

メタボロミクスの国際連携と データ共有

有田正規（遺伝研、理研CSRS）

arita@nig.ac.jp

社会的関心

[5 March 2014] Human Longevity Inc.

寿命延長のためにマイクロバイーム、メタボローム、臨床データと組み合わせた年間最大4万人のヒトゲノムDB構築
(Metabolon, J・クレイグ・ベンター研究所)

[Nov 2013] NIH Metabolomics Collaboration

2012年のメタボロミクス・コアグラントの継続。

6センター、50 M USD が稼働中。

<http://metabolomicsworkbench.org/>

メタボロミクスの利点

- 細胞の中身を計測できる
 - GC-MS ガスクロ EIイオン化。揮発性、炭化水素向き
 - LC-MS 液クロ ESIイオン化。脂質、高分子向き
 - CE-MS キャピラリー イオン性低分子向き
- 最新機器だと質量精度は 10 ppm
(質量数 1000 のとき0.01 レベル)
- クロマトグラムの理論段数は $10^3 - 10^4$
(これくらい分離できる)

メタボロミクスの現状

細胞内の代謝物を網羅できるわけではない

量、濃度の問題: 質量分析計のダイナミックレンジ 10^5

多いアミノ酸 Gln, Tau など 10^{-2} M / L

少ないアミノ酸 Met など 10^{-4} M / L

ホルモン、サイトカインなど 10^{-8} M / L

個体、部位の問題: オルガネラ単位の計測は無理

代謝を完全に止める方法はない(光、温度、etc.)

食事、共生、個体により大きな差

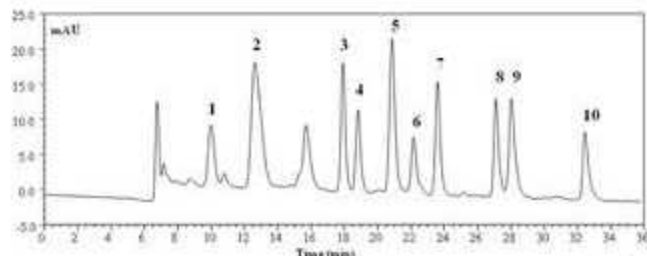
抽出、計測の問題: 網羅的に抽出できる方法はない

抽出は水・エタノール・クロロホルム=1:2:1 が王道 (Bligh-Dyer)

EI, ESIなどイオン化の特徴、クロマトグラムの特徴...

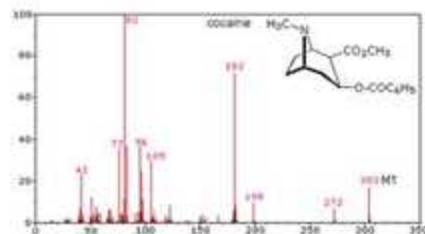
代謝物を「同定」する作業

- クロマトグラムで、ピークが出てくる時間を検証



純品を測っておく

- マススペクトルが純品と一致するか検証



EI-MSの場合はライブラリが存在

ESI-MSのライブラリは純品を測っておく

他に同位体分布やNMRなどもあるが、
とどのつまり、保有する標準品の総数 = 検出力

メタボロミクス学会のガイドライン

論文には安易に「同定した」(identified)とは書けない

Level 1	Identified	同定した	同一機器で測定した標準物質と保持時間・スペクトル・同位体分布など、少なくとも2つ以上の分析結果が一致するもの
Level 2	Putatively annotated	推定した	質量スペクトルやその他の物理化学的特徴から構造が確定したもの
Level 3	Putatively characterized (class)	クラス推定した	構造は確定できないがスペクトル等が構造の類似性を示すもの
Level 4	Unknown	未知	代謝物の存在は明らかだが構造や種類がわからないもの

「同定」の難しさ

- 既知の天然物構造は20万を超えるが、生物界における分布はほとんど不明。標準品も無い。
 - 誰も正解を知らない
 - 検証不能のため、論文は著者の一存による表記
- 研究室間でのデータ比較が不可能。
 - 機器が違えば結果も違う。生データはサイズが大きく、交換すら不可能(次世代シーケンサの結果に似る)
 - 代謝物名の統合すらできないのが現状

これまでとこれから

- MassBank データベースの構築 (PI:西岡孝明)
標品情報を集めるプロジェクト
- Bio-MassBankデータベースの構築 (PI:金谷重彦)
生体情報を集めるプロジェクト



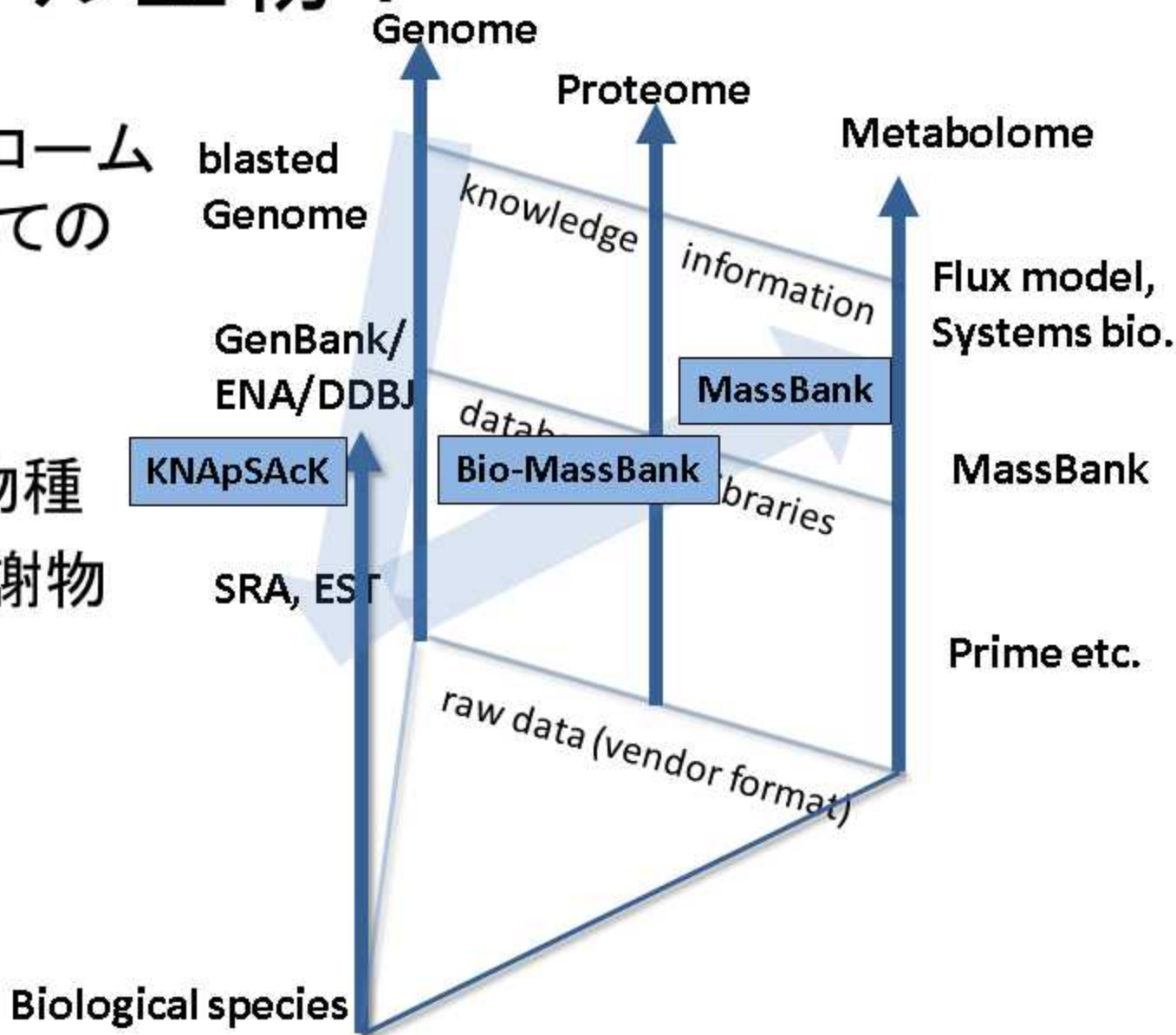
- モデルメタボロームの構築 (PI:有田正規)
 - 今まで集めたものを標準化し、整理する作業
 - 登録数をいかに少なくできるかが鍵標準化のフォーマット、プロトコル: かずさDNA 櫻井さん
整理するのに必要な分子、文献情報: 奈良先 金谷さん



なぜモデル生物？

ゲノム、メタボローム
をつなぐ軸としての
「生物種」


ゲノムから生物種
生物種から代謝物




「同定」の難しさを超える

- 既知の天然物構造は20万を超えるが、生物界における分布はほとんど不明。標準品も無い。
 - 文献より生物種情報を収集 (KNApSAck)
 - 収集した標準スペクトルの整理 (MassBank)
- 研究室間でのデータ比較が不可能。
 - 国際連携 (MassBank)
 - フォーマット、プロトコルの統一 (Metabolonote)
 - アノテーション作業の自動化、ルール化 (MassBase, KomicMarket)


KNApSAcK Update (Kanaya)






“KNApSAcK” Family

Since 2008.07




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
 <p>Lunch Box 食用データベース <small>Since 2008.07</small></p>	 <p>WorldMap 世界の薬用植物データベース <small>Since 2009.06</small></p>	 <p>Biological Activity Natural Activity <small>Since 2011.06</small></p>
 <p>DietNavi 病気予防データベース <small>Since 2012.11</small></p>	 <p>KAMPO 漢方薬、生薬データベース <small>Since 2008.09</small></p>	 <p>Biological Activity Metabolite Activity <small>Since 2013.01</small></p>
 <p>FoodProcessor 加工食品データベース <small>Since 2012.11</small></p>	 <p>JAMU IndonesiaHerbデータベース <small>Since 2009.11</small></p>	
 <p>DietDish 食べ合わせデータベース <small>Since 2012.11</small></p>	 <p>Tea Pot ハーブデータベース <small>Since 2013.09</small></p>	
 <p>MARCHÉ 旬データベース <small>Since 2014.04</small></p>	 <p>Skewered KNApSAcK 串刺し検索 <small>Since 2010.10</small></p>	


Picnic Gene Annotation


 <p>Arabidopsis <small>Since 2008.04</small></p>
 <p>Bacillus <small>Since 2008.05</small></p>
 <p>Human <small>Since 2009.03</small></p>

Strap Correlation Coefficient


 <p>Arabidopsis <small>Since 2009.09</small></p>
 <p>Bacillus</p>

Pickaxe Metalloprotein Database

 <p>MetalMine <small>Since 2009.08</small></p>
--



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

● Nervous system agent

- Antipyretic
- Analgesic
- Antiarthritic
- Anesthetic
- Sedative
- Antispasmodic
- Anticonvulsant
- Antidementic
- Antidepressant
- CNS Stimulant
- Diaphoretic
- Emetic
- Antiemetic
- Antigout
- Antimigraine
- Antimyasthenic
- Antiparkinson
- Antipsychotic
- Muscle relaxant

● Cardiovascular agent

- Antidiabetic
- Hemostatic
- Antithrombotic
- Cardiotonic
- Antiarrhythmic
- Diuretic
- Hypotensive
- Atrial fibrillation agent
- Antianemic
- Other cardiovascular agent

● Respiratory tract agent

- Antitussive
- Expectorant
- Antiasthmatic
- Other respiratory tract agent

● Digestive organ agent

- Antidiarrheic
- Carminative
- Stomachic
- Laxative
- Choleric
- Antihepatotoxic
- Other digestive organ agent

● Genitourinary agent

- Oxytocic
- Antifertility
- Abortifacient
- Other genitourinary agent

● Anti-cancer agent

- Antioxidant
- Anticancer
- Antitumor
- Antineoplastic
- Antimutagenic

● Anti-inflammatory agent

- Antiinflammatory
- Antiallergic
- UV shield
- Antidermatitic
- Antiedemic

● Immunological agent

- Immunosuppressant
- Immunostimulant
- Immunomodulative

● Nutrient

- Nucleic acid
- Essential amino acid
- Nonessential amino acid
- Vitamin
- Nutrient
- Tonic

● Not therapeutic agent

- Solvent
- Flavor
- Odor
- Pigment
- Emulsifying agent
- Antiseptic

● Anti-microbial

- Antibacterial
- Antituberculous
- Antifungal
- Inhibit spore germination
- Antimicrobial

● Anti-viral

- Antiviral
- Antileprotic
- Antihepatic
- Anti-HIV
- Anti-HSV

● Anti-parasitic

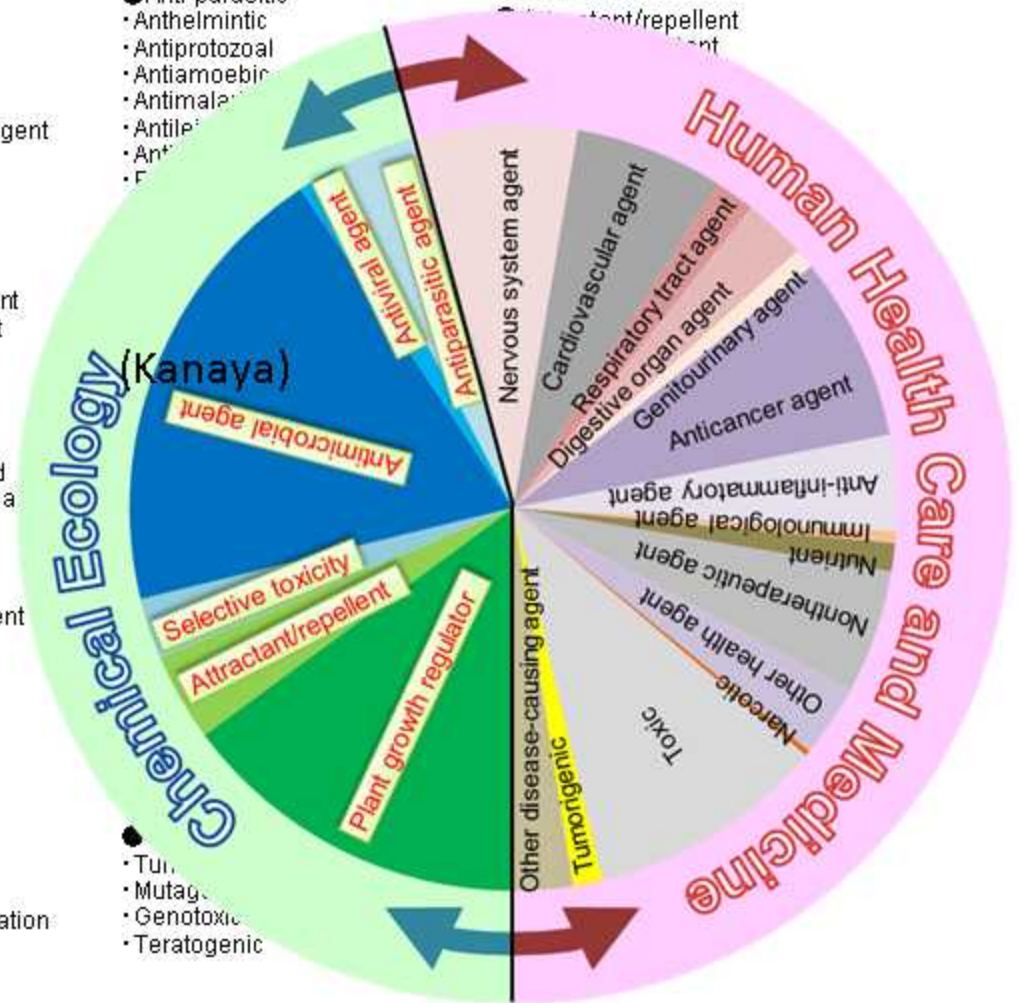
- Anthelmintic
- Antiprotozoal
- Antiamoebic
- Antimalarial
- Antileishmanicidal
- Antitrypanosomal
- Fungicide

● Other disease-causing

- Psychotomimetic
- Hemolytic
- Allergenic
- Irritant
- Dermatitic
- Edematous

● Stimulant/repellent

- Stimulant/repellent



- Tumorigenic
- Mutagenic
- Genotoxic
- Teratogenic

国際連携その1

- MetabolomeXchange (or MetX) @ Leiden
with Metabolights, MetabolomicsWorkbench, GMD,
Bordeaux MD, HMDB, MassBank

データの公開、共有のためのコンソーシアム。DB側は
private, public の区別を付ける必要あり。privacy
protection lawにも準拠。(本申請ではヒトを載せず。)

メタデータをMetX XML にてportalに提供。

10/10に締結予定

国際連携その2

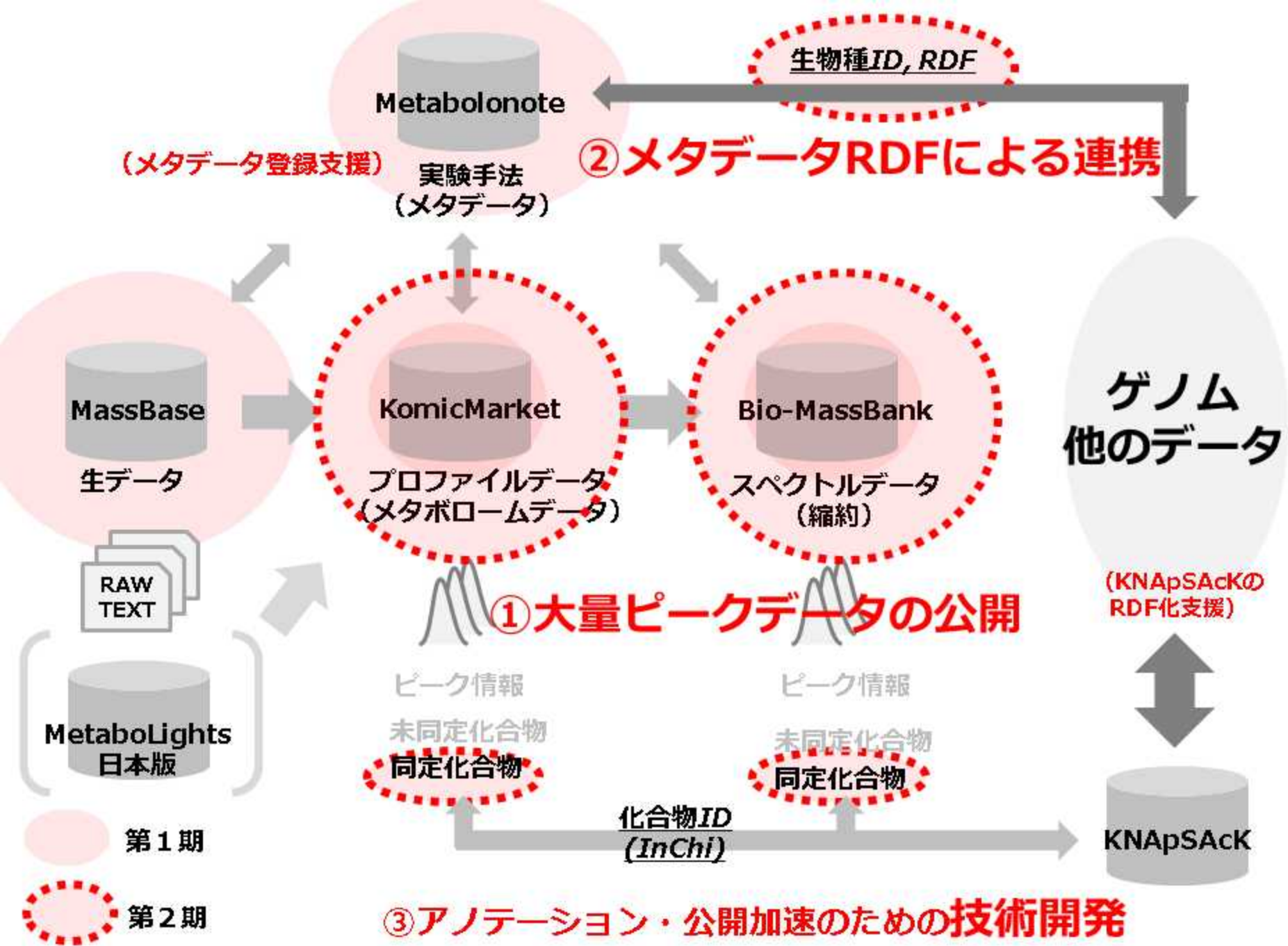
- Norman MassBank @ eawag, UfZ

チュートリアルを毎年開催する環境物質のスペクトルサーバー。データを Subversion にて共有。

- MassBank of North America @ Davis US

O.Fiehnグループのスペクトルを MassBank ブランチとして来年度より公開予定 (JST-NSF 予算)。

MassBankの中で、メタデータおよびMS²の整理が必要。



生物種ID, RDF

② メタデータRDFによる連携

Metabolonote

実験手法 (メタデータ)

(メタデータ登録支援)

MassBase

生データ

RAW TEXT

MetaboLights 日本版

KomicMarket

プロファイルデータ (メタボロームデータ)

Bio-MassBank

スペクトルデータ (縮約)

ゲノム 他のデータ

(KNApSackKの RDF化支援)

① 大量ピークデータの公開

ピーク情報
未同定化合物
同定化合物

ピーク情報
未同定化合物
同定化合物

化合物ID (InChi)

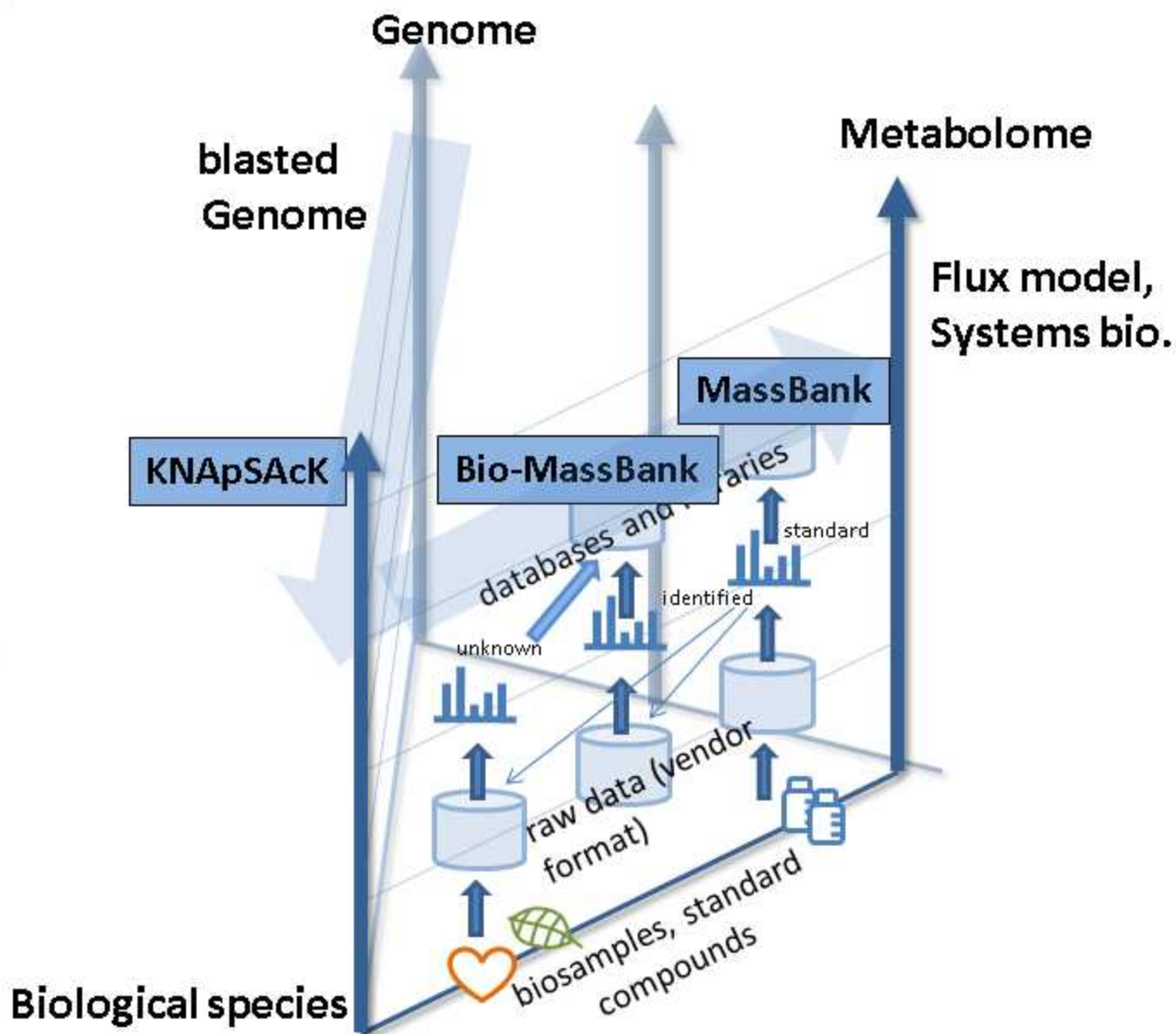
KNApSackK

③ アノテーション・公開加速のための技術開発

第1期

第2期

これから



見通し,まとめ

- 代謝物を同定する際のプロトコルを国内でまとめて海外へも提案
 - 国際メタボロミクス学会のワーキンググループ
 - MetabolomeXchange
- 特徴ピークの情報から、スペクトルに基づく化合物推定が飛躍的に向上
 - 母核構造ごとに仮想ライブラリを作成
 - リソースを MassBank その他から公開

これらの作業は、学問として必須の要件

謝辞

コアメンバー: 金谷重彦、櫻井望、津川裕司、Ramon Mejia, 福島敦史

Massbank, KNApSAcK, Metabolonote, MassBase,
MetabolomeXchange, Norman MassBank