

# メタボローム・データベースの開発

金谷 重彦

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  
・比較ゲノム学講座

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23年6月10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

機能性

配合

薬用植物

ゲノム  
トランスクリプトーム  
プロテオーム

メタボローム

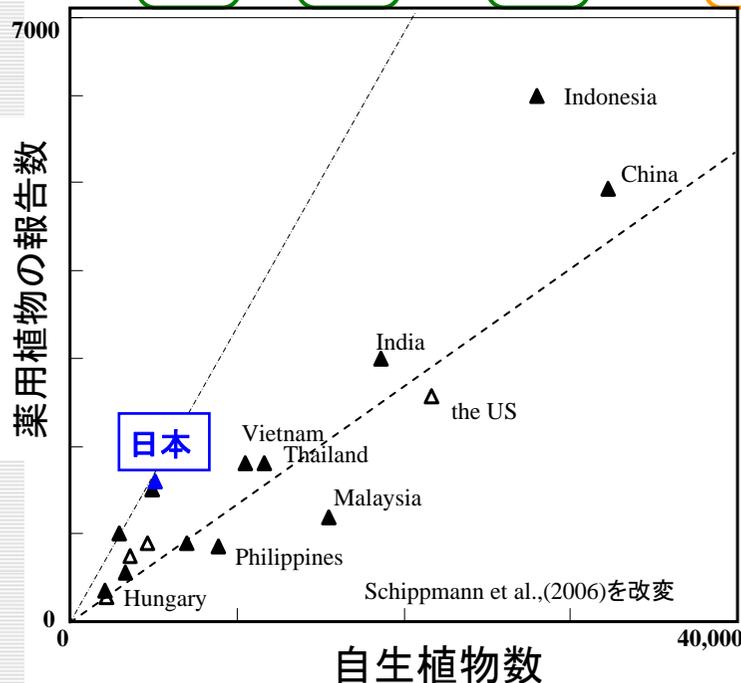
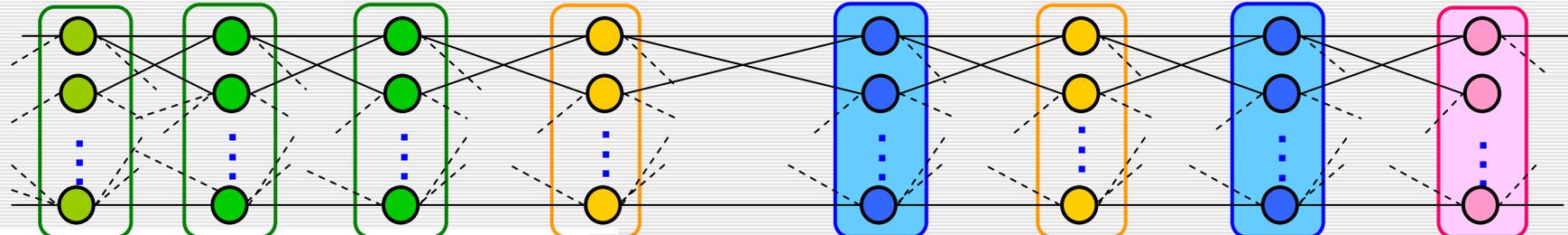
ゲノム  
トランスクリプトーム  
プロテオーム

メタボローム

生理活性

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

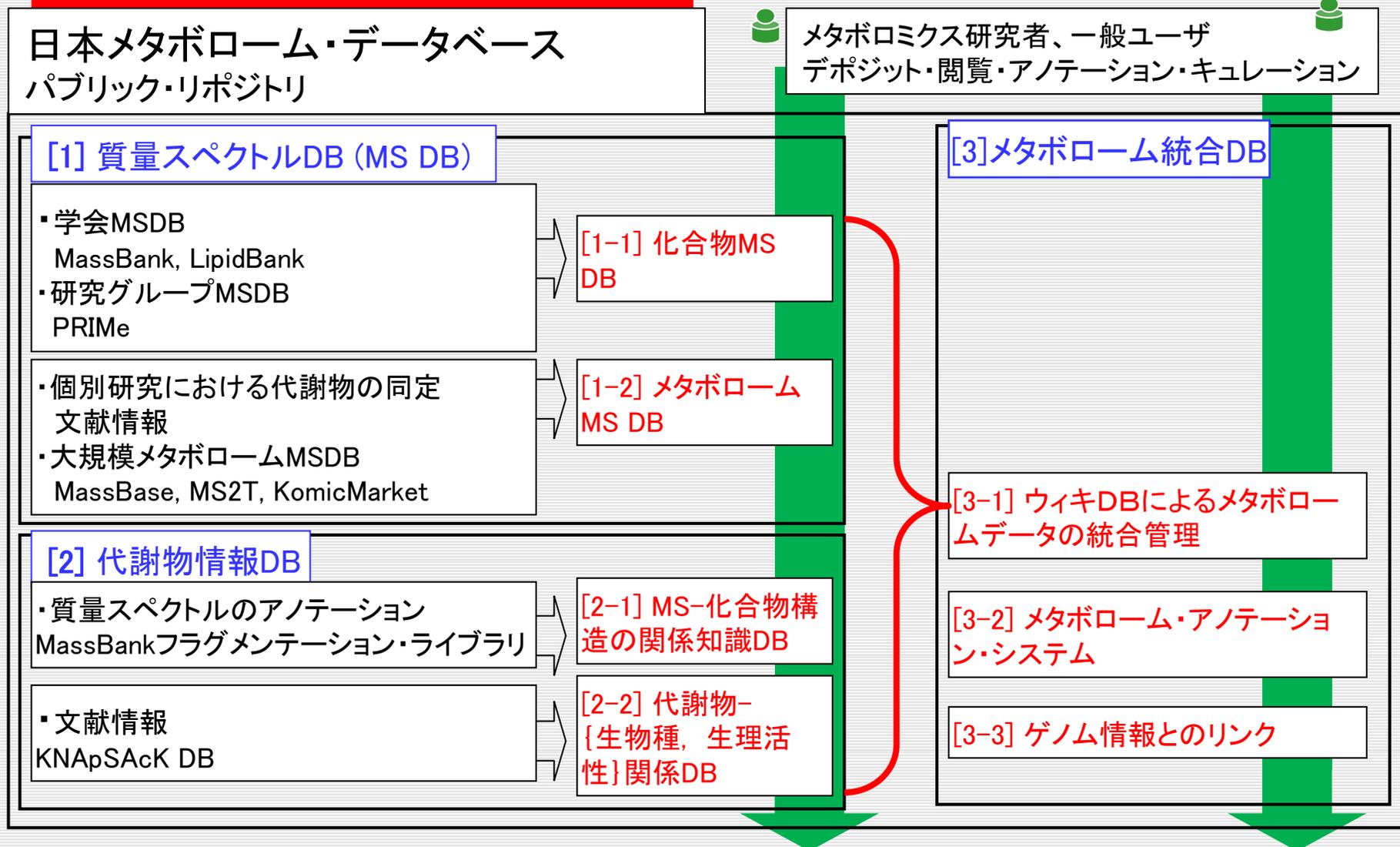
2次代謝物は生理活性の宝庫

(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

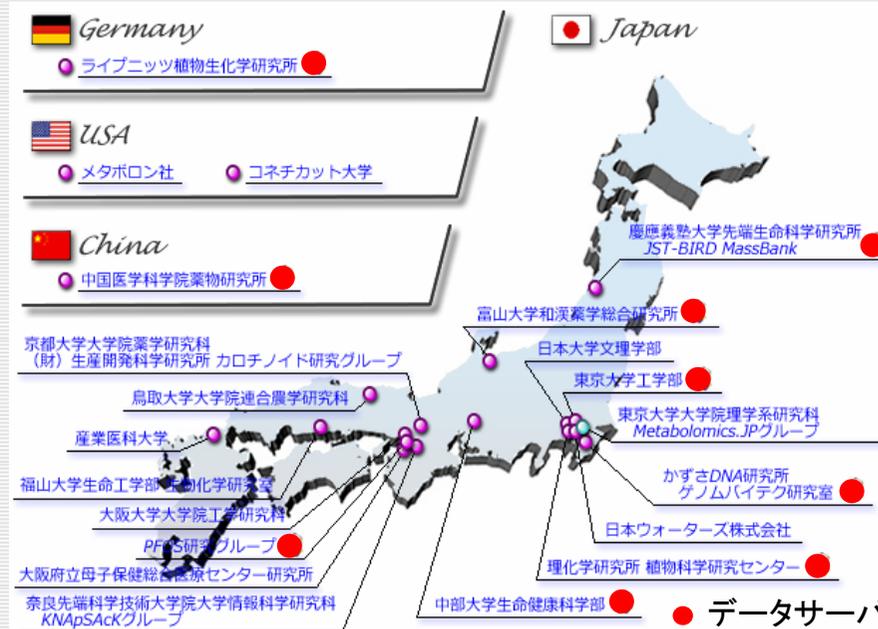
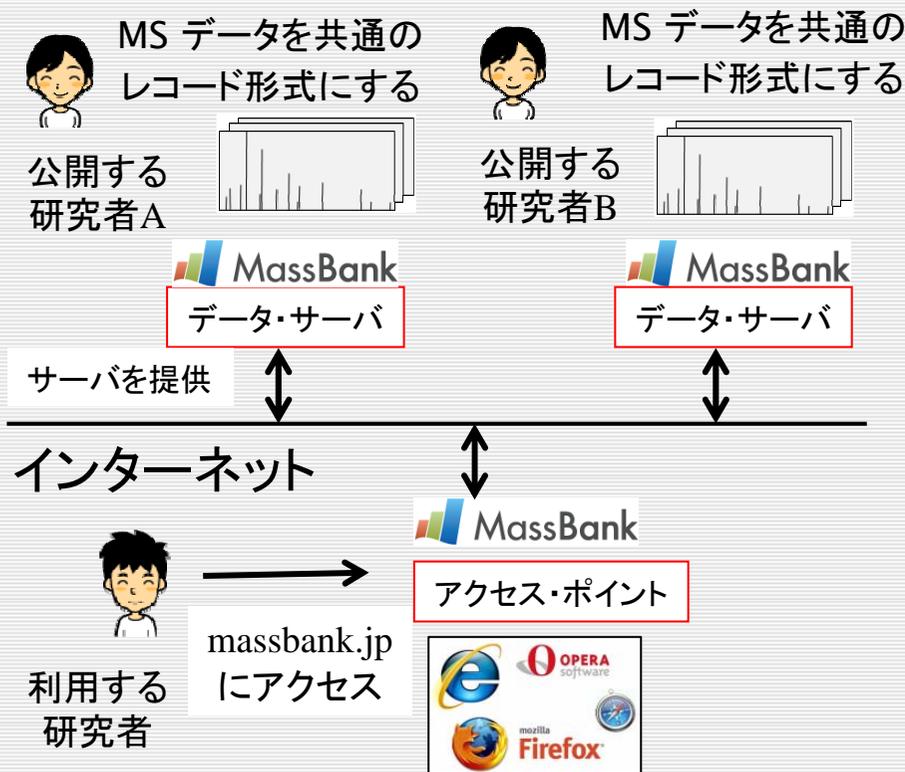
質量スペクトルから生理活性情報の統合化をめざす。



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] 質量スペクトルDB (MS DB) 分散型DB(MassBank)による世界貢献

## 化合物MS DB (BIRD II期)



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



○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)

○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

# [1-2] 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)



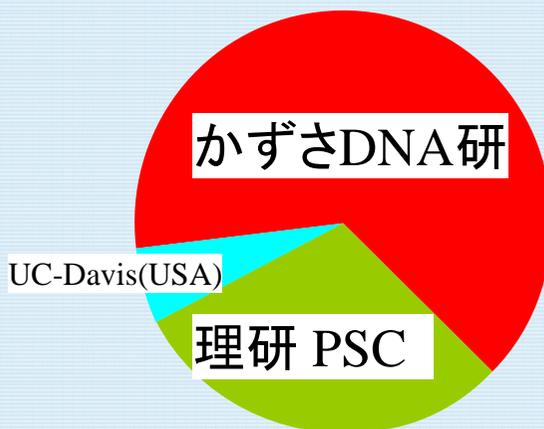
キャピラリー電気泳動  
質量分析装置  
(CE-MS)



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)



DB公開件数: 2.4万件  
(世界一の蓄積量)

KOMIC  
Market



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*, (2011)  
(accepted)  
AMDORAP: Non-targeted metabolic profiling  
based on high-resolution LC-MS

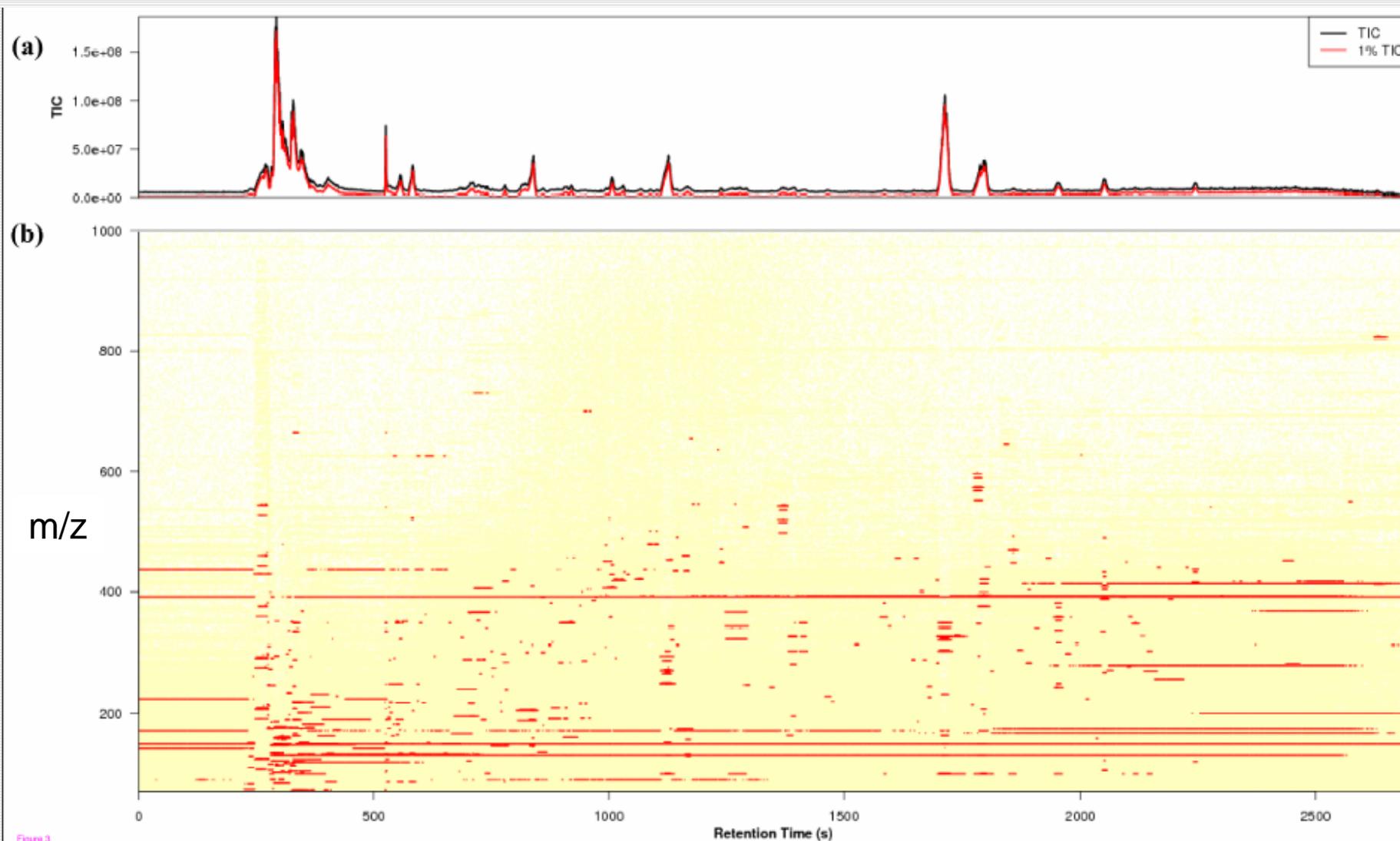


Figure 3

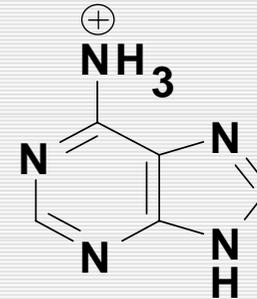
# [2-1] 代謝物情報DB: MS-化合物構造の関係知識DB

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

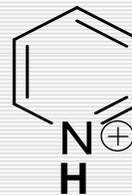


(BIRD II期)

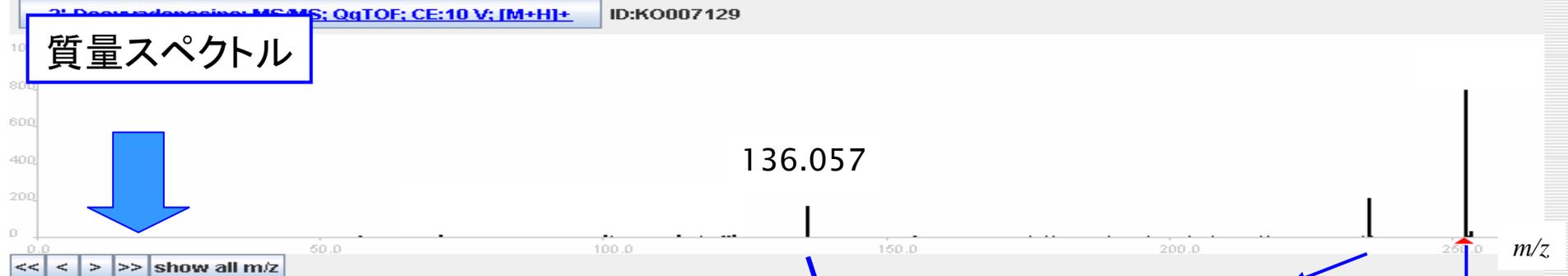
標品サンプル: 3000種  
報告代謝物数: 5-6万種

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

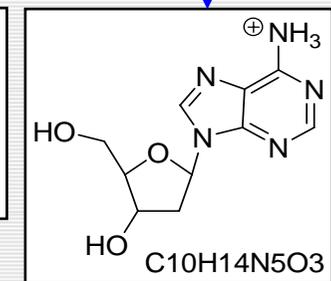
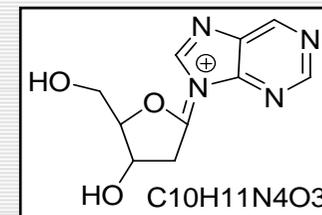
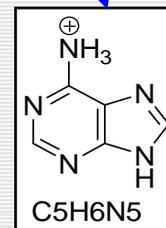
0.76  
0.74



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

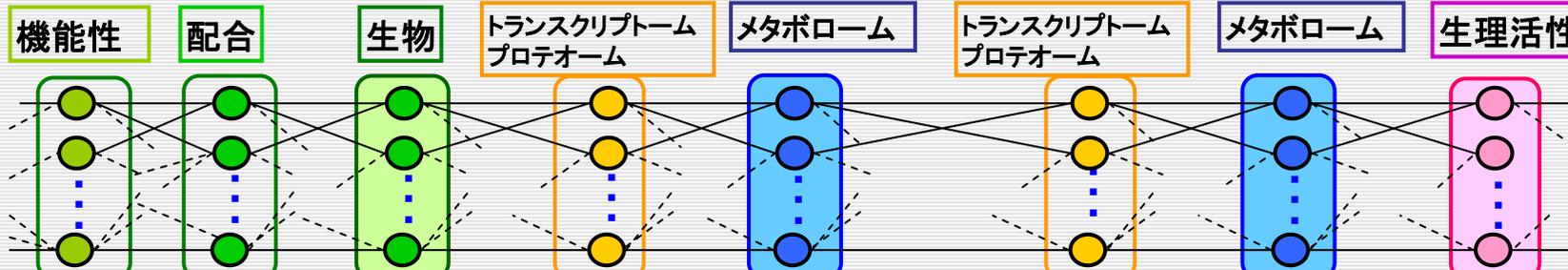
# [2-2] 代謝物情報DB: 代謝物-生理活性DB

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



配合例A

- クミン
- コリアンダー
- 唐辛子
- 胡椒
- ニンニク
- ショウガ
- ウコン
- サフラン
- パプリカ
- クローブ
- シナモン
- カルダモン
- ナツメグ
- オールスパイス
- キャラウェイ
- フェンネル
- フェヌグreek
- ジャガイモ
- タマネギ
- ニンジン

- クルクミン
- クルクモール
- アズレン
- カプサイシン
- アリシン

- TRPV1
- ビタミンB1
- アリチアミン

- エネルギー代謝促進
- 抗ガン作用・ガン抑制
- 解毒機能
- リラックス効果
- 解毒
- 解熱
- 強壮
- 駆風
- 高血圧予防・血圧降下
- 血行を高める
- 健胃
- 抗菌・殺菌
- 動脈硬化予防
- 滋養強壮
- 消化促進
- 食欲増進
- 新陳代謝亢進
- 整腸
- 鎮咳
- 疲労回復

代謝物-生理活性DB

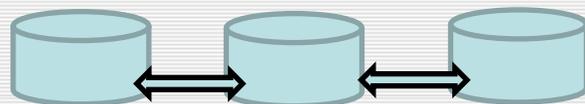
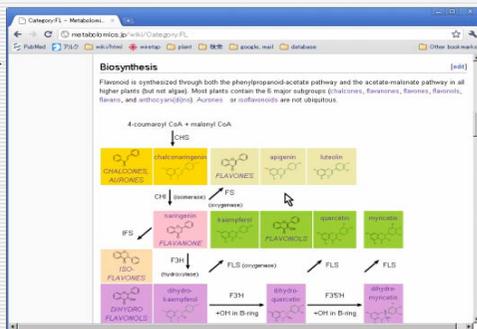
生物種-生理活性DB  
(38,555 レコード)

# [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



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

専門家による  
注釈、編集



DB型ウィキ(**BIRD II期**; [metabolomics.jp](http://metabolomics.jp))

「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、

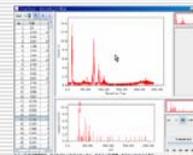
○持続可能なDB、

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

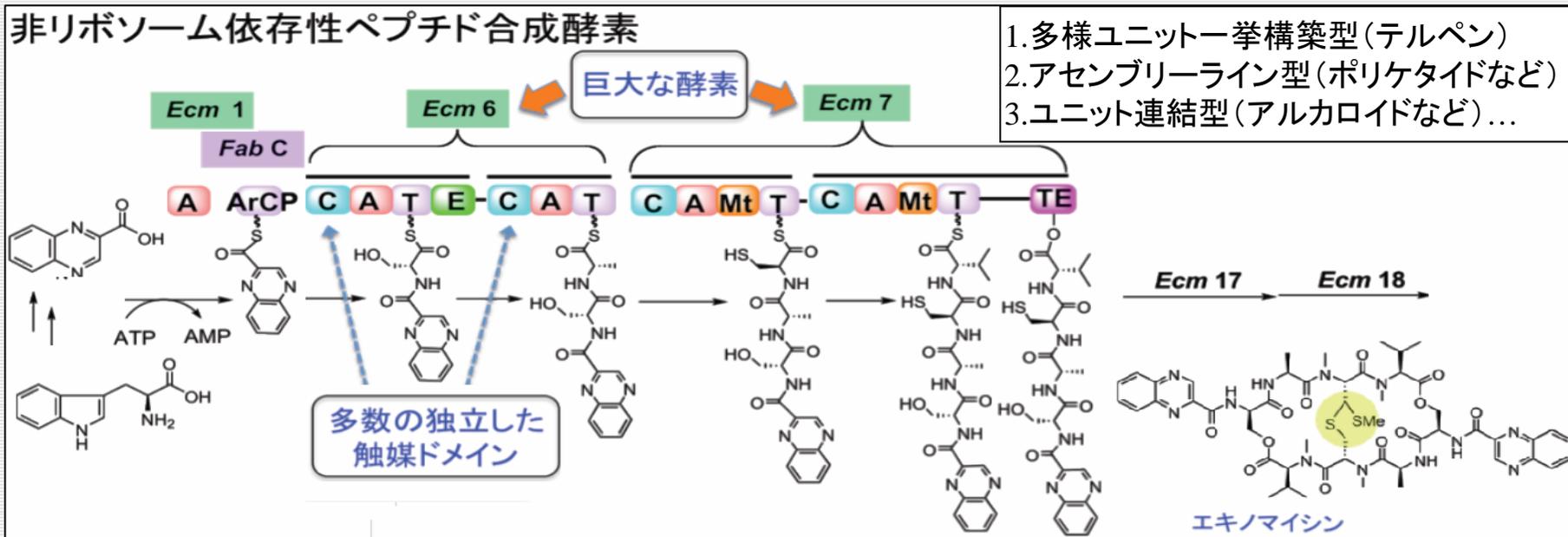
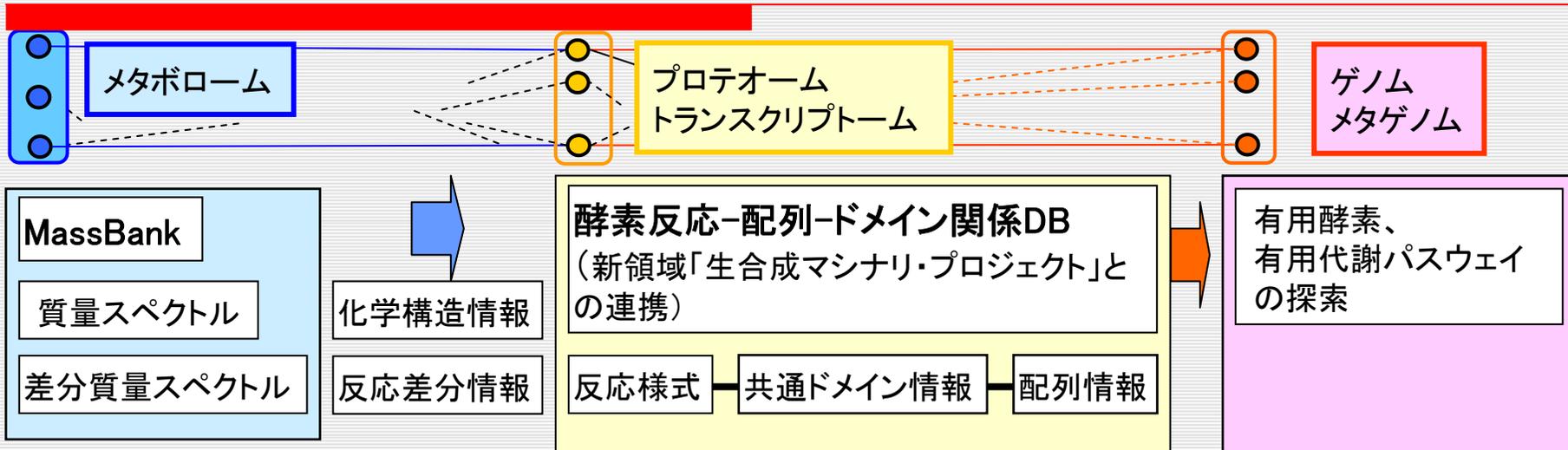


[3-2] メタボローム・アノ  
テーション・システム



# [3-3] ゲノム情報とのリンク

## 2次代謝物から遺伝情報とのリンク



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。

# [3-3] ゲノム情報とのリンク

## 2次代謝物から遺伝情報とのリンク(新領域学術研究バイオマシナリーと連携)



たすけて〜〜〜〜!

反応ID	酵素名	KEGG番号	EC番号	反応	生物種名	遺伝子名	アミノ酸配列	DDBJ/G enBank/ NCBI(Accession)	文献
[あけておく]	任意	任意	任意	必須	必須	任意	任意		必須
	diphosphate phosphohydrolase; pyrophosphate phosphohydrolase	R00004	3.6.1.1	*Diphosphate* + *H2O* <=> 2 *Orthophosphate*	<i>Bacillus subtilis</i>	hnpP	MSDKQVTTILFDLDGTLNLTNELDASF LHTLEHYYPKIKYREDVLAFIGP SLF DTFSSMDPKCEDMIAMYRAYNHQ MHD SLVTEYETVYETLDLALQAGFTL GVTTKLRDTVNM GUKLTGCEFFETV VTLDVYVNAKPDPEPVLLALKQLGS		Elberle, Microbial, 155, (2009), 1758
	名前が複数あるときには、半角「;」で区切る。			代謝物を半角の「*」で囲む  反応が可逆のとき半角で「<=>」、 不可逆のとき半角で「=>」 最低限: 基質と生成物があればよい。 (化学量論的にまっさり書く必要はない)					筆頭著名字, 雑誌名, 巻(年), 最初のページ 被験文献がある場合は半角「;」で区切る
	acetylacetyl-CoA synthase; acetyl-CoA:malonyl-CoA C-acetyltransferase (decarboxylating)		2.3.1.19 4	*acetyl-CoA* + *malonyl-CoA* => *acetylacetyl-CoA* + *CoA* + *CO2*	<i>Streptomyces</i> sp. CL190	nphT7	MTDVRFRJDTGAYVPERIVSNDEVG APAGVDDDWITRKTGRRRWAADD QATSD LATAAGRAALKAAGITPEQL TVJAVATSTPD RP QPP TAA'YVGHHL GATGTAAPDVNACSGTVFALSSVA GTLVYRGGYALVJGAD LYSRLN PAD RKTVYVLFQD GAGAMV LIGPTSTGTG PIVRRVALHTFGGLTDURVPA GGSR QPLD TDGLDAGLQYFAMD GREVRR	AB540131	Okamura, Proc. Natl. Acad. Sci. USA, 107 (2010) 11265. [PMID: 20534558]

# 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所 (Dr. Dirk Walther)

[2-1] MS-化合物構造の  
関係知識 DB (NAIST)

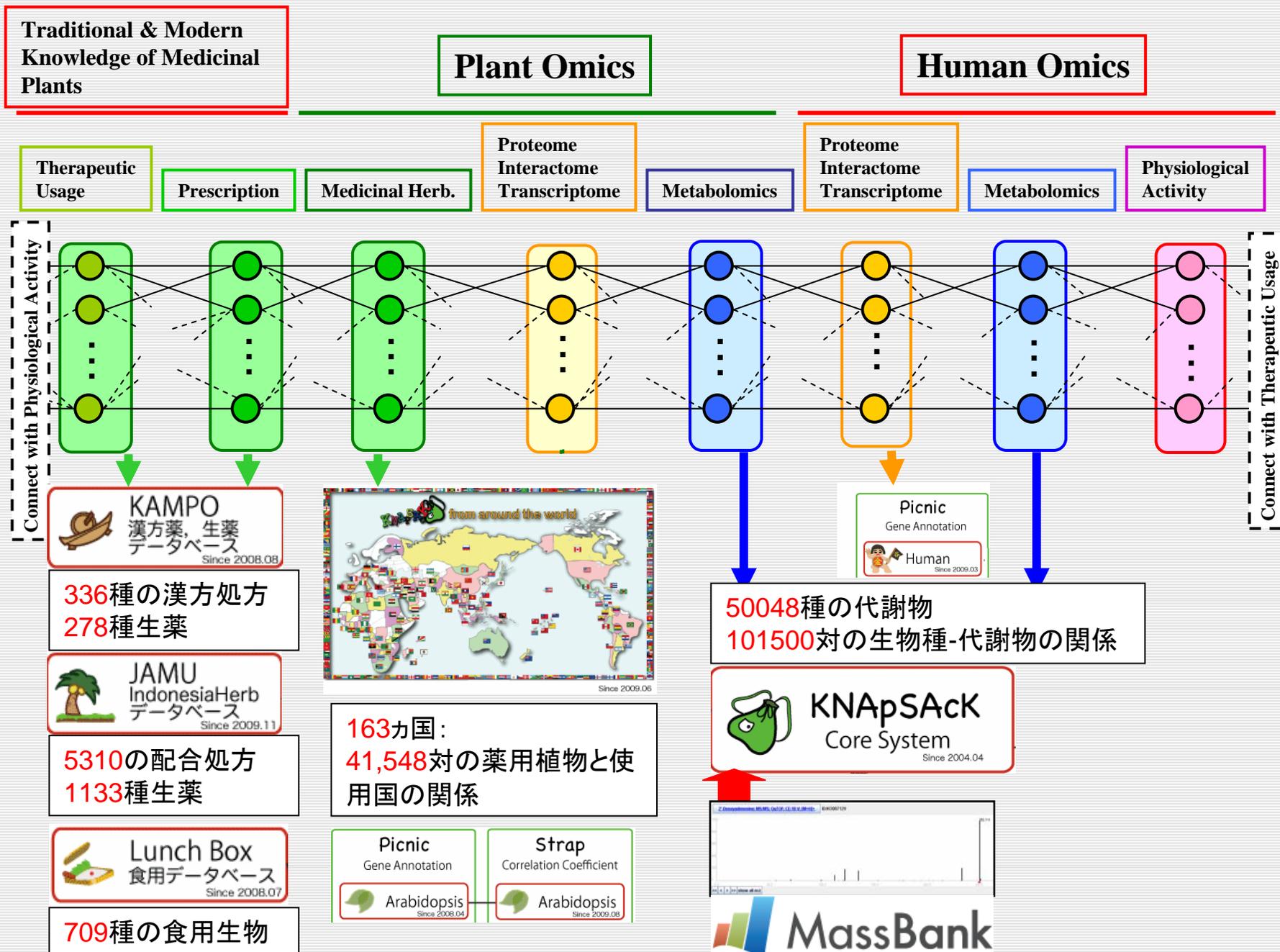
[2-2]代謝物-{生理活性, 生物  
種}関係DB (NAIST)

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)

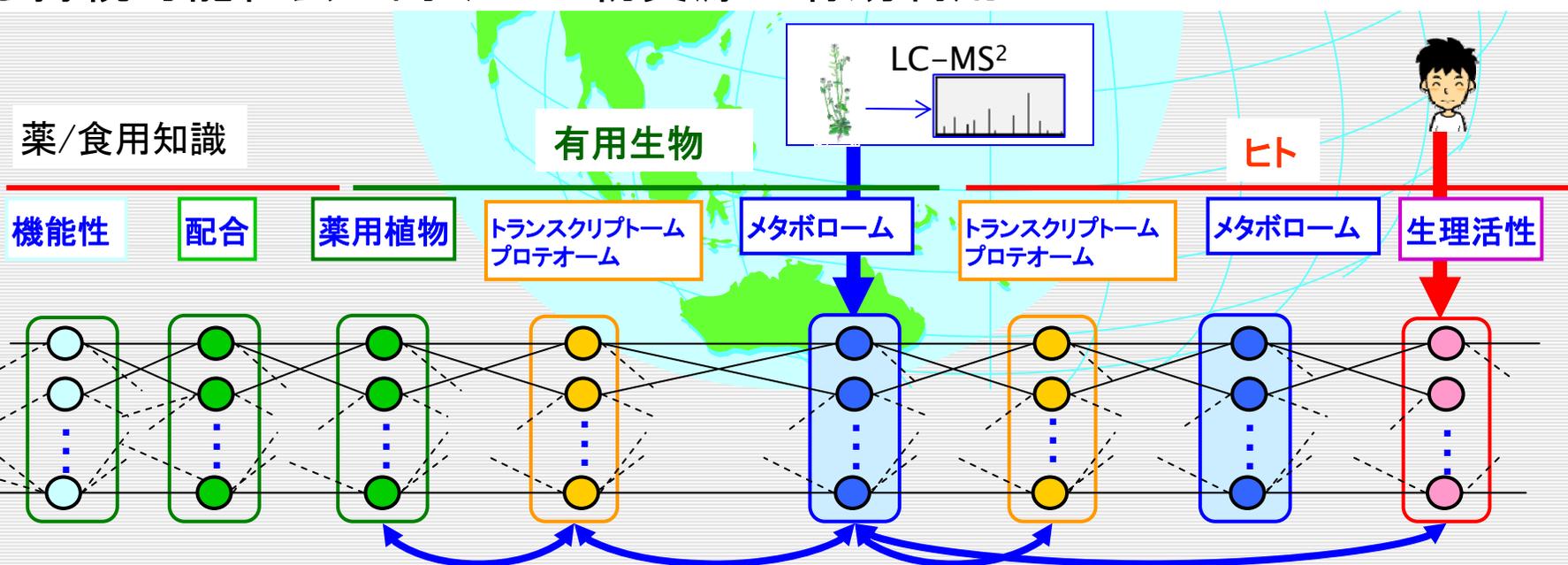
研究開発項目	H 2 3 年度	H 2 4 年度	H 2 5 年度
<b>1. 質量スペクトル DB (化合物 MSDB とメタボローム MSDB) の拡充</b> 1.1 化合物 MS DB の開発 (奈良先端大・理研・かずさ DNA 研グループ) ・プロトタイプシステムの開発 ・データの DB への蓄積ならびに公開 1.2 メタボローム MS DB の開発 (奈良先端大・理研・かずさ DNA 研グループ) ・DB プロトタイプシステムの開発 ・データの DB への蓄積ならびに公開		公開 ↓	
<b>2. 代謝物質情報DBの構築</b> 2.1 代謝物質と生物活性の関係データベース (奈良先端大グループ) ・データベース構築 ・データ蓄積ならびに公開 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 (理研・奈良先端大グループ) ・システム開発 ・データの充実ならびに公開		公開 ↓	
<b>3. メタボローム統合 DB</b> (理研グループ) ・wikiを中心としたデータ統合技術開発 ・wikiを中心とした統合データの公開			公開 ↓

# 伝統知識と現代的知識の融合: KNApSack Family



# 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)
- BIRD II期の成果の一部は、Nature誌ならびにその姉妹誌に3回紹介された。
- 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

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  
・比較ゲノム学講座

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23年6月10日

# おしまい

