

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

理化学研究所

環境資源科学研究センター

平成25年10月5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

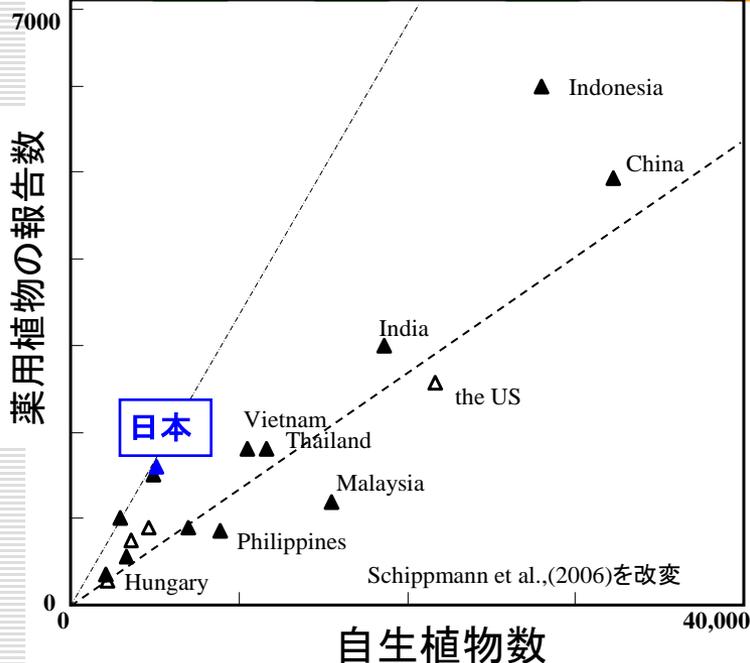
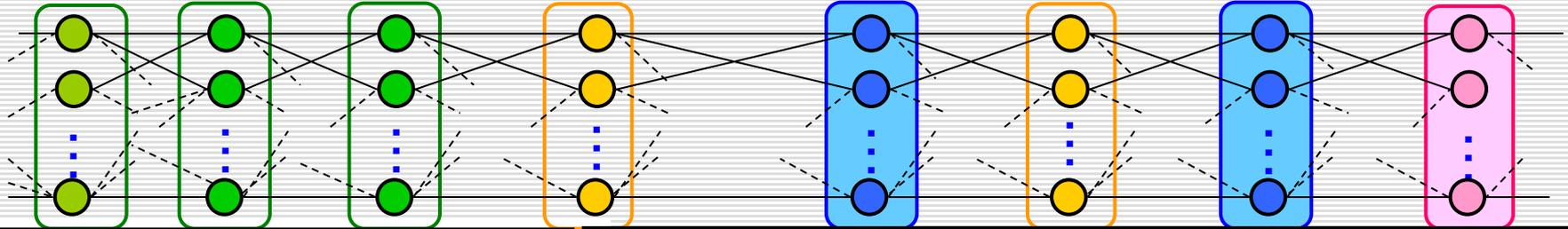
トランスクリプトーム
プロテオーム

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

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

(3)2次代謝DBとゲノム情報のリンク

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

[1-1] 化合物MS
DB

・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

[1-2] メタボローム
MS DB

[2] 代謝物情報DB

・文献情報
KNApSAcK DB

[2-1] 代謝物-
{生物種, 生理活
性}関係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

[2-2] 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

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

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

[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

メタボロームデータの統合管理

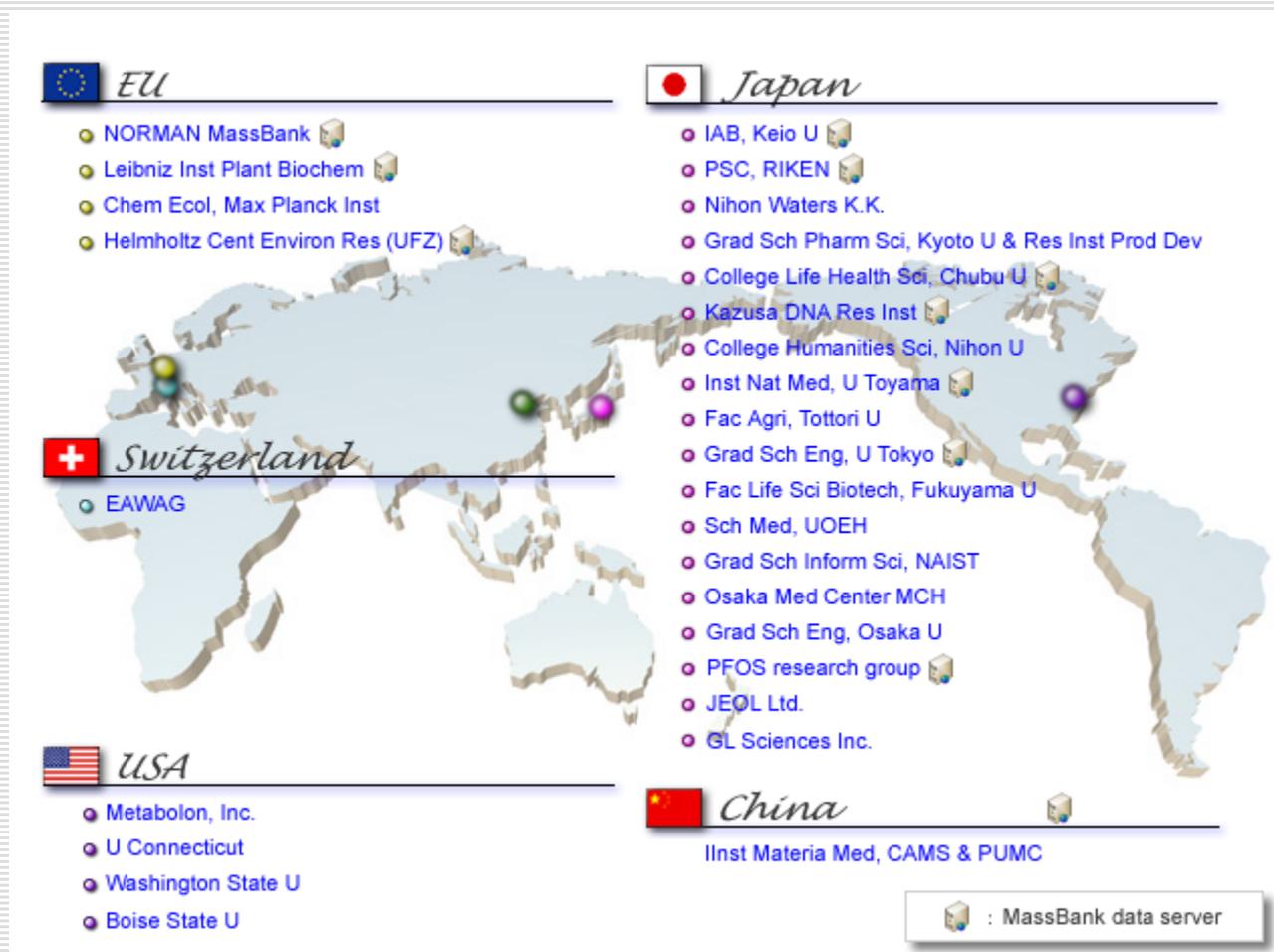
[3-2] メタボローム・アノテーション・システム

[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MS DB

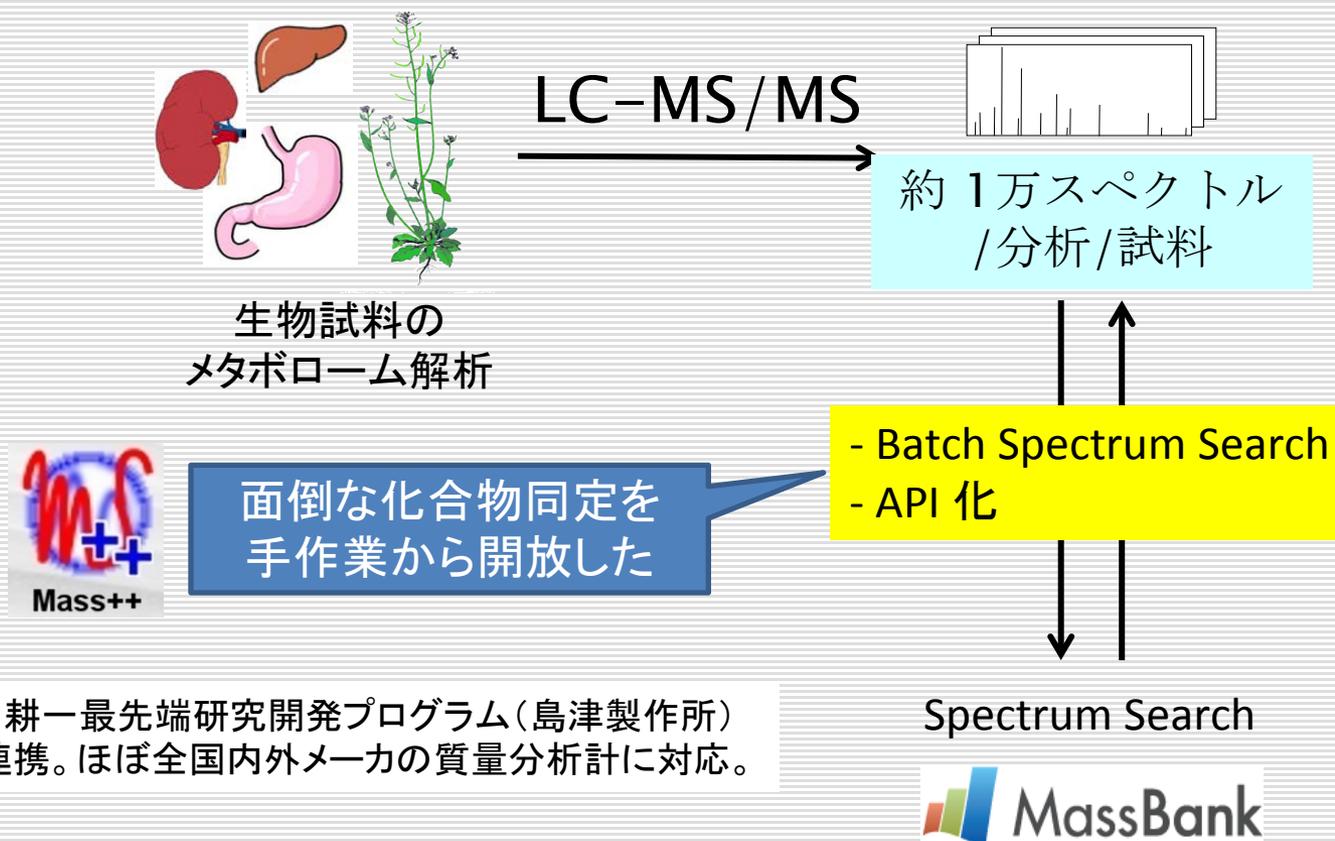
平成25年度 MassBank 開発内容: MassBankは世界標準



27 研究グループ(19 Japan, 4 EU, 2 USA, 1 China, and Switzerland) が9つのデータサーバから、15,483 化合物について分析した 39,467 の高品質なマススペクトルを提供している。訪問者は1,100 unique access / 日である (July 2013)。

[1-1]化合物MS DB

平成25年度 MassBank 開発内容:大量クエリの検索に対応



Third party によるツール開発が増加:
例: NORMAN グループの RMassBank

[1-1]化合物MS DB

平成25年度 MassBank 開発内容:分子式表現データの必要性

質量分析計はどんどん高精度化し続けている。

	質量分析計	m/z 測定値	測定誤差 (ppm)
1990 年代末	ESI-Q-Q-MS/MS	287.1	3,500
2000 年初頭	ESI-Q-TOF-MS/MS	287.06362	30
2012 年	ESI-Q-TOF-MS/MS	287.055317	1

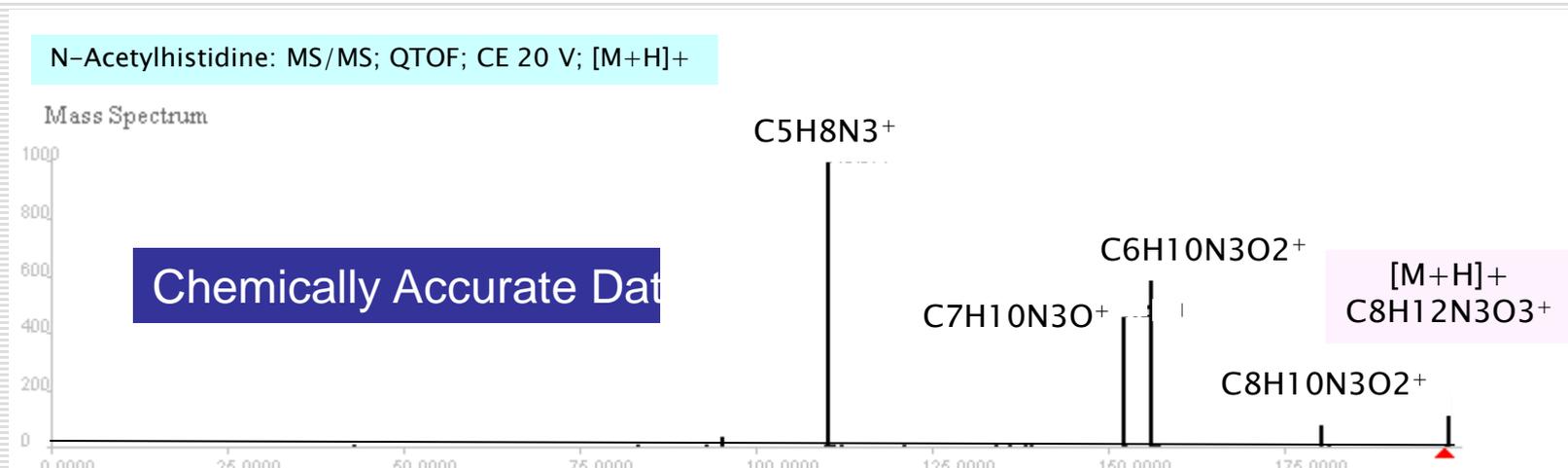
例: $[M+H]^+ = m/z$ 287.055017 (理論値) の場合

MassBank データの m/z 測定精度は 1 - 30 ppm である。

将来、MassBank データがメタボローム研究の超高精度化に役立たなくなる！

[1-1]化合物MS DB

平成25年度 MassBank 開発内容:分子式表現したデータの提供



Chemically Accurate Data を提供することによって
超高精度化と化学情報化を実現する

[1-1]化合物MS DB

平成25年度 MassBank 開発内容:分子式表現したデータの作成

MassBank に登録されている精度の高いデータ(2013年7月)

質量分析計 のタイプ	ESI- IT-FT	ESI- IT-TOF	ESI- Q-TOF	APCI- IT-FT	APCI- Q-TOF	MALDI- TOFTOF
データ数	8,573	253	5,246	1,199	633	45

合計 15,949 マスペクトル: 1,370化合物(正イオン)、1,110化合物(負イオン)



m/z 数値データ ---> 分子式データ
Chemically Accurate Data

Chemically Accurate Data の数

	分子式の総数	Unique な 分子式の数
分子式に置換えができた m/z (ピーク) 数	33,564	9,605
中性脱離分子の分子式 (= Molecular ion – Product ions)	31,063	11,243

[1-1]化合物MS DB

平成25年度 MassBank 開発内容: 分子式表現したデータ利用例



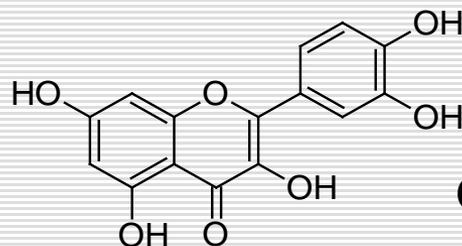
Query (m/z 数値データ)

検証

(1)
MassBank データ
Chemically Accurate Data を検索

(2)
 m/z 数値 データを検索

「(1)、(2) どちら化学構造式を識別する
能力が高いか？」 比較した。



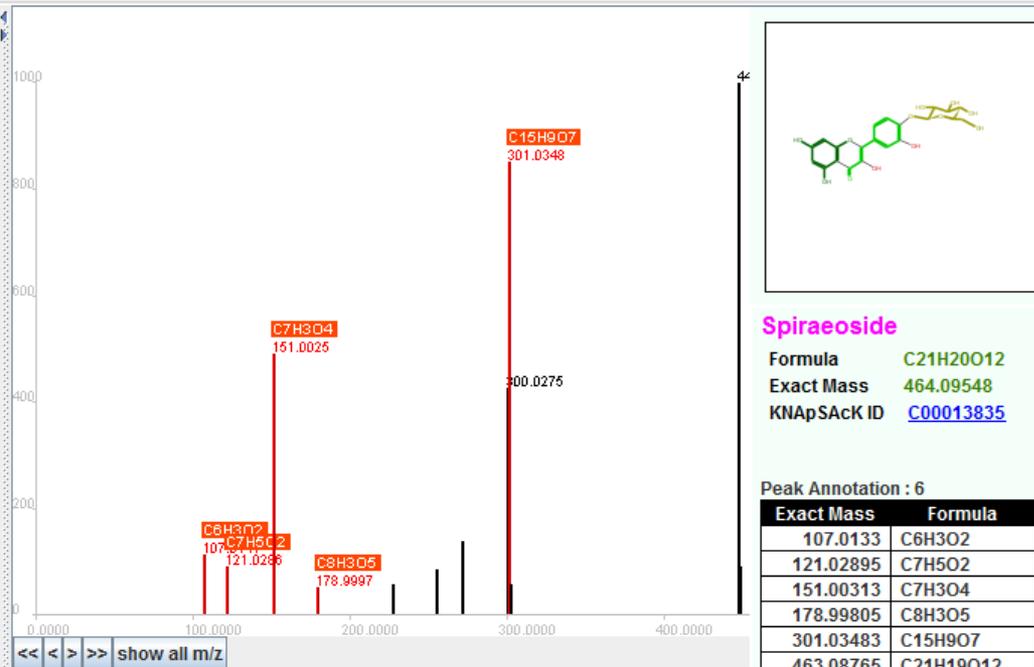
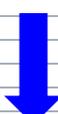
Quercetin 骨格

No.	Name	ID
24...	1750. Homoorientin [PR040130]	US002481
24...	1750. Homoorientin [PR040130]	US002482
24...	1750.	2483
24...	1752.	2484
24...	1752.	2485
24...	1752.	2486
24...	1752.	2487
24...	1752. Quercitrin [PR040196]	US002488
24...	1752. Quercitrin [PR040195]	US002489
24...	1752. Quercitrin [PR040194]	US002490
24...	1754. Quercetin-7-O-rhamnoside [P00993]	US002491
24...	1756. Cosmosiin [TY000187]	US002492
24...	1756. Cosmosiin [PR100649]	US002493
24...	1756. Cosmosiin [PR040066]	US002494
24...	1758. Callistephin [PR100672]	US002498
24...	1758. Callistephin [PR040160]	US002499
25...	1758. Callistephin [PR040159]	US002500
25...	1758. Callistephin [PR040158]	US002501
25...	1758. Callistephin [PR040157]	US002502
25...	1760. Sophoricoside [BML01310]	US002503

クエリのマススペクトル
(m/z 数値)



m/z 数値 --> 分子式変換

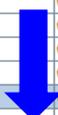


Spiraeoside
 Formula C21H20O12
 Exact Mass 464.09548
 KNApSAcK ID C00013835

Peak Annotation : 6

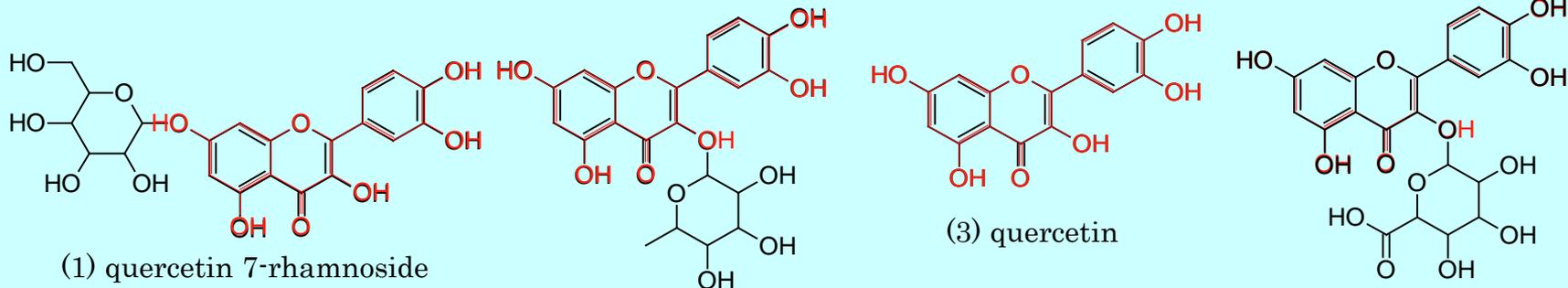
Exact Mass	Formula
107.0133	C6H3O2
121.02895	C7H5O2
151.00313	C7H3O4
178.99805	C8H3O5
301.03483	C15H9O7
463.08765	C21H19O12

MassBank
Chemically Accurate Data
を検索

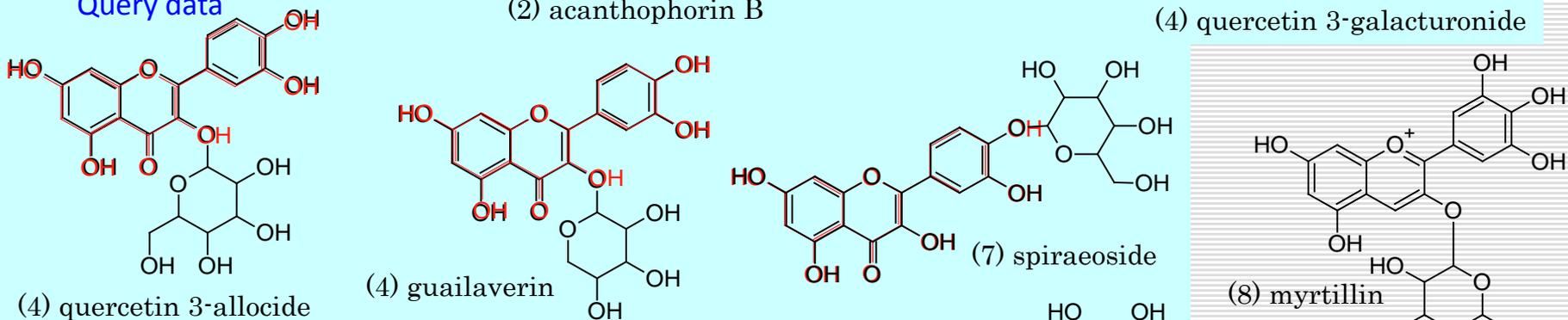


検索結果
ヒットした類似マススペクトル

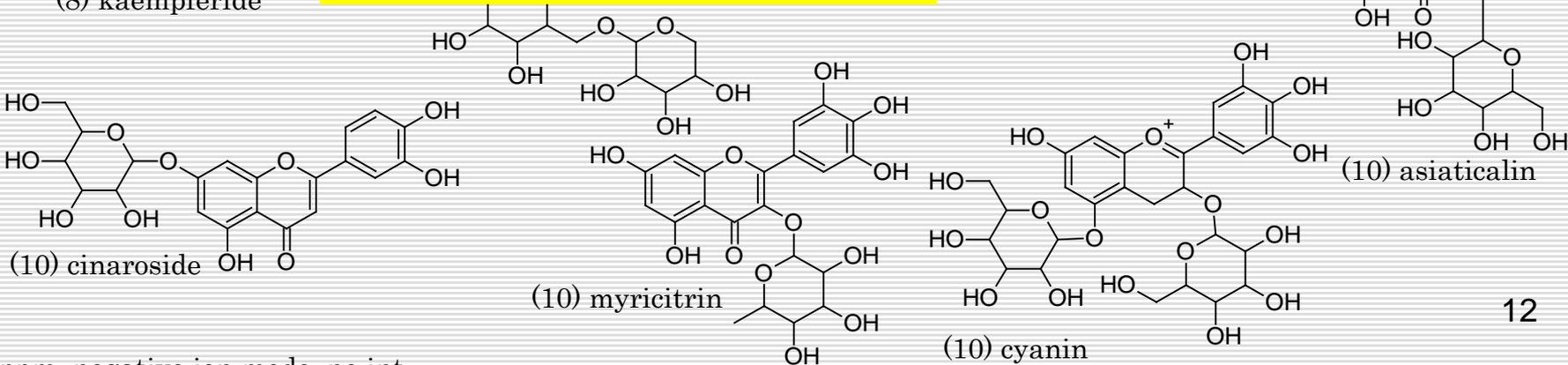
No.	Name	Formula	Exact Mass	Count	107.013551...	121.028051...	151.001951...	178.999151...	227.032351...	255.026951...	271.022751...	301.034251...	447.092151...
1	Quercetin	C15H10O7	302.04265	7	matching								
2	Quercetin 3-galacturonide	C21H18O13	478.07474	6	matching								
3	Quercetin 3-alloside	C21H20O12	464.09548	6	matching								
4	Guajaverin	C20H18O11	434.08491	6	matching								
5	Spiraeoside	C21H20O12	464.09548	5	matching								
6	Empetrin	C21H21O12	465.1033	4	matching								
7	Quercetin 3-O-galactoside	C21H20O13	480.09039	3	matching								
8	Cyanin	C27H31O16	611.16121	3	matching								
9	Petunidin 3-O-beta-D-galactopyranoside	C22H23O12	479.11895	3	matching								
10	Luteolin	C15H10O6	286.04774	3	matching								
11	Cyanidin 3-O-glucoside	C21H21O11	449.10839	3	matching								
12	Dihydroquercetin	C15H12O7	304.0583	3	matching								



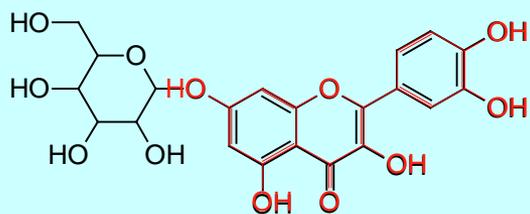
Query data



化学構造の識別能が高い。

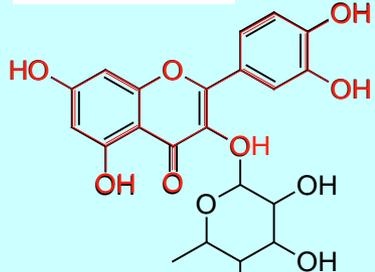


> Score 0.68

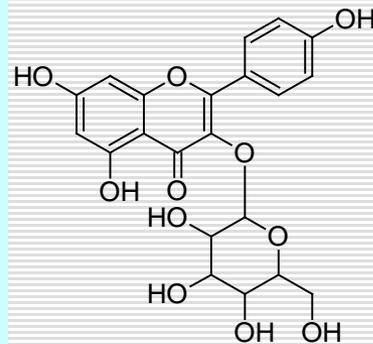


(1) quercetin 7-rhamnoside

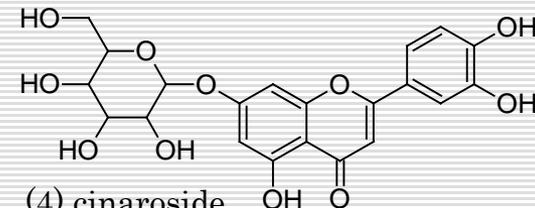
Query data



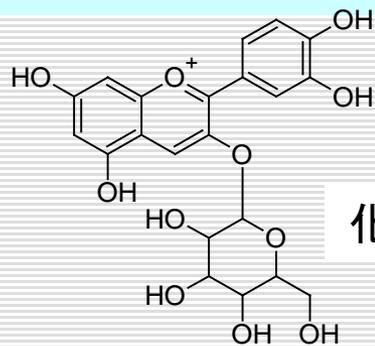
(2) quercitrin



(3) astragalin

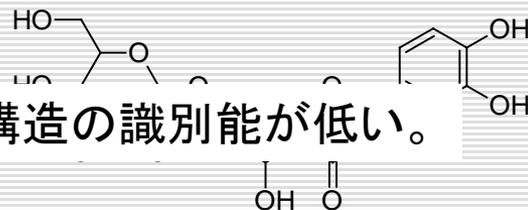


(4) cinaroside

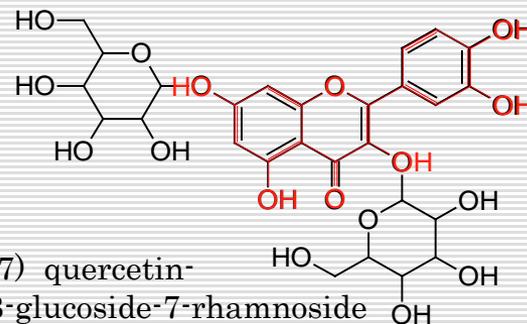


(5) cyanidin-3-galactoside

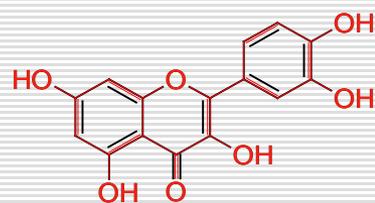
化学構造の識別能が低い。



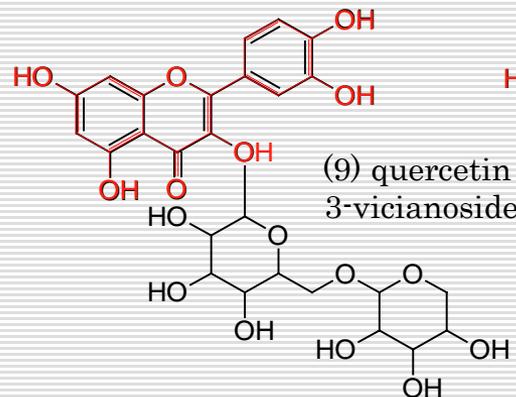
(6) luteolin-7-glucoside



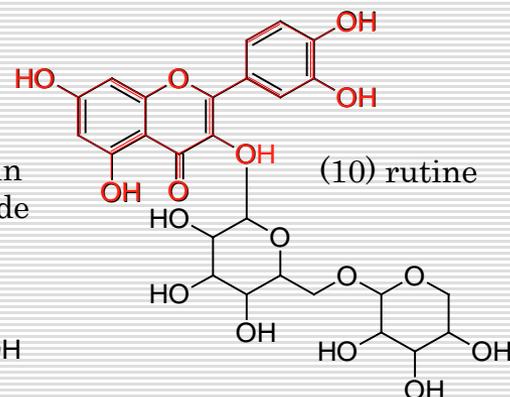
(7) quercetin-3-glucoside-7-rhamnoside



(8) quercetin



(9) quercetin 3-vicianoside



(10) rutin

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

[3]メタボローム統合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

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

メタボロー

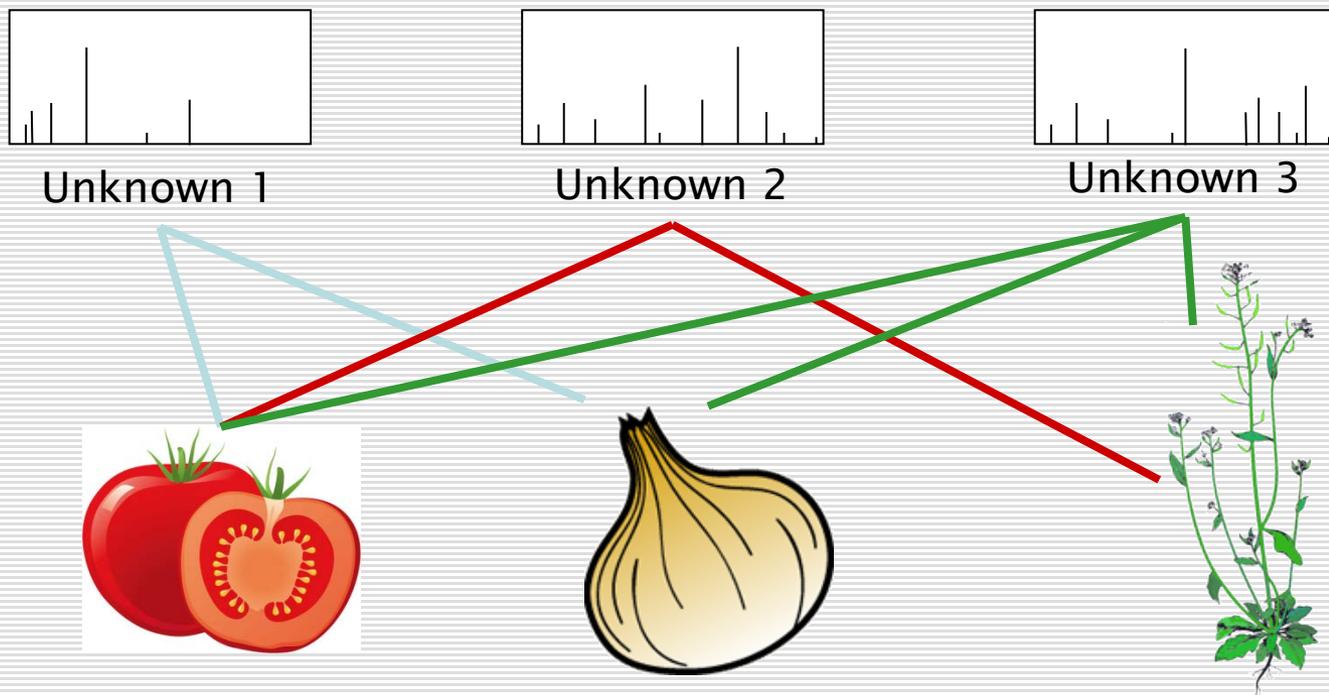
アノテーショ

バンク

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

[1-2]メタボロームMS DB

平成25年度メタボロームデータの整理と公開: Bio-MassBank



トマトには
unknown 1
unknown 2
unknown 3
がある。

タマネギには
unknown 1
unknown 3
がある。

シロイヌナズナには
unknown 2
unknown 3 がある。

マススペクトルの類似性を利用して、同じ unknown があるかどうか、
を調べることができる。

[1-2]メタボロームMS DB

平成25年度メタボロームデータの整理と公開: Bio-MassBank

平成25年度までに Bio-MassBank で公開したデータ

植物

生物種	組織など	研究グループ	データ数
シロイヌナズナ	葉	かずさ DNA 研	870
ミヤコグサ	花卉	かずさ DNA 研	908
キャベツ, 3栽培品種		かずさ DNA 研	1,702
トマト, 12栽培品種	果実	かずさ DNA 研	20,303
ハウレンソウ, 4栽培品種	葉	かずさ DNA 研	17,611
ナンヨウアブラギリ, 4栽培品種	果実	かずさ DNA 研	5,481
タマネギ	食用部	理研、植物科学センター	72
ディクソニア・アンタルクティカ、	葉	かずさ DNA 研	434
ヒメツリガネゴケ	葉	かずさ DNA 研	484
ゼニゴケ, 2種	葉	かずさ DNA 研	1,107

[1-2]メタボロームMS DB

平成25年度メタボロームデータの整理と公開: Bio-MassBank

微生物

生物種	組織など	研究グループ	データ数
シアノバクテリア PCC6803	細胞	かずさ DNA 研	124
クラミドモナス	細胞	かずさ DNA 研	146
マイタケ	子実体	かずさ DNA 研	1,424

哺乳動物

生物種	組織など	研究グループ	データ数
ハツカネズミ	3臓器、脂質	中部大学	2,250
ヒト (* 公開準備中)	尿	CEA - Centre d'Etude de Saclay, Gif-sur-Yvette, France	13,306

総計 66,272 データ

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

[1-1] 化合物MS
DB

・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

[1-2] **メタボローム
MS DB**

[2] 代謝物情報DB

・文献情報
KNAPSAcK DB

[2-1] 代謝物-
{生物種, 生理活
性}関係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

[2-2] MS-化合物構
造の関係知識DB

[3]メタボローム統合DB

ウィキDBによるメタボロー
ームの統合管理

[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

櫻井 望
(財)かずさDNA研究所・
産業基盤開発研究部

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

[1-2]メタボロームMS DBの構築 目標

大量データの公開と、公開を加速する技術開発

[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



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)

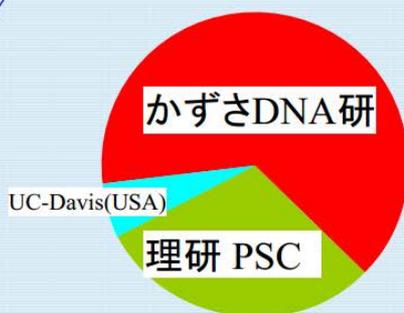


液体クロマトグラフィー
質量分析装置
(LC-MS)



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)

メタボロームデータの定式化
(代謝物ピークの抽出)



DB公開件数:2.4万件
(世界一の蓄積量)

代謝物の同定
アノテーション



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。

[1-2]メタボロームMS DB データ公開の概要



Metabolonote

メタデータ専用のデータベース

28生物種、375分析を764の解析データに関するメタデータを新規に公開。
7つの他データベースと連携



MassBase

生データのリポジトリ

132生物種、38,344分析を新規に公開



KomicMarket

検出ピーク情報のデータベース

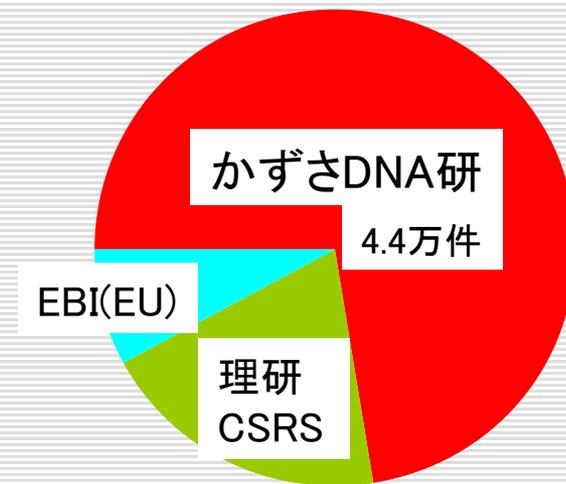
17生物種、200分析分を新規に公開



Bio-MassBank

生物由来のMSスペクトルデータベース

16実験セット、50,644ピークのスペクトルを新規に公開



メタボロームデータの公開数

[1-2]メタボロームMS DB (参考)その他の成果

メタボローム情報の普及

- ・ポータルサイトKOMICSの構築
- ・学会発表・出展による広報

分子生物学会、農芸化学会、細胞工学会、
質量分析学会、植物細胞分子生物学会、
植物生理学会、食品科学工学会

データ解析ツールの開発・公開

PowerGet (LC-MS用)、FragmentAlign
(GC-MS用)等のバージョンアップ

論文発表

MFSearcher: 質量分析データから、迅速
な組成式演算・データベース検索を行う
システム

Sakurai et al. (2013) Bioinformatics 29 (2): 290–291



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

[1-1] 化合物MS
DB

・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

[1-2] メタボローム
MS DB

[2] 代謝物情報DB

・文献情報
KNAPSAcK DB

[2-1] 代謝物-
[生物種, 生理活
性)関係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

[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

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
・計算システムズ生物学講座

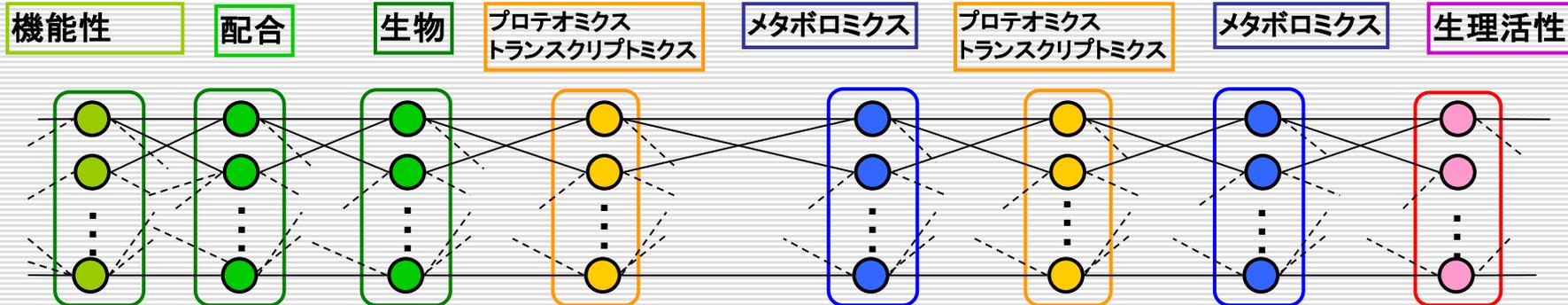
基礎研究: 持続可能社会に向けた生物
産業応用: 新規有用代謝物質の探索、ゲノム育種による有用/新規代謝物質の生産 など

1.1 生物間・ゲノムバイオロジー研究・プラットフォーム

薬用・食用
知識ベース

有用生物

動物

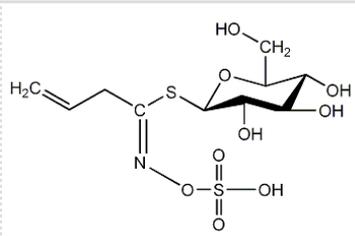


モンシロチョウ
VS
アブラナ科植物

アブラナ科植物



sinigrin



摂食刺激
(モンシロチョウ)

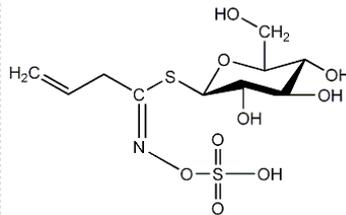


ヒト
VS
アブラナ科植物

アブラナ科植物



sinigrin



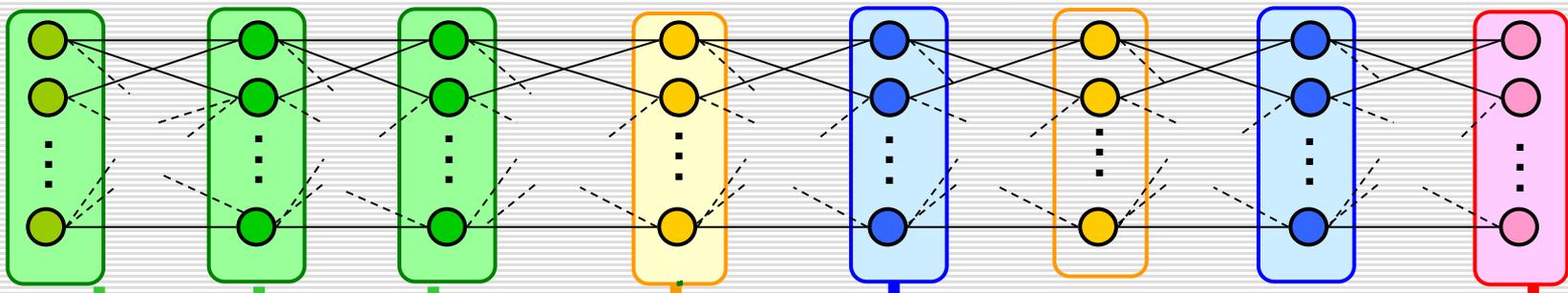
消化促進作用,
利尿作用。
ガン細胞のアポトー
シス作用?
(ヒト)

伝統知識と現代的知識の融合: KNApSAcK Family データベース

薬用・食用 知識ベース

植物/微生物Omicis

ヒトOmicis



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

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

309種加工食品
261種食材

DietDish
食べ合わせデータベース
Since 2012.11

3708種の組み合わせ
414種食材

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

2,500 生物種-代謝反応の
関係

217カ国
48,256対の薬用植物と使用国の関係

Lunch Box
709種の食用生物

DietNavi
病気予防データベース

602種の食用植物
403種の代謝物

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

KNApSAcK
Core System
Since 2004.04

50048種の代謝物
101500対の生物種-代謝物の関
係

MassBank



統合DBプロジェクト

Biological Activity
Metabolite Activity
Since 2012.02

9584対の代謝物-生物活性の関係

Biological Activity
Natural Activity
Since 2011.08

33703 対の生物種-活性の関係
1949 種の活性種
1533 種の生物種

Metabolite Activity DB (二次)代謝物-生物活性關係DB

9,584 代謝物-活性關係

2,356 metabolites 140 activity categories

2,963 biological activities 778 target species.

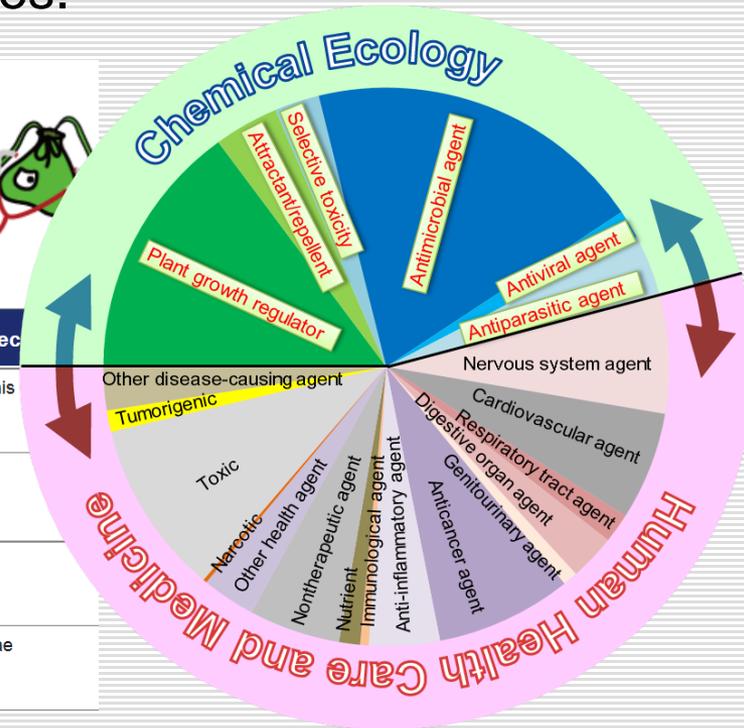


Metabolite Activity

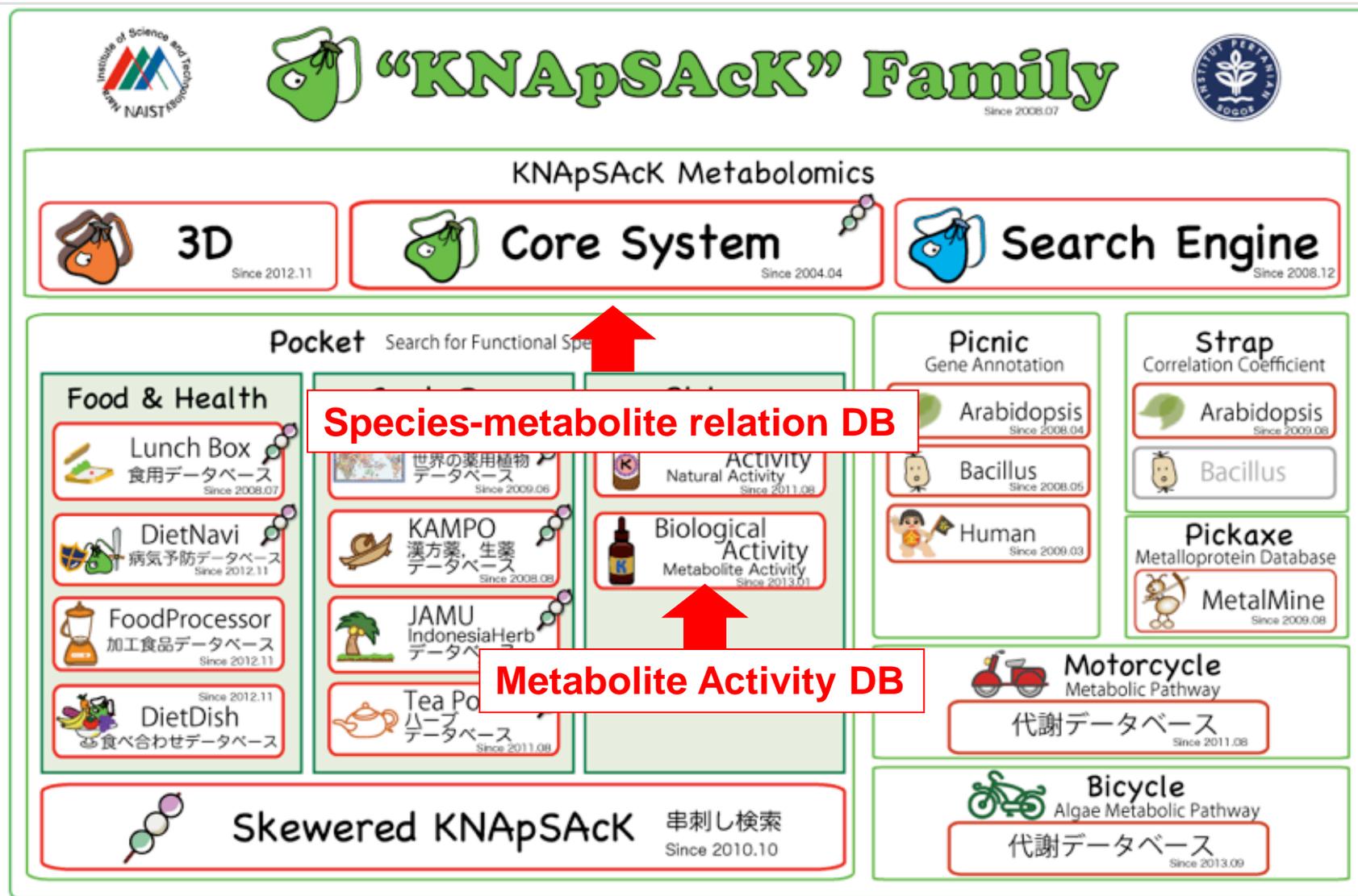


INPUT WORD = [Match Type : Partial , Metabolite : glucosinolate]

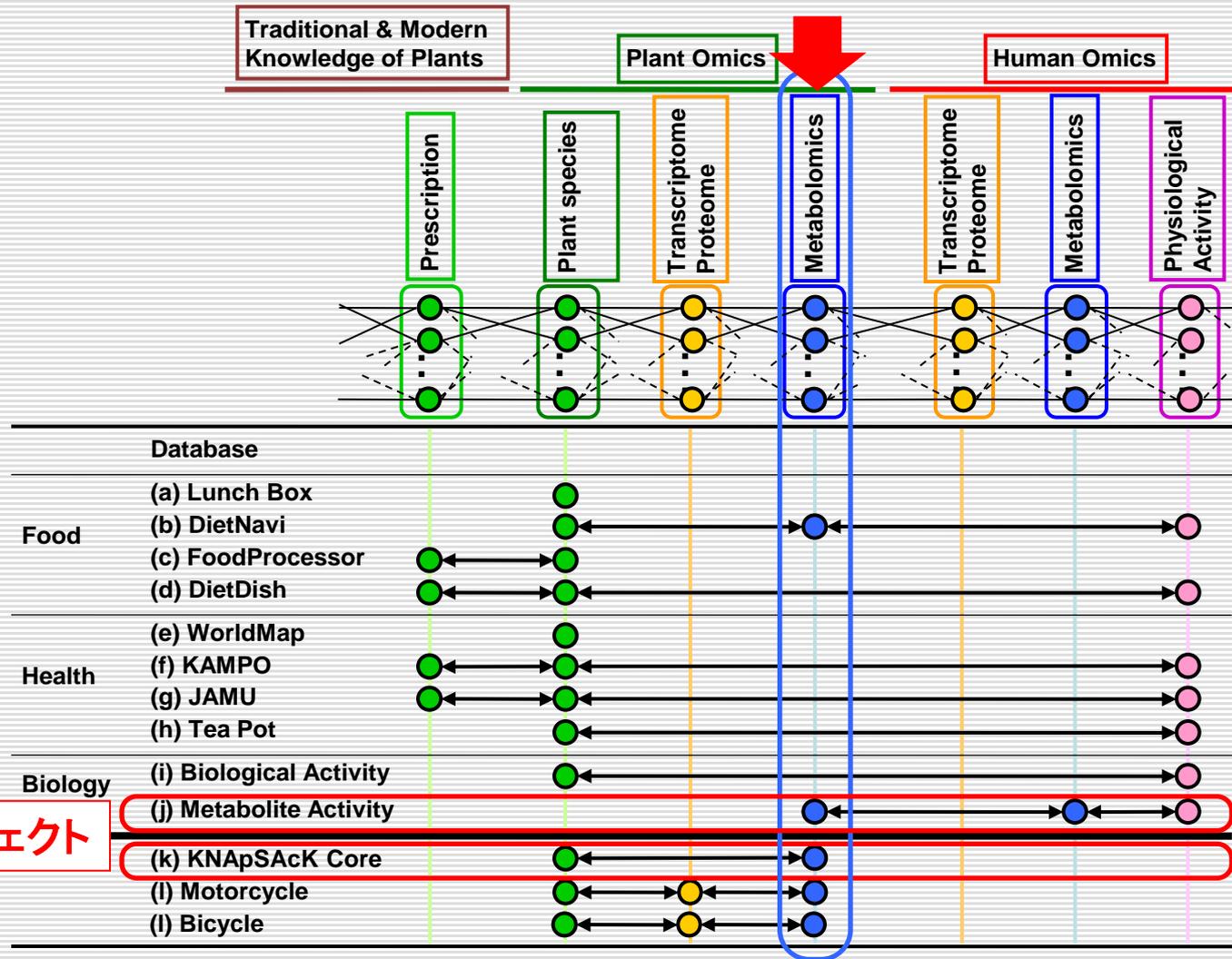
C_ID	Metabolite Name	Activity Category	Biological Activity (Function)	Target Species
C00001486	Progoitrin 2(R)-2-Hydroxy 3-butenyl glucosinolate	Feeding attractant	stimulate feeding	Plutella maculipennis
C00001488	Sinigrin 2-Phenylethyl glucosinolate Allyl glucosinolate	Attractant	used as attractants in field traps	Delia brassicae
C00001488	Sinigrin 2-Phenylethyl glucosinolate Allyl glucosinolate	Oviposition attractant	attract and stimulate egg laying	Pieris brassicae
C00001488	Sinigrin 2-Phenylethyl glucosinolate Allyl glucosinolate	Oviposition attractant	attract and stimulate egg laying	Hymlemya brassicae



Main window of KNApSAcK family: http://kanaya.naist.jp/KNApSAcK_Family/



Current status of KNAPsAck family



Publication List of KNApSAcK Family DB

KNApSAcK Metabolite Activity DB

Nakamura Y., Afendi MA, Parvin AK, Ono N, Tanaka K, Morita HA, Sato T, Sugiura T, Amin M, Kanaya S, KNApSAcK Metabolite Activity Database for Retrieving the Relationships between Metabolites and Biological Activities

Plant Cell Physiol. (2013)(to be accepted)

KNApSAcK Core (Species-metabolite DB)

Nakamura K, Shimura N, Otabe Y, Morita HA, Nakamura Y, Ono N, Amin M, Kanaya S, KNApSAcK-3D: A Three-Dimensional Structure Database of Plant Metabolites, *Plant Cell Physiol.* **54**: e4.1-8(2013)

Afendi FM, Okada T, Yamazaki Morita HA, Nakamura Y, Nakamura K, Ikeda S, Takahashi H, Amin M, Darusman LK, Saito K, Kanaya S

KNApSAcK Family Databases: Integrated Metabolite-Plant Species Databases for Multifaceted Plant Research

Plant Cell Physiol. **53**: e1.1-12(2012)

その他: KNApSAcK Family関係

Ikeda S, Abe S, Nakamura Y, Kibinge N, Morita HA, Nakatani A, Ono N, Ikemura T, Nakamura K, Amin M, Kanaya S, Systematization of the Protein Sequence Diversity in Enzymes Related to Secondary Metabolic Pathways in Plants, in the Context of Big Data Biology Inspired by the KNApSAcK Motorcycle Database, *Plant Cell Physiol.* **54**, 711-727 (2013)

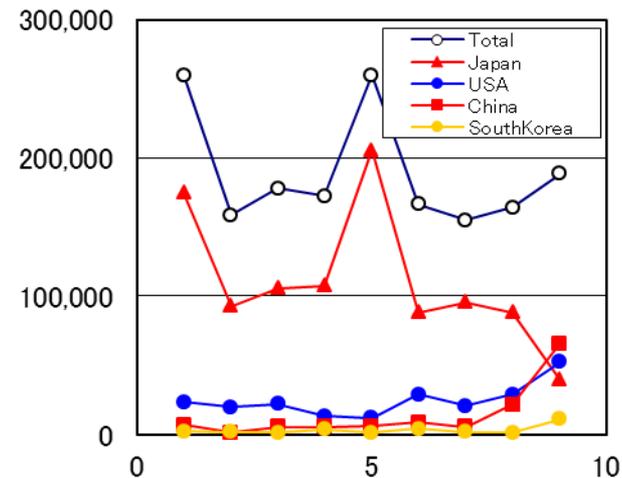
池田 俊、桂樹 哲雄、小野 直亮、中谷 淳至、中村由紀子、森田 晶、金谷 重彦、オミックス・プラットフォーム：バイオ・ビッグ・データに挑む、*生物工学* **90**:777-781(2013)

Afendi FM, Ono N, Nakamura Y, Nakamura K, Darusman LK, Kibinge N, Morita HA, Tanaka K, Horai H, Amin M, Kanaya S

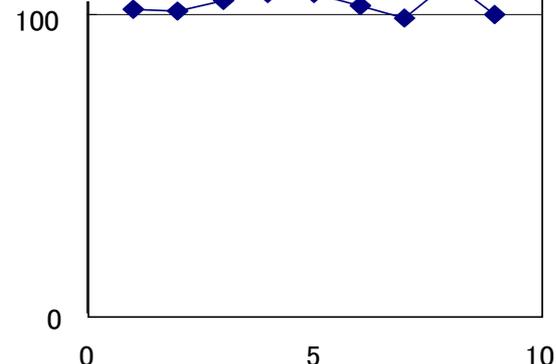
Data Mining Methods for Omics and Knowledge of Crude Medicinal Plants toward Big Data Biology *Comput. Struct. Biotech. J.*, 4: e201301010 1-14 (2013)

1-9月のアクセスログ

件数



地域数



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

[1-1] 化合物MS
DB

・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

[1-2] メタボローム
MS DB

[2] 代謝物情報DB

・文献情報
KNApSAcK DB

[2-1] 代謝物-
{生物種, 生理活
性}関係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

[2-2] MS-化合物構
造の関係知識DB

[3]メタボローム統合DB

有田 正規
理化学研究所
環境資源科学研究センター

[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

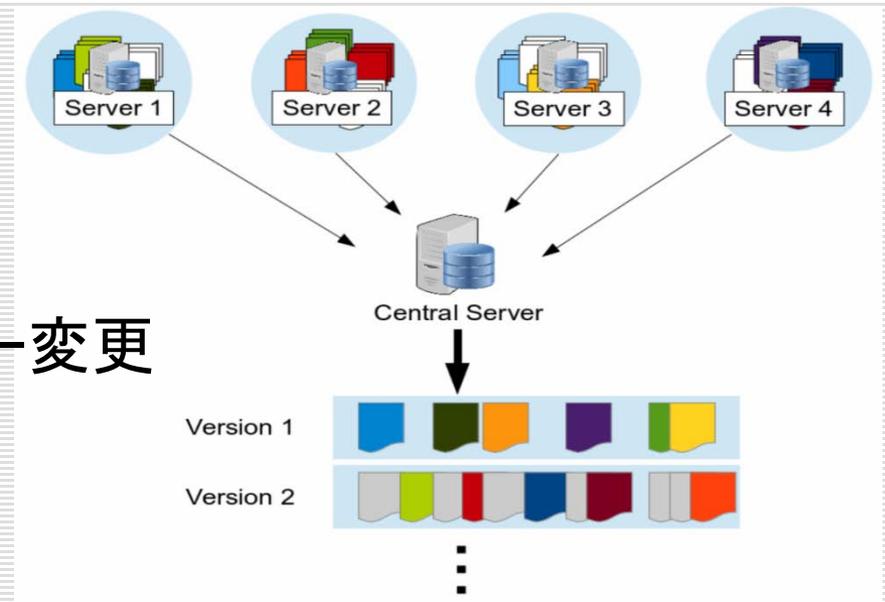
[3-3] ゲノム情報とのリンク

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

ネットワークによるデータ共有

- コミュニティによる共有: データ提供を義務付けるのではなく、クローラーが各サイトを訪れてデータを収集
- 自動化: スペクトルを自動的にコンパイルしてサマリー作成
- 再分散: ネットワークを通して、サマリー (compiled spectra) を配布

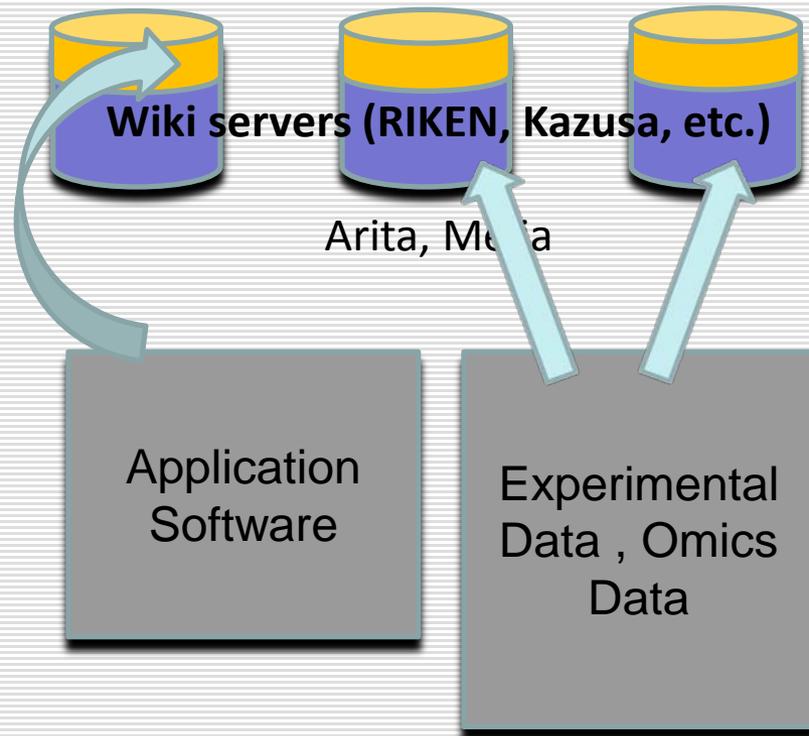
ネットワークへの参加は自由
サマリーを作成すれば、メンバー変更
による影響を最小化できる



システム概要

Networking Layer (Akie Mejia)

1. ネットワークレイヤはRESTプロトコルに基づき、JSONファイルを収集。



2. 研究グループはwikiその他のサーバを用意
3. MassBank_{light}などのプログラムは各サーバーから利用可能
4. データの検索は、サマリーファイルを用いて各サーバーで処理

コンセプトの変化

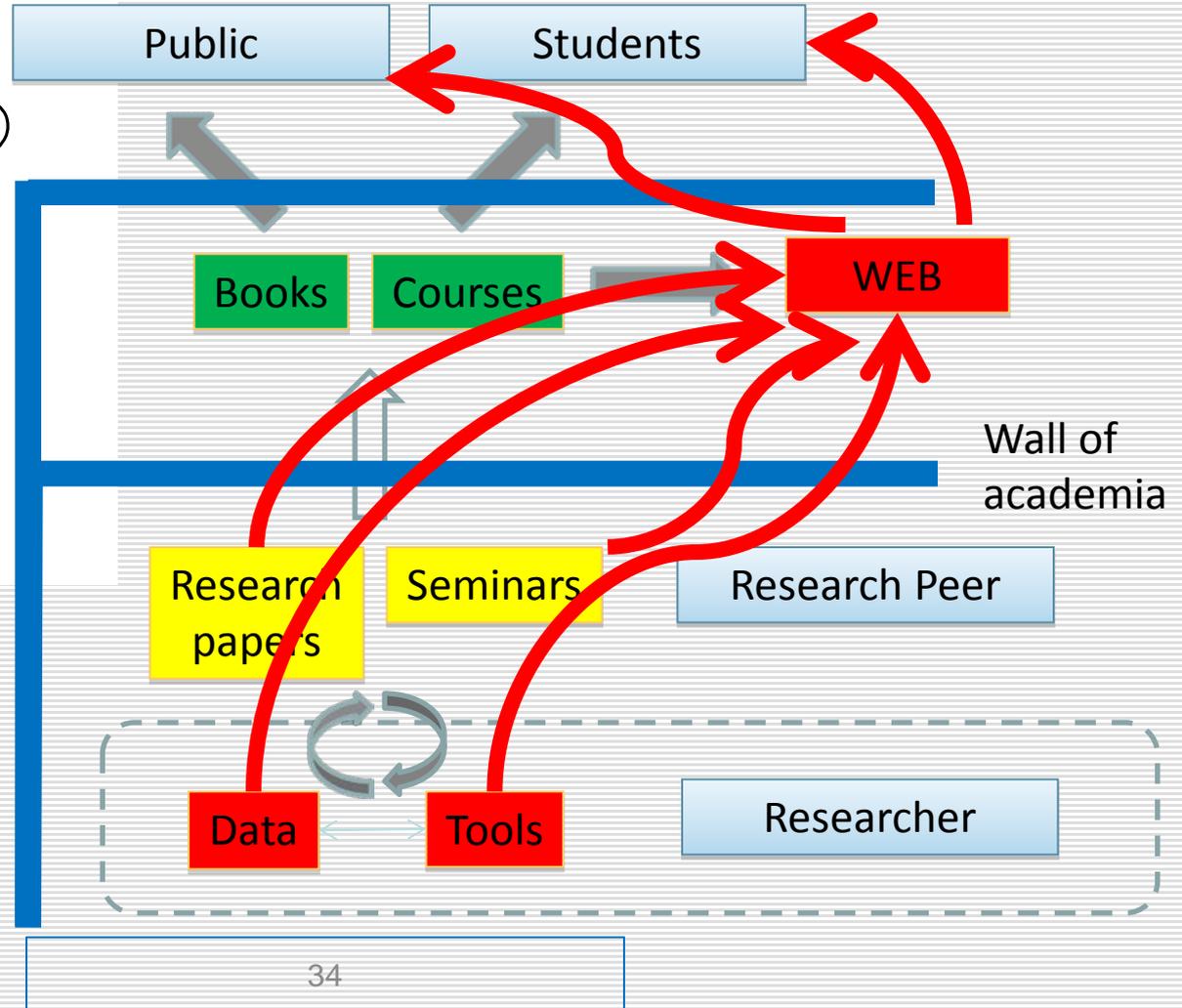
ウェブページによるデータ公開

ウェブのみだと
インセンティブ
が無い(大変なだけ)

研究の範疇を

e-Scienceに拡大

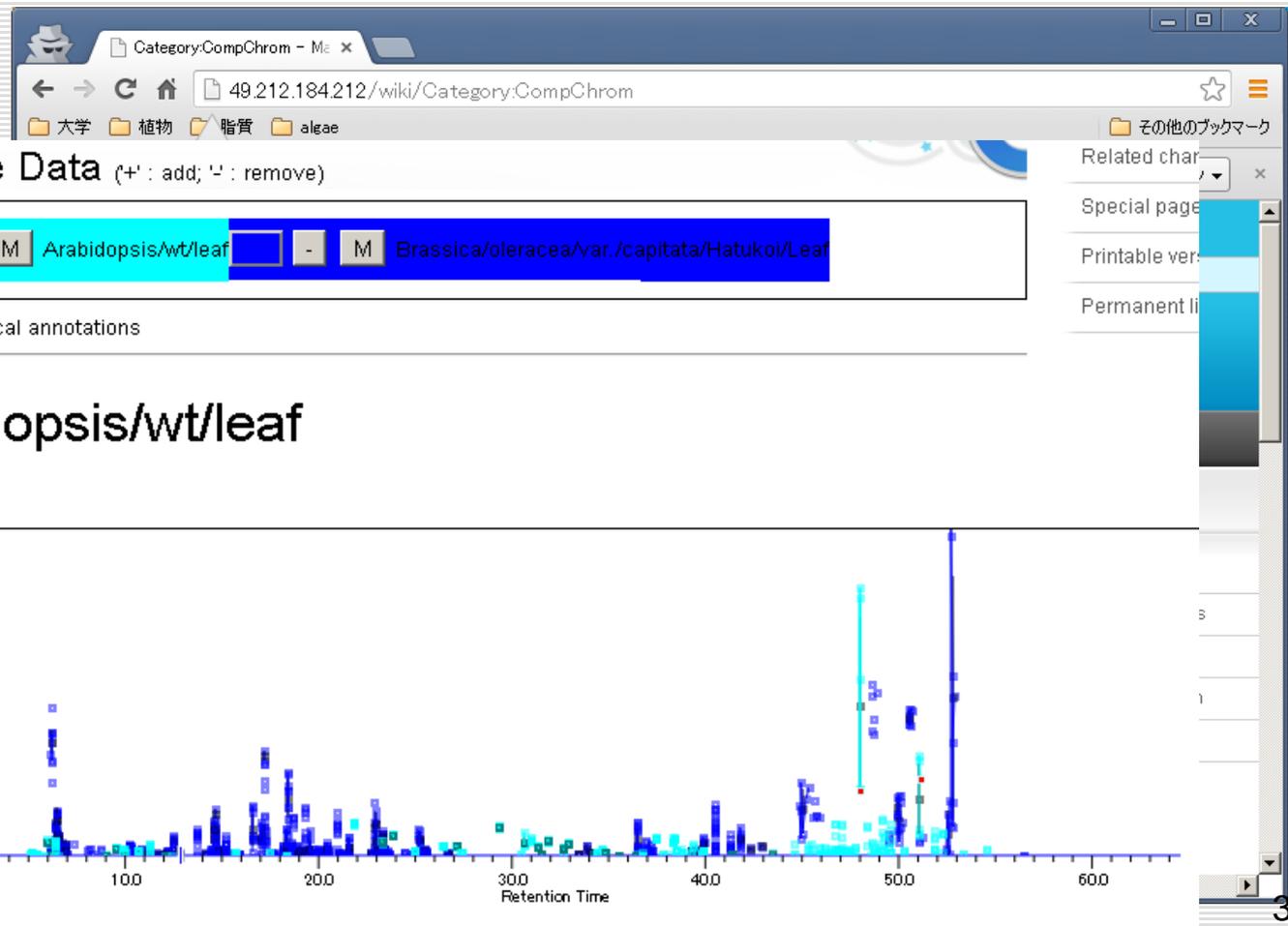
論文や授業の代わりに
そのコンテンツを利用



Server 例: Spectral data

REFSPEC server as the next version of MassBank

<http://49.212.184.212/wiki/Category:CompChrom>



データを“シェア”するためのコツ

1. コミュニティで管理

- 一人の“administrator”を作らないようにする
- GRIDによる管理 (e.g. PDB, GenBank)

2. 自動化

- よく整理したデータを最初に作成し、後は自動化
- 管理コストを下げることが重要

3. 分散化

- 追加データは各自が管理するようにする

メタボロームデータベースの統合

:第4の科学に向けて

サンプル

検証実験

機器分析

生データ



ピーク検出
(半定量)



PowerGet

FragmentAlign



ピークデータ

MassBank



Bio-MassBank

データマイニング

RefSpec



化合物
アノテーション
(定性)

ReSpect

MS2T

構造推定システム



生物情報



ゲノム情報



代謝パスウェイ

メタボロームデータ



化合物情報
生理活性
文献情報



転写トーム

プロテオーム

疾患

作業仮説構築