じたメタボロミクス研究基盤
- 新規有用代謝物質の探索の基盤
- ゲノム育種による有用/新規代謝物質の生産へ向けた基盤
- ゲノム科学への入り口としての啓発活動による社会貢献
- 持続可能社会に向けた生物資源の有効利用



# メタボローム・データベースの開発

金谷 重彦・西岡孝明

奈良先端科学技術大学院大学 (NAIST)

情報科学研究科・情報生命科学専攻

・計算システムズ生物学講座

櫻井 望

(財)かずさDNA研究所・

産業基盤開発研究部

有田 正規

(独)理化学研究所

植物科学研究センター

平成24年2月24日