

生物種メタボロームモデル・ データベースの構築

有田正規 (理研CSRS)

金谷重彦 (NAIST)

櫻井望 (かずさDNA)

前期までの主要結果

- Bio-MassBankの公開(奈良)

| 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 | 花弁 | 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 |

前期までの主要結果

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

前期までの主要結果

- Metabolonote等の公開(かずさ)

SE1:/S01/M01/D01

Sample Set Information

| | |
|-------------|--|
| ID | SE1 |
| Title | Arabidopsis thaliana leaf metabolite analysis |
| Description | Investigation of Arabidopsis thaliana leaf metabolites. 6 replicated data are examined for each sample. |
| Authors | Takashi Arai, Ryohei Sakai, Hitaru Enomoto, Ritsuro Sakurai, Hideyuki Suzuki, Daisuke Shibata, Kazusa DNA Research Institute |
| Reference | Direct Submission |
| Comment | version 5 |

Sample Information

| | |
|-----------------|---|
| ID | SE1 |
| Title | Arabidopsis wt leaf |
| Organization | Arabidopsis thaliana |
| Organization ID | KOJI2 taxonomy: F702 |
| Compound ID | |
| Compound Name | |
| Preparation | Arabidopsis thaliana Col-0 seeds are sown on pots filled by vermiculite and are grown in an incubator with 16h light/8h dark and 22 degree C condition. After 2 months later, all leaves are harvested. |
| Comment | [KomicMarket ID] <KSA_13> |

Analytical Method Information

| | |
|---------------|----------------------------|
| ID | M01 |
| Method and ID | M01 |
| Sample Amount | 0.7 mg |
| Comment | [MassBase ID] M01_CL_05711 |

Analytical Method Details Information

| | |
|--------------------|--|
| ID | M01 |
| Title | LC-FT-SEC-MS ESI positive method 1 |
| Instrument | Agilent1100 HPLC (Agilent), LTQ-FT (Thermo Fisher Scientific) |
| Instrument type | LC-FT/SEC-MS |
| Ionization | ESI |
| Ion Mode | Positive |
| Description | Harvested sample is frozen by liquid N2 and resulting powder (100mg) are solved in 300uL 80% acetonitrile and is injected into HPLC after 0.2um membrane filter treatment. HPLC conditions: Agilent 1100 (Column: TQ8000 (4.6 x 250 mm, 5 micrometer) TQ8000), Solvent: A: 0.2% formic acid in water, B: 100% acetonitrile; Gradient: (0-10 to 25.0 min), 20 to 90% (25.0 to 40.0 min), 90% (40.0 to 50.0 min), 2% (50.0 to 57.0 min), Column temp: 30 degree C, Flow rate: 0.3mL/min, P0 (60s), P1 (200 ms condition), Filter: 1, FTMS is a cation topology by m/z: 20000 to 20000 (1500.0), 2: normal scan rate (Dep MS/MS) Most intense ion from (1), rejected (mass: 256.2000, 266.0000, 284.2000, 300.2000, 326.2000, 344.2000, 372.2000, 388.2000, 416.3000) |
| Comment of details | |

Data Analysis Information

| | |
|-----------------------------------|--------|
| ID | D01 |
| Data Analysis Set ID | D01 |
| Recommended default places of m/z | ITMS 2 |
| Comment | |

Bio-MassBank The mass spectrum data are available at Bio-MassBank.

NEW KomicMarket TEMPORARY The peak data files written in TugMD format are downloaded by the new KomicMarket temporary web site.

Data Analysis Details Information

| | |
|----|-----|
| ID | D01 |
|----|-----|

Bio-MassBank

Mass Record: KZSAT0000000024
 Arabidopsis wt leaf / SE1_S01_M01_D01_P0016; LC-ESI-ITFT; MS2; [M+H]⁺; RT:9.576756 min

Mass Spectrum

ACCESSION: KZSAT0000000024
 RECORD TITLE: Arabidopsis wt leaf / SE1_S01_M01_D01_P0016; LC-ESI-ITFT; MS2; [M+H]⁺; RT:9.576756 min
 DATE: 2019.05.15 (Created: 2011.12.09)
 AUTHORS: Takashi Arai, Ryohei Sakai, Hitaru Enomoto, Ritsuro Sakurai, Hideyuki Suzuki, Daisuke Shibata, Kazusa DNA Research Institute
 LICENSE: CC BY-NC-SA
 PUBLICATION: Direct Submission
 COMMENT: [Metabolonote ID] SE1_S01_M01_D01_P0016
 COMMENT: [MassBase ID] M01_CL_05711
 COMMENT: [MS] KZSAT0000000024
 COMMENT: MS_TYPE details is ITMS.

OSNAME: Glucosaminolactone
 COMPOUND_CLASS: Natural Product
 CHEMFORMULA: C15H21NO10S
 CHEXACT_MASS: 437.0841
 CHEMFILES: N/A
 CHEMPACT: N/A
 CHEMLINK: KEGG 000419
 CHEMLINK: PubChem 120007545 00007545

SPECIES: Arabidopsis thaliana
 SPECIES: NCBI: TAXONOMY: 3109
 SPESAMPLE: Arabidopsis wt leaf

ACQUISUMENT: Agilent1100 HPLC (Agilent), LTQ-FT (Thermo Fisher Scientific)
 ANALYTICMETH: FTMS, LC-ESI-ITFT

KnapSack
Core System
Since 2004.04

KFCG
Kazusa Functional Compound
Genome and Database

前期までの主要結果

- KNApSack Family DB の拡充(奈良)
 - 9,584 代謝物-活性関係
 - 生物活性のオントロジー
 - 食品中の代謝物

“KNApSack” Family Since 2008.07

KNApSack Metabolomics

- 3D** Since 2012.11
- Core System** Since 2008.04
- Search Engine** Since 2008.12

Pocket Search for Functional Species

| | | |
|---|---|--|
| Food & Health <ul style="list-style-type: none"> 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 食食べ合わせデータベース Since 2012.11 | Crude Drug <ul style="list-style-type: none"> WorldMap 世界の薬用植物データベース Since 2009.04 KAMPO 漢方薬データベース Since 2009.04 JAMU IndonesiaHerb データベース Since 2009.11 Tea Pot 茶データベース Since 2011.09 | Biology <ul style="list-style-type: none"> Biological Activity Natural Activity Since 2011.08 Biological Activity Metabolite Activity Since 2013.01 |
|---|---|--|

Picnic Gene Annotation

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

Strap Correlation Coefficient

- Arabidopsis Since 2009.09
- Bacillus

Pickaxe Metalloprotein Database

- MetalMine Since 2009.08

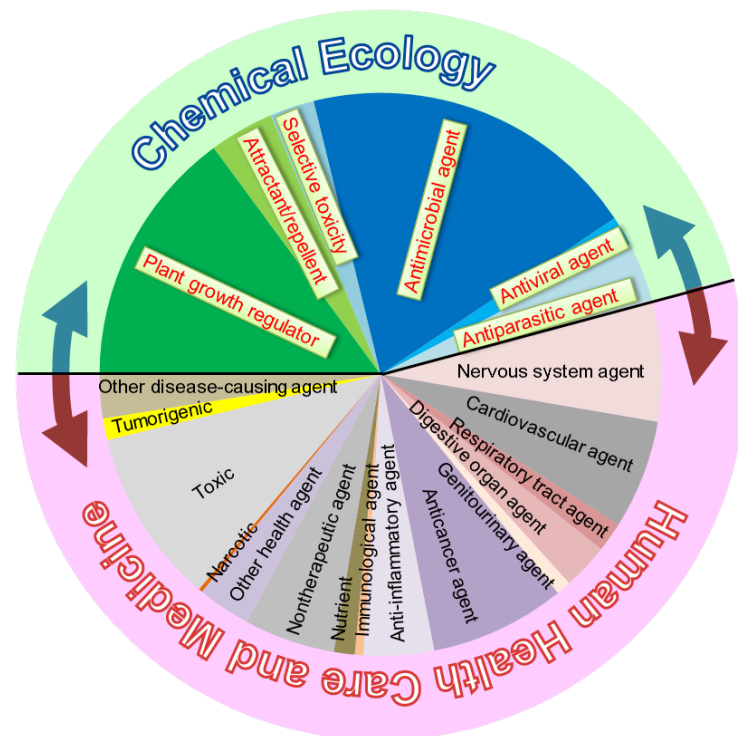
Motorcycle Metabolic Pathway

- 代謝データベース Since 2011.08

Bicycle Algae Metabolic Pathway

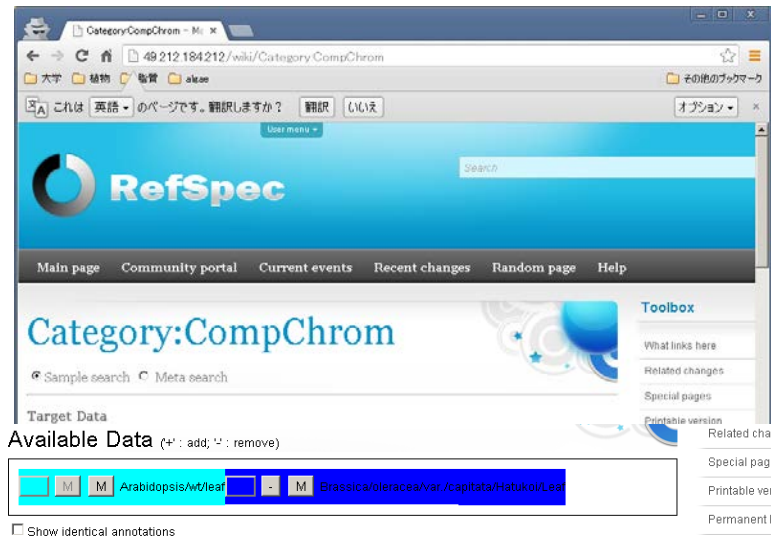
- 代謝データベース Since 2013.09

Skewered KNApSack 串刺し検索 Since 2010.10

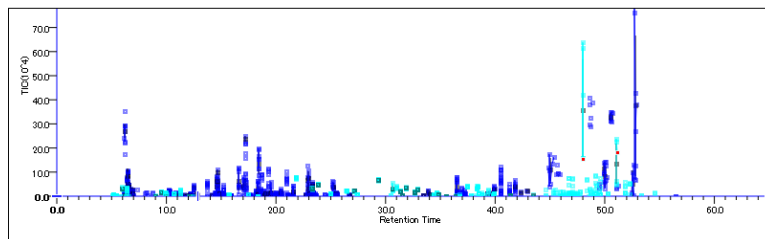


前期までの主要結果

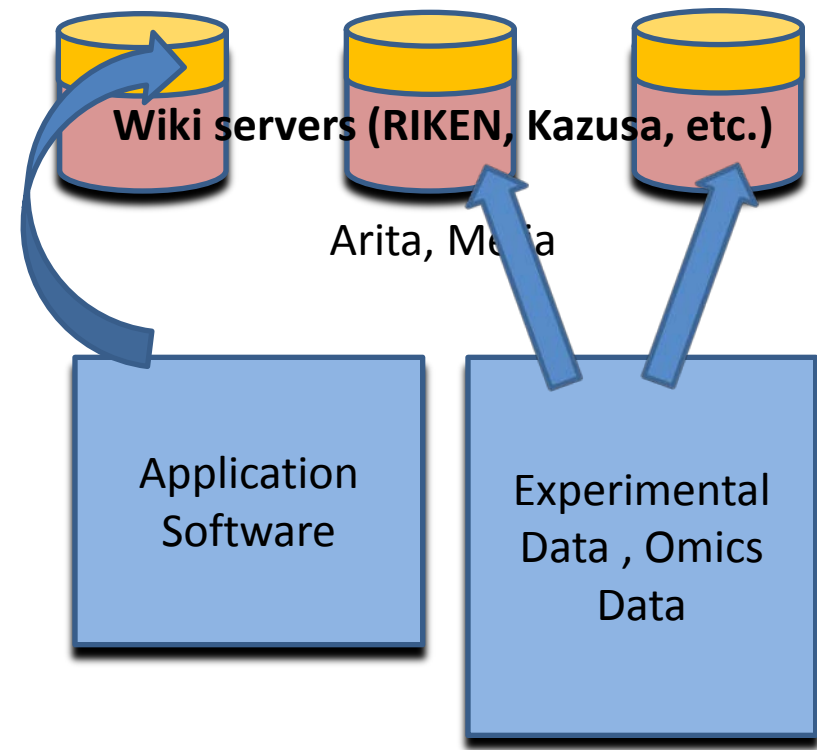
データ共有ソフトの作成 (理研)



Arabidopsis/wt/leaf

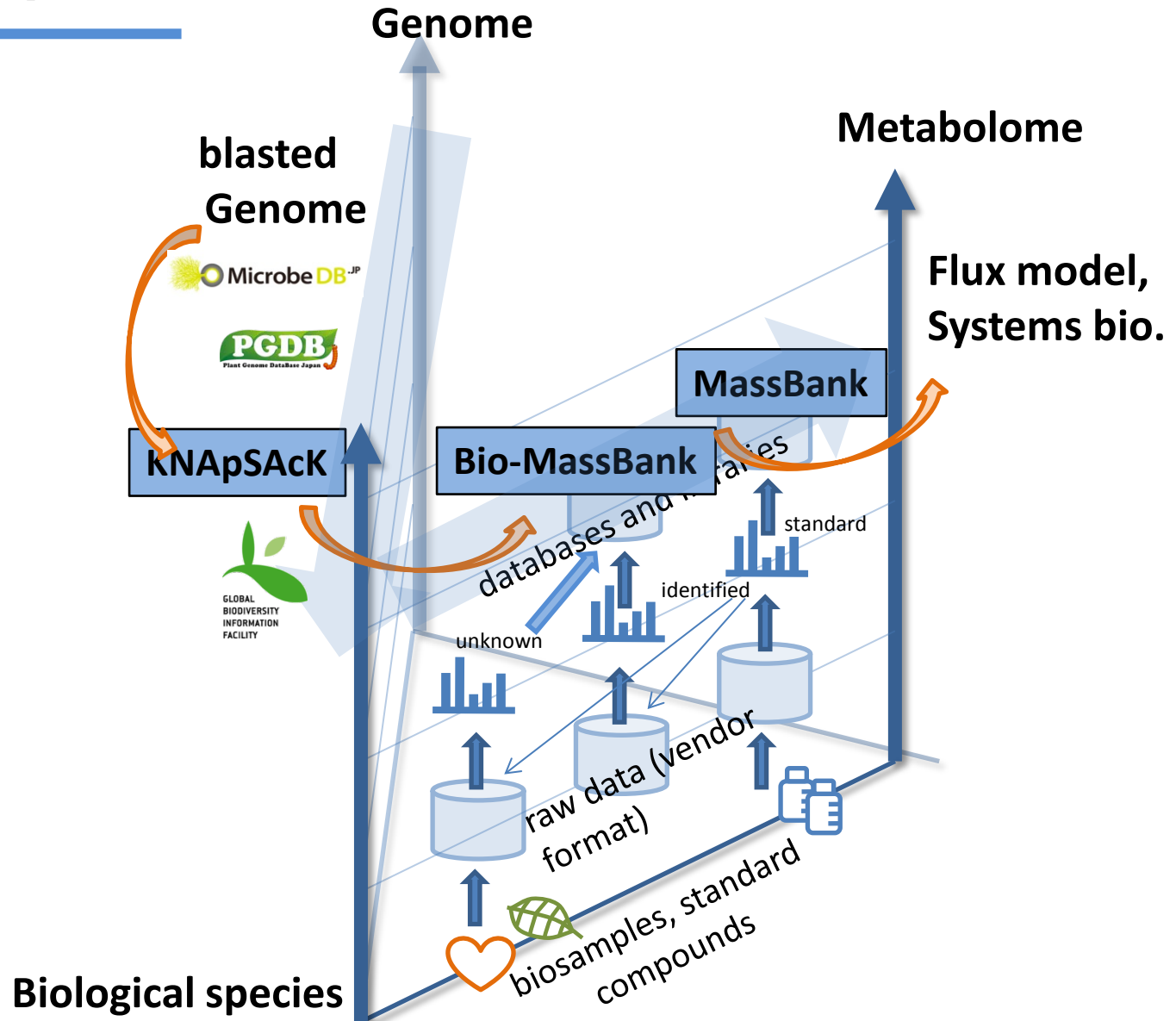


Networking Layer (Akie Mejia)



今回の提案

ゲノム
+ 文献情報
↓
代謝物リスト
↓
代謝物同定
スペクトル
↓
システム生物学



目標

- 化合物情報とスペクトル情報を生物種ごとに整理・公開する

モデル含む20種程度、特に植物

- 活性情報を付した化合物を、代謝マップを通してゲノム情報とリンクする

二次代謝物を中心に、脂質をカバー

- オントロジーと標準フォーマットを整備

アノテーションRDF、整理活性RDFで連携

メタボモデル解説

例： トマト

Asteridae -

Acids, Bases[m]

:

alanine etc.

Solanaceae -

Alkaloids [m]

:

Solanum lycopersicon

tomatine [m]

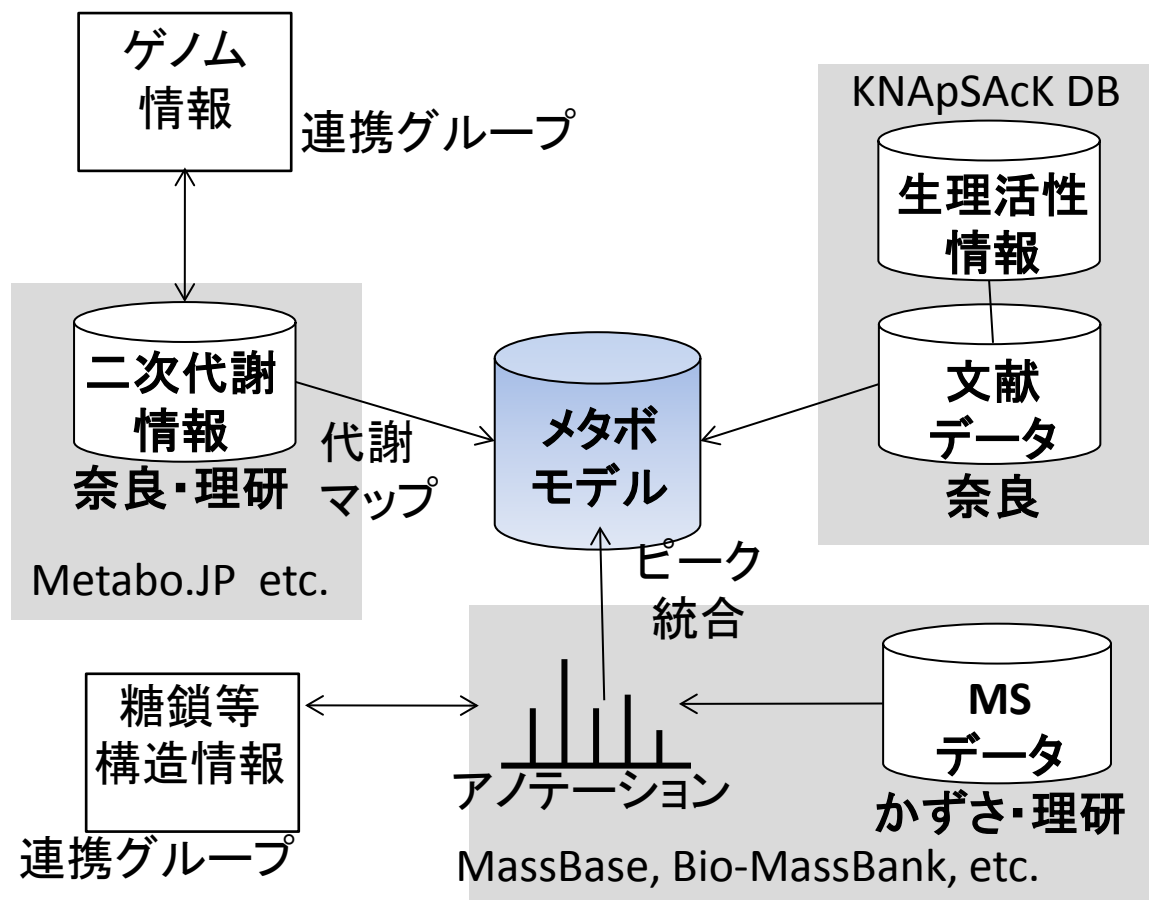
:



検索するDBから、見るDBへ。Bio-MassBank の系統分類版。
各代謝物に、文献に拠る裏付けおよび統合スペクトル情報を付与
物質の生理活性も(分かる場合は)付与 (リンクアウト)

研究体制

旧 MassBank プロジェクト雇用者(鶴岡メンバー)は解散。
KNApSAcKとシステム開発以外は新規雇用。



各グループの研究内容

理研

1. 生物種メタボモデルの構築
2. データ共有基盤ソフトウェアの構築
3. アノテーションソフトウェアの構築

奈良

1. 生物種と代謝物の関係データベースの充実
2. ケミカルエコロジーを中心とした代謝物と生物活性の関係データベース
3. 代謝物の構造情報によるスペクトル情報にもとづいた構造推定法

かずさ

1. メタボロームデータの大規模整備
2. メタデータのRDFによる外部データベースとの連携
3. アノテーションおよびデータ公開加速のための技術開発

目標のまとめ

1. (Bio-)MassBankのスペクトルを拡充・統合し、生物の系統毎に整理、公開
2. KNApSAcKを介して、代謝物とゲノム情報、生理活性をリンク
3. 必要なフォーマット、オントロジー、ツールの作成

