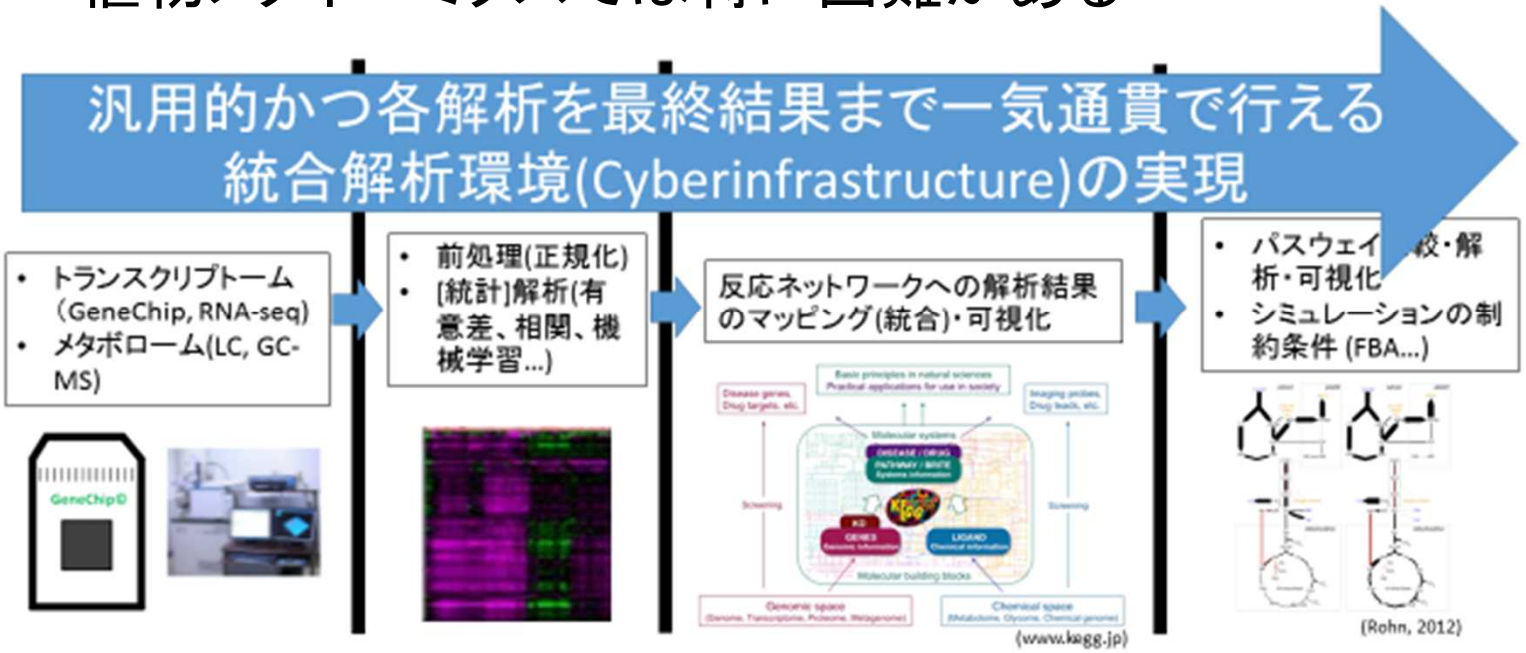
生化学反応ネットワーク統合解析環境の拡張

理化学研究所
生命システム研究センター
西田孝三

1. 研究概要
2. ユースケースと成果
3. 新展開と今後

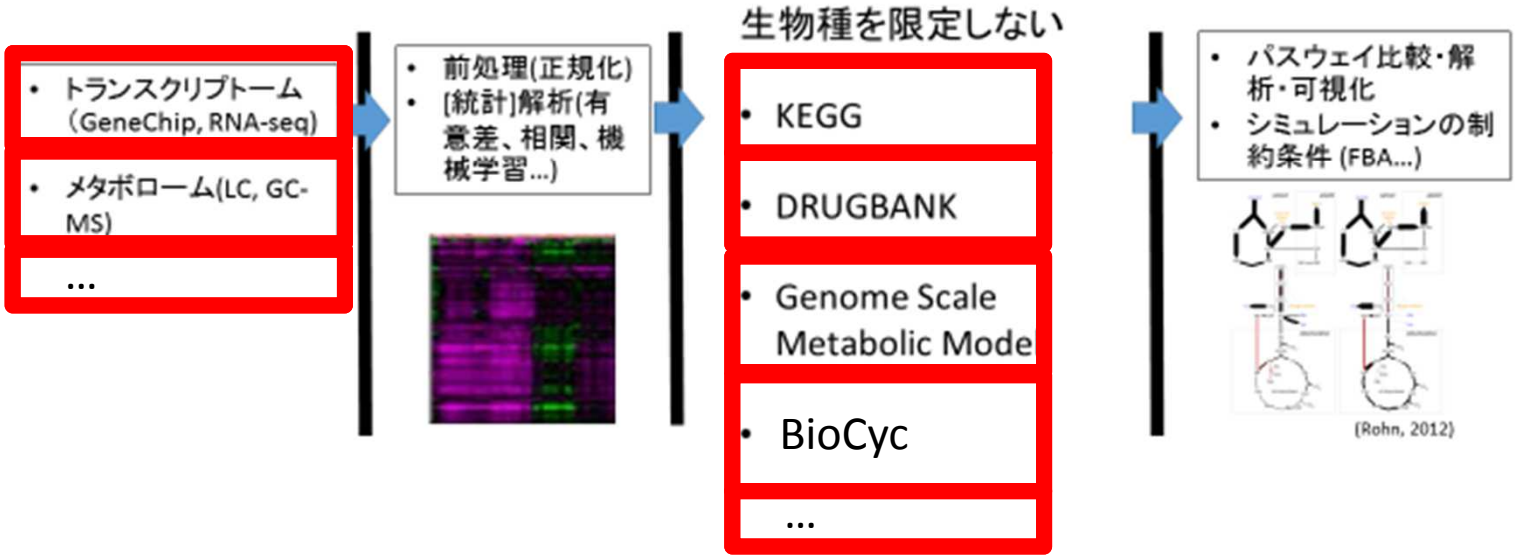
背景、目的

- オミクス実験結果からパスウェイ解析まで一通り行えるフリーな統合解析環境(Cyberinfrastructure)の実現
  - 統合化推進プログラムのデータベースを対象
  - 各異種解析を統合する環境
  - 植物メタボロミクスでは特に困難がある



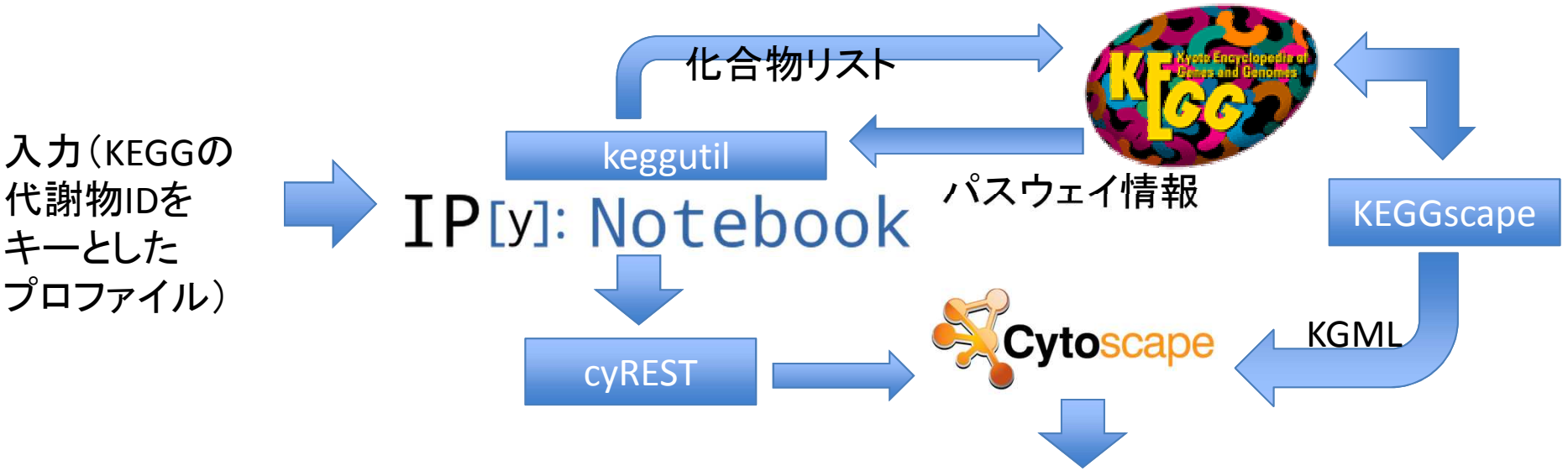
前年度との違い

- 入力情報の拡張（発現プロファイルから代謝物プロファイル）
  - Pathview(Luo, 2013) でも未熟
- 再現性が高く、拡張の用意な解析パイプラインの提供
  - 解析の変更、追加も容易に(リバイズ時など)



研究開発成果概要

- 代謝物プロファイル可視化パイプライン
  - KEGGマッピング用Pythonパッケージkeggutilと Cytoscape app (cyREST, KEGGscape)を組み合わせたIPython Notebookパイプライン(github.com/kozo2/keggutil, togotrial2014)
  - なぜGalaxy(Goecks, 2010)を使わなかったか



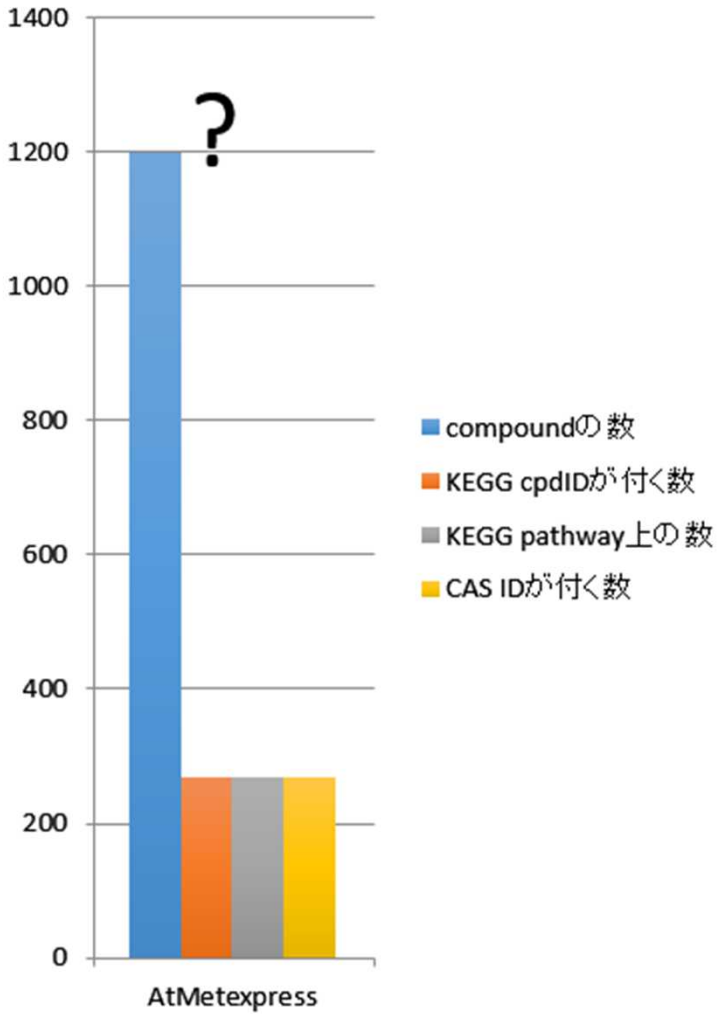
ユースケース

- AtMetExpress
  - 前年度統合解析トライアルの成果DB
  - シロイヌナズナの代謝物~1,200個のプロファイルデータベース
  - パスウェイ統合は未達成
  - データセット名Matsuda, 2010のみマルチオミクス解析可能な発現プロファイル有
    - 実験条件が合うプロテオミクスデータセット発見できず

結果

- Col0-timecourseのleaf-root間時系列プロファイル比較
  - http://nbviewer.ipython.org/github/kozo2/togotrial2014/blob/master/leaf-root.ipynb
  - https://github.com/kozo2/togotrial2014/blob/master/leaf-root.svg

達成できたこと、できなかったこと



できたこと

- 再現性が高く、詳細な化合物プロファイルの可視化

できなかったこと

- 多くの化合物のDB統合
  - 化合物オントロジーを用いたエンリッチメント解析
  - ゲノムスケール代謝モデルとの比較
- パイプラインの汎用化、RDFの活用、データフォーマットの定義

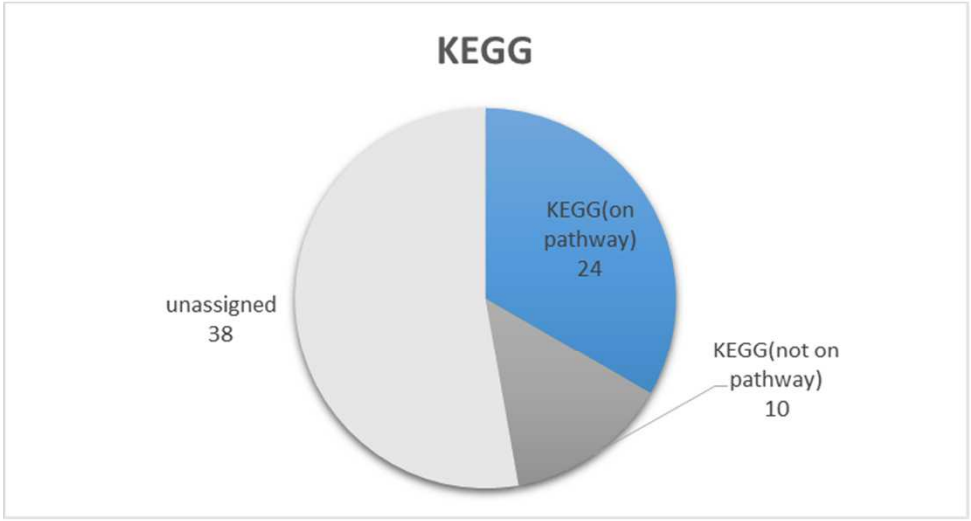
AtMetExpressの パスウェイマッピングにおける難題

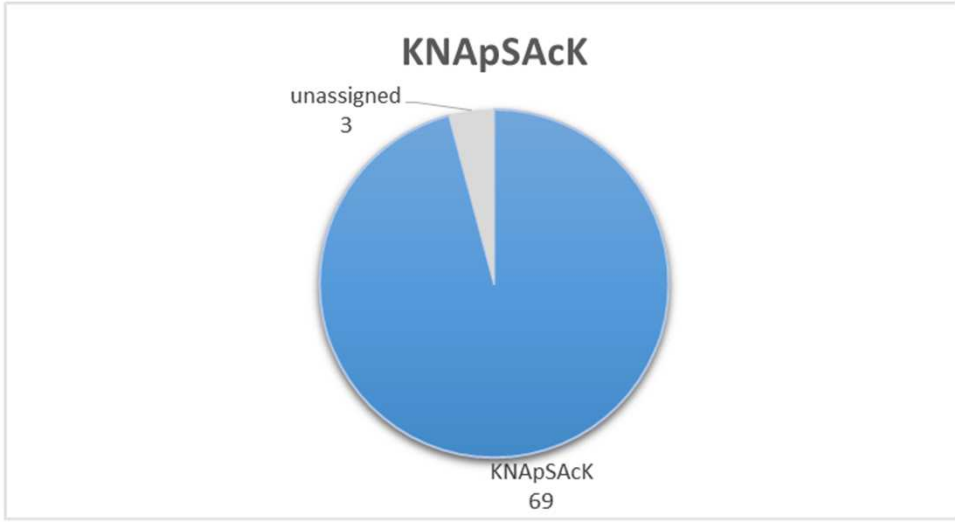
- 化合物におけるIDの問題
  - 代謝物のアノテーションのレベルの差
    - 機器の違い、標品の有無、立体化学など
    - 曖昧な化合物名がIDとして用いざるを得なくなる
  - 植物特有の事情(二次代謝パスウェイの重要性)
    - エントリ自体パスウェイデータベースには無いことがある
- マニュアルキュレーションが必要
  - 現状の情報基盤では機械的処理は困難

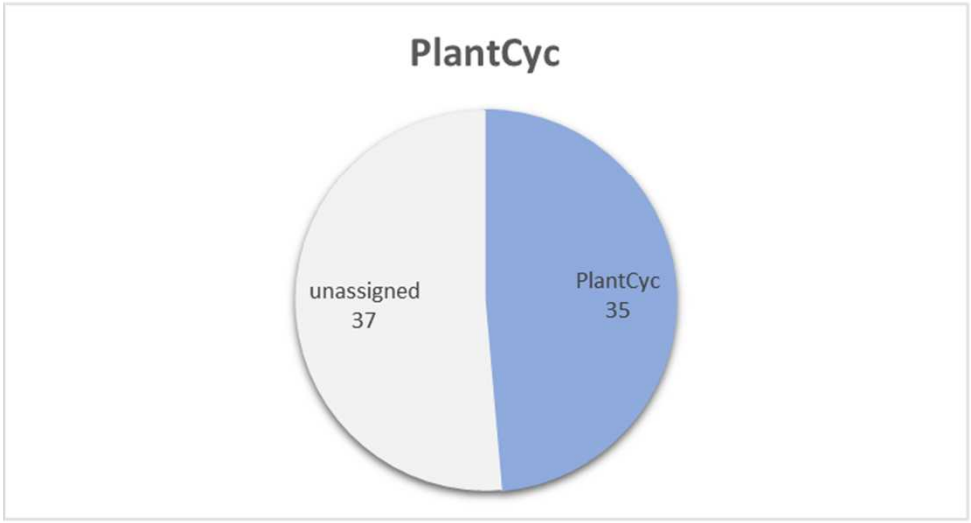
新展開から生まれた開発計画

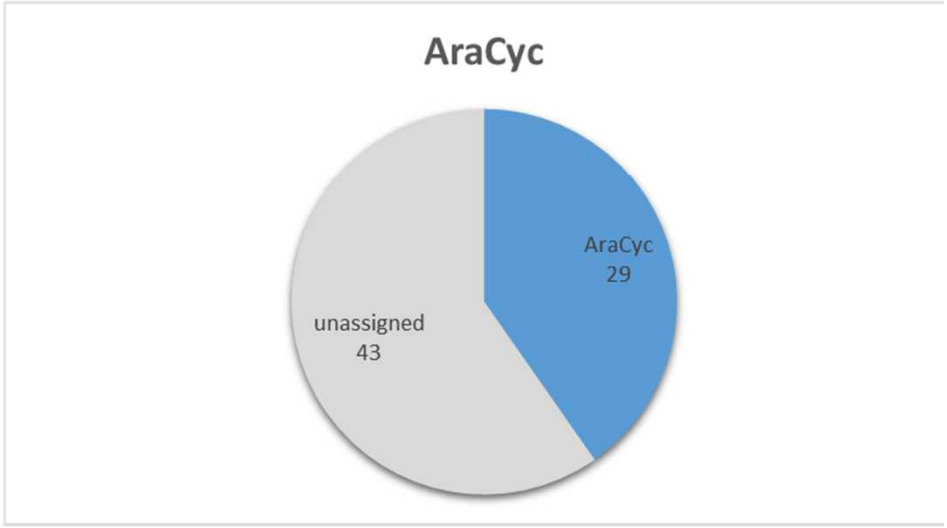
- 代謝物のマニュアルキュレーション(DBCLS 時松氏に拠る)
  - 二次代謝(Flavonoid)にフォーカス
  - 天然物や二次代謝に特化したKNApSAcKデータベースを中心に複数DBと照合
  - Saito, Plant Physiol Biochem. 2013 のscaffold structureに基づいたパスウェイ図の作成

72 Flavonoidsのデータベース参照可否の図

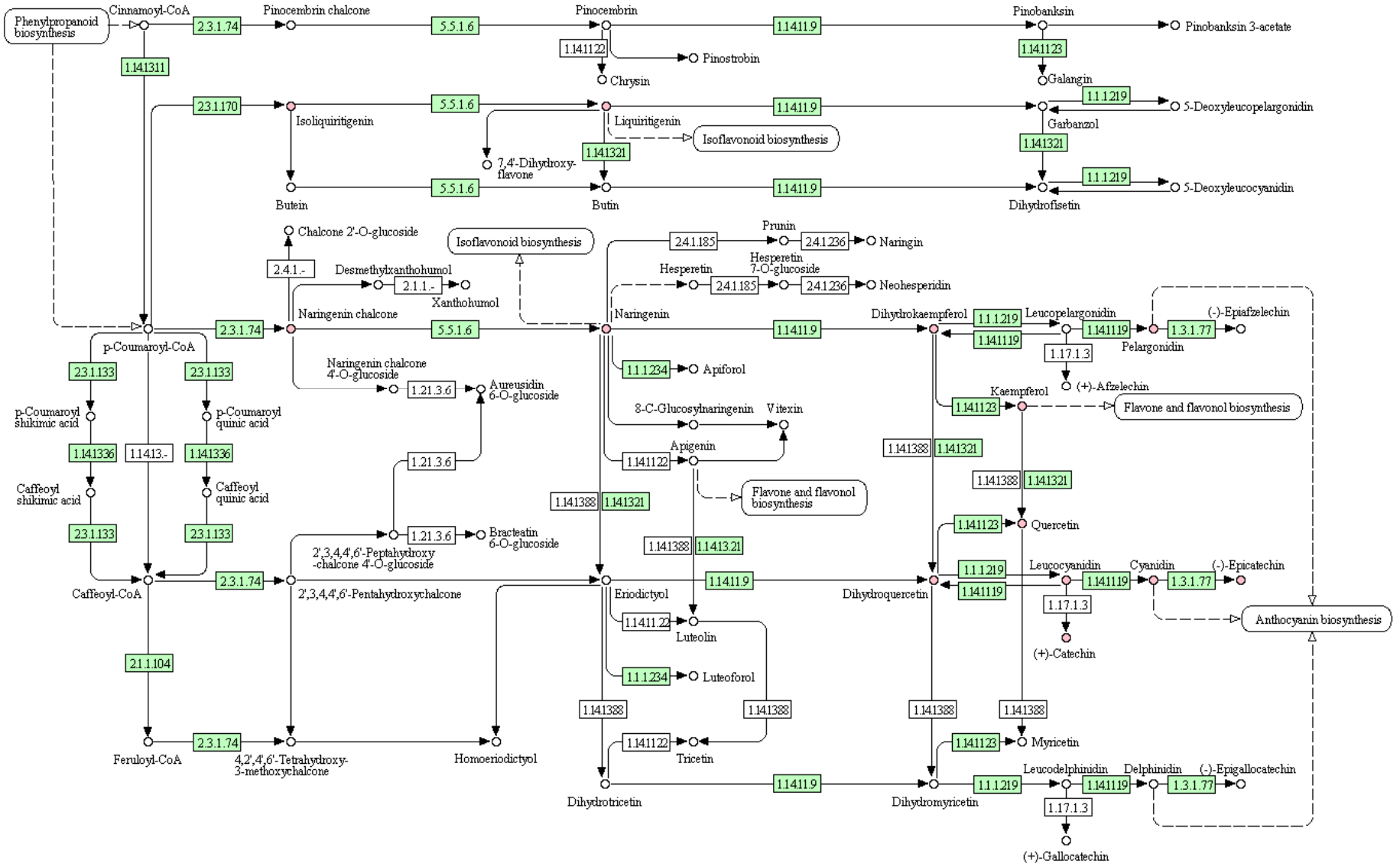


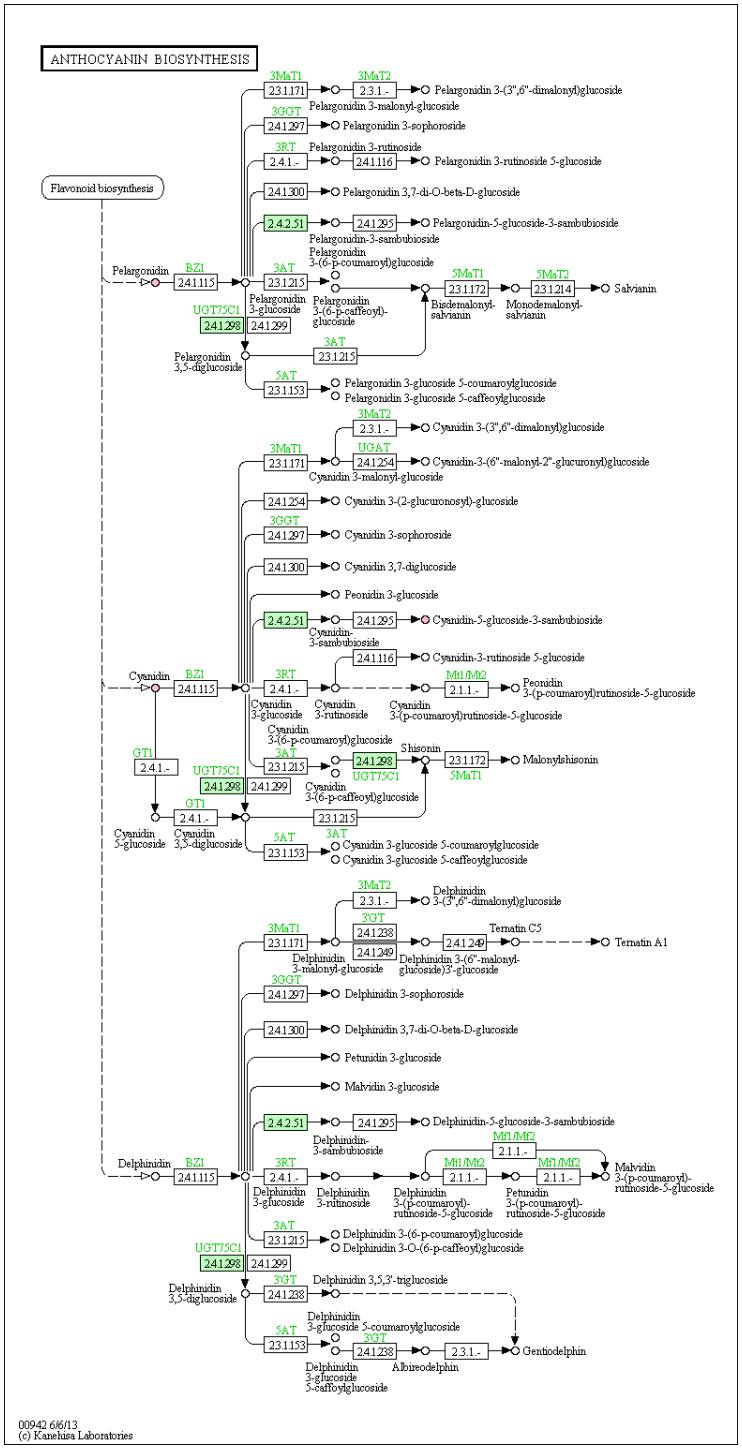


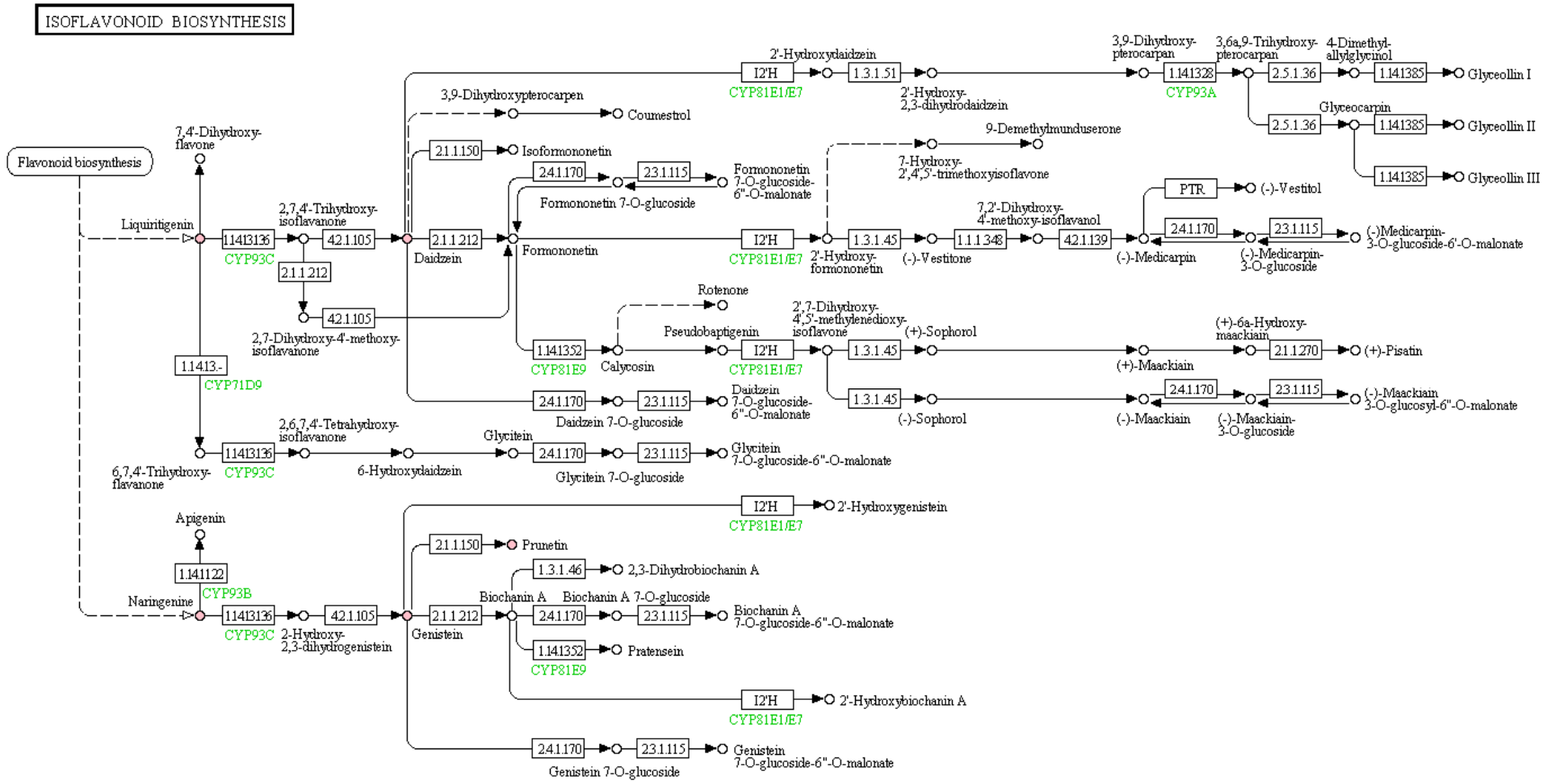


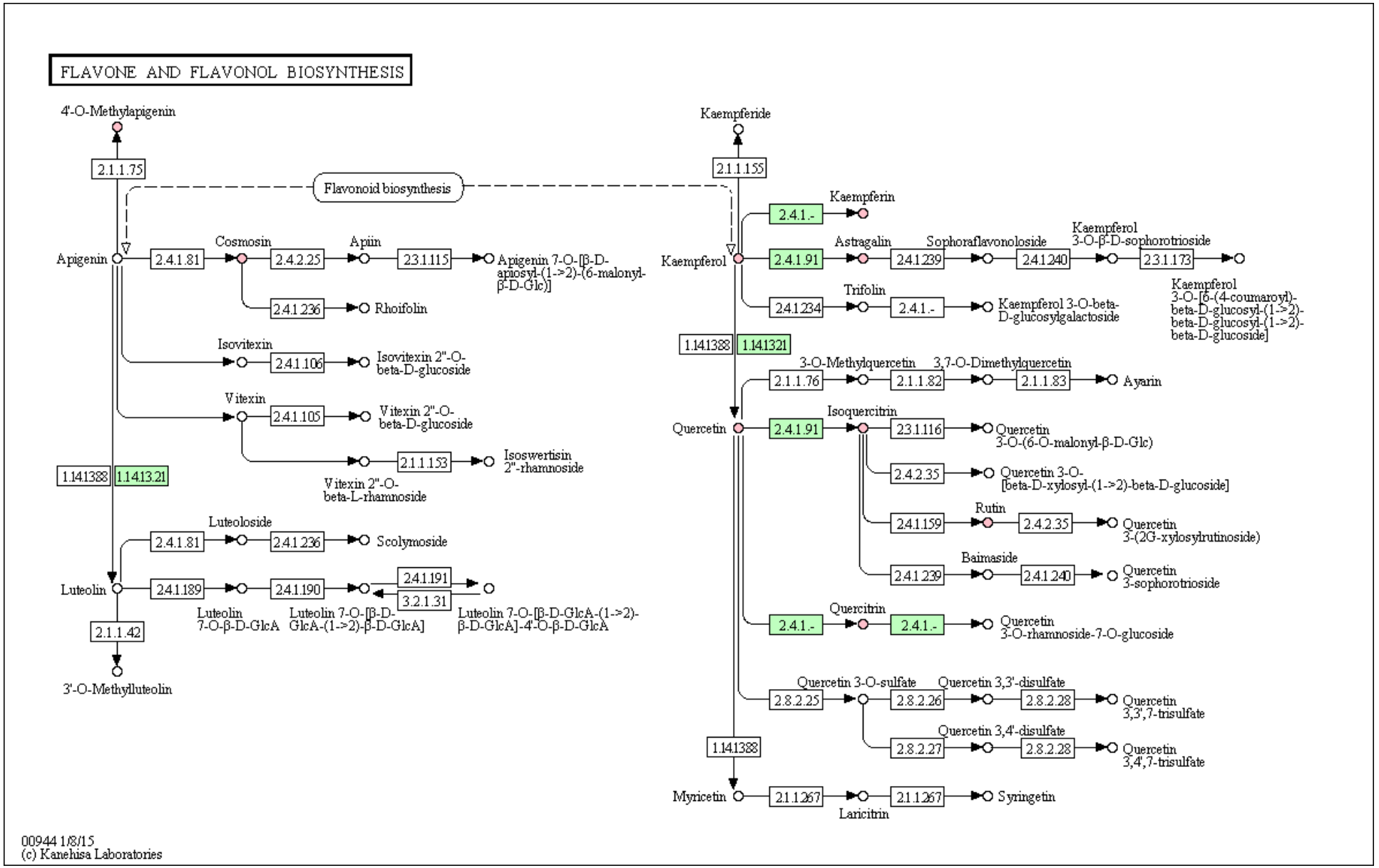


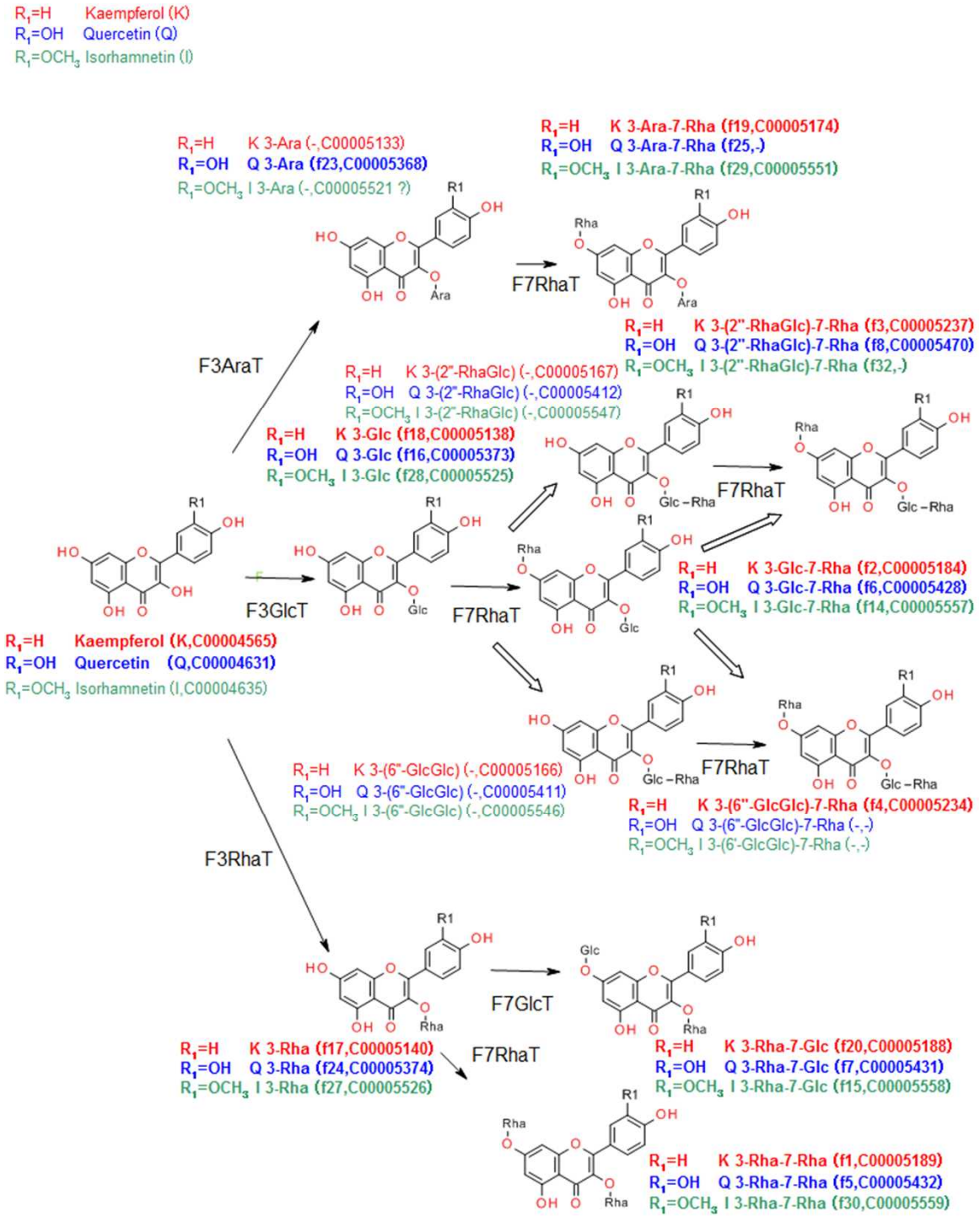
FLAVONOID BIOSYNTHESIS




ANTHOCYANIN BIOSYNTHESIS
Flavonoid biosynthesis
Pelargonidin
Cyanidin
Delphinidin
Pelargonidin 3-(3",6"-dimalonyl)glucoside
Pelargonidin 3-malonyl-glucoside
Pelargonidin 3-sophoroside
Pelargonidin 3-rutinoside
Pelargonidin 3-rutinoside 5-glucoside
Pelargonidin 3,7-di-O-beta-D-glucoside
Pelargonidin-5-glucoside-3-sambubioside
Pelargonidin-3-sambubioside
Pelargonidin 3-(6-p-coumaroyl)glucoside
Pelargonidin 3-glucoside
Pelargonidin 3-(6-p-caffeoyl)-glucoside
Pelargonidin 3,5-diglucoside
Pelargonidin 3-glucoside 5-coumaroylglucoside
Pelargonidin 3-glucoside 5-caffeoylglucoside
Salvianin
Bisdemalonyl-salvianin
Monodemalonyl-salvianin
Cyanidin 3-(3",6"-dimalonyl)glucoside
Cyanidin-3-(6"-malonyl-2"-glucuronyl)glucoside
Cyanidin 3-malonyl-glucoside
Cyanidin 3-(2-glucuronosyl)-glucoside
Cyanidin 3-sophoroside
Cyanidin 3,7-diglucoside
Peonidin 3-glucoside
Cyanidin-5-glucoside-3-sambubioside
Cyanidin-3-sambubioside
Cyanidin-3-rutinoside 5-glucoside
Peonidin 3-(p-coumaroyl)rutinoside-5-glucoside
Cyanidin 3-glucoside
Cyanidin 3-rutinoside
Cyanidin 3-(p-coumaroyl)rutinoside-5-glucoside
Cyanidin 3-(6-p-coumaroyl)glucoside
Shisonin
Malonylshisonin
Cyanidin 3-(6-p-caffeoyl)glucoside
Cyanidin 3-glucoside 5-coumaroylglucoside
Cyanidin 3-glucoside 5-caffeoylglucoside
Delphinidin 3-(3",6"-dimalonyl)glucoside
Ternatin C5
Ternatin A1
Delphinidin 3-malonyl-glucoside
Delphinidin 3-(6"-malonyl-glucoside)3'-glucoside
Delphinidin 3-sophoroside
Delphinidin 3,7-di-O-beta-D-glucoside
Petunidin 3-glucoside
Malvidin 3-glucoside
Delphinidin-5-glucoside-3-sambubioside
Delphinidin-3-sambubioside
Malvidin 3-(p-coumaroyl)-rutinoside-5-glucoside
Delphinidin 3-glucoside
Delphinidin 3-rutinoside
Delphinidin 3-(p-coumaroyl)-rutinoside-5-glucoside
Petunidin 3-(p-coumaroyl)-rutinoside-5-glucoside
Delphinidin 3-(6-p-coumaroyl)glucoside
Delphinidin 3-O-(6-p-caffeoyl)glucoside
Delphinidin 3,5,3'-triglucoside
Delphinidin 3-glucoside 5-coumaroylglucoside
Gentiodelphin
Albureodelphin
Delphinidin 3-glucoside 5-caffoylglucoside
00942 6/6/13
(c) Kanehisa Laboratories


ISOFLAVONOID BIOSYNTHESIS
Flavonoid biosynthesis
7,4'-Dihydroxy-flavone
Liquiritigenin
1.14.13.36
CYP93C
2,7,4'-Trihydroxy-isoflavanone
4.2.1.105
2.1.1.212
Daidzein
2,7-Dihydroxy-4'-methoxy-isoflavanone
1.14.13.-
CYP71D9
6,7,4'-Trihydroxy-flavanone
1.14.13.36
CYP93C
2,6,7,4'-Tetrahydroxy-isoflavanone
6-Hydroxydaidzein
Apigenin
1.14.11.22
CYP93B
Naringenine
1.14.13.36
CYP93C
2-Hydroxy-2,3-dihydrogenistein
4.2.1.105
Genistein
3,9-Dihydroxypterocarpen
Coumestrol
2.1.1.150
Isoformononetin
2.4.1.170
2.3.1.115
Formononetin 7-O-glucoside-6"-O-malonate
Formononetin 7-O-glucoside
Formononetin
I2'H
CYP81E1/E7
2'-Hydroxydaidzein
1.3.1.51
2'-Hydroxy-2,3-dihydrodaidzein
9-Demethylmunduserone
7-Hydroxy-2',4',5'-trimethoxyisoflavone
2'-Hydroxy-formononetin
1.3.1.45
(-)-Vestitone
1.1.1.348
7,2'-Dihydroxy-4'-methoxy-isoflavanol
4.2.1.139
(-)-Medicarpin
PTR
(-)-Vestitol
2.4.1.170
2.3.1.115
(-)-Medicarpin-3-O-glucoside
(-)-Medicarpin-3-O-glucoside-6'-O-malonate
Rotenone
Pseudobaptigenin
1.14.13.52
CYP81E9
Calycosin
2',7-Dihydroxy-4',5'-methylenedioxy-isoflavone
(+)-Sophorol
(-)-Sophorol
(+)-Maackiain
(-)-Maackiain
(+)-6a-Hydroxy-maackiain
2.1.1.270
(+)-Pisatin
(-)-Maackiain-3-O-glucoside
(-)-Maackiain-3-O-glucosyl-6"-O-malonate
Daidzein 7-O-glucoside
Daidzein 7-O-glucoside-6"-O-malonate
Glycitein
Glycitein 7-O-glucoside
Glycitein 7-O-glucoside-6"-O-malonate
2'-Hydroxygenistein
Prunetin
1.3.1.46
2,3-Dihydrobiochanin A
Biochanin A
Biochanin A 7-O-glucoside
Biochanin A 7-O-glucoside-6"-O-malonate
1.14.13.52
Pratensein
2'-Hydroxybiochanin A
Genistein 7-O-glucoside
Genistein 7-O-glucoside-6"-O-malonate
3,9-Dihydroxy-pterocarpan
3,6a,9-Trihydroxy-pterocarpan
CYP93A
2.5.1.36
4-Dimethyl-allylglycinol
1.14.13.85
Glyceollin I
Glyceocarpin
Glyceollin II
Glyceollin III
00943 11/19/13
(c) Kanehisa Laboratories


FLAVONE AND FLAVONOL BIOSYNTHESIS
4'-O-Methylapigenin
2.1.1.75
Flavonoid biosynthesis
Apigenin
2.4.1.81
Cosmosin
2.4.2.25
Apiin
2.3.1.115
Apigenin 7-O-[β-D-apiosyl-(1->2)-(6-malonyl-β-D-Glc)]
2.4.1.236
Rhoifolin
Isovitexin
2.4.1.106
Isovitexin 2"-O-beta-D-glucoside
Vitexin
2.4.1.105
Vitexin 2"-O-beta-D-glucoside
2.1.1.153
Isoswertisin 2"-rhamnoside
Vitexin 2"-O-beta-L-rhamnoside
1.14.13.88
1.14.13.21
Luteoloside
2.4.1.81
2.4.1.236
Scolymoside
Luteolin
2.4.1.189
2.4.1.190
2.4.1.191
3.2.1.31
Luteolin 7-O-β-D-GlcA
Luteolin 7-O-[β-D-GlcA-(1->2)-β-D-GlcA]
Luteolin 7-O-[β-D-GlcA-(1->2)-β-D-GlcA]-4'-O-β-D-GlcA
2.1.1.42
3'-O-Methylluteolin
Kaempferide
2.1.1.155
Kaempferin
2.4.1.-
Kaempferol
Astragalin
2.4.1.91
Sophoraflavonoloside
2.4.1.239
Kaempferol 3-O-β-D-sophorotrioside
2.4.1.240
2.3.1.173
Kaempferol 3-O-[6-(4-coumaroyl)-beta-D-glucosyl-(1->2)-beta-D-glucosyl-(1->2)-beta-D-glucoside]
Trifolin
2.4.1.234
2.4.1.-
Kaempferol 3-O-beta-D-glucosylgalactoside
1.14.13.88
1.14.13.21
3-O-Methylquercetin
2.1.1.76
3,7-O-Dimethylquercetin
2.1.1.82
2.1.1.83
Ayarin
Quercetin
Isoquercitrin
2.4.1.91
23.1.116
Quercetin 3-O-(6-O-malonyl-β-D-Glc)
2.4.2.35
Quercetin 3-O-[beta-D-xylosyl-(1->2)-beta-D-glucoside]
Rutin
2.4.1.159
2.4.2.35
Quercetin 3-(2G-xylosylrutinoside)
Baimaside
2.4.1.239
2.4.1.240
Quercetin 3-sophorotrioside
Quercitrin
2.4.1.-
2.4.1.-
Quercetin 3-O-rhamnoside-7-O-glucoside
Quercetin 3-O-sulfate
Quercetin 3,3'-disulfate
2.8.2.25
2.8.2.26
2.8.2.28
Quercetin 3,3',7-trisulfate
Quercetin 3,4'-disulfate
2.8.2.27
2.8.2.28
Quercetin 3,4',7-trisulfate
1.14.13.88
Myricetin
2.1.1.267
2.1.1.267
Syringetin
Laricitrin
00944 1/8/15
(c) Kanehisa Laboratories



原子の由来を意識した
パスウェイ図
(DBCLS時松氏に拠る)

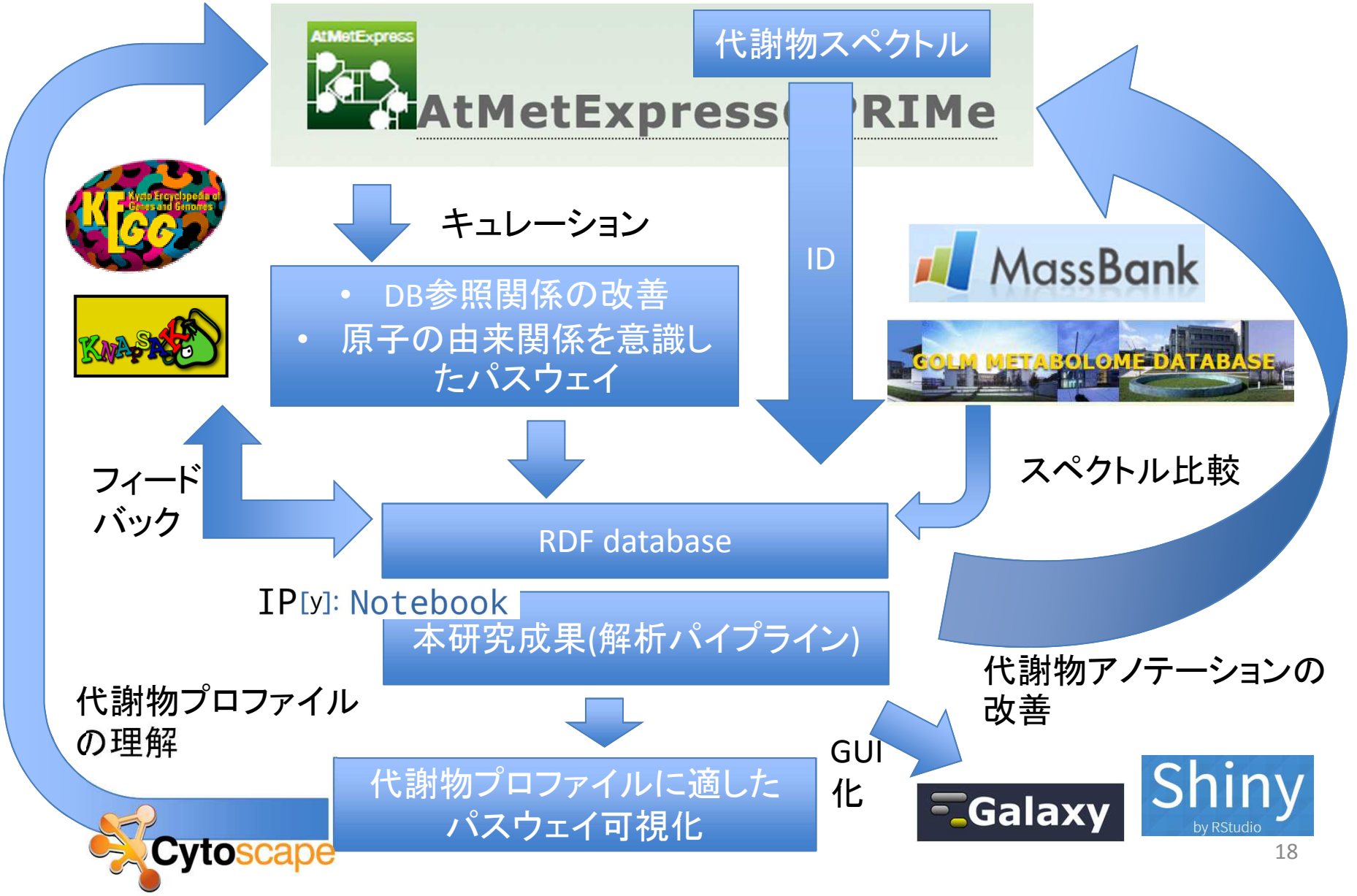
開発の今後

- Glucosinolate, alkaloid, lipidへのキュレーション、パスウェイ図作成(必要に応じ)
- スペクトル情報のリンク、活用
  - Massbank, Golm metabolome database
- RDFデータベース化、公開と解析パイプラインの統合

なぜ代謝物プロファイルの活用にRDFが必要か

- DB参照情報がプロファイルDBで提供できていない
  - 組織を超えた情報リンクを推進する必要がある。
  - データ共有が発現プロファイルより難しい。遅れている
- 自動処理に統一的なデータ保存、取得方法が必要
  - 事前にフォーマットを決める必要がない。代謝物情報を整理するためのフォーマットを定めることは難しい
  - 独自フォーマットはデータ共有を困難にする
- アノテーションの明確化、ID問題解決への手がかり
  - 曖昧な化合物名への依存からの脱却。スペクトル情報をURIとしアノテーションを定義するなど
  - 単一の情報源には依ることができない。複数の情報源を参照する必要がある

本研究の将来性


AtMetExpress
代謝物スペクトル
AtMetExpress
RIMe
キュレーション
ID
MassBank
• DB参照関係の改善
• 原子の由来関係を意識したパスウェイ
GOLM METABOLOME DATABASE
フィードバック
スペクトル比較
RDF database
IP[y]: Notebook
本研究成果(解析パイプライン)
代謝物プロファイルの理解
代謝物アノテーションの改善
GUI化
代謝物プロファイルに適したパスウェイ可視化
Cytoscape
Galaxy
Shiny
by RStudio

まとめ

- 本研究成果パイプラインは代謝物プロファイルの実験間の詳細な可視化を可能にする(時系列など)
  - しかしながらメタボローム においてはデータベース統合環境の改善が必要

- 代謝物アノテーション改善のためのRDFを用いたデータベース統合の必要性を示した
  - 原子の由来に基づいた生合成経路オントロジーの土台を提示

謝辞(敬称略)

- 福島敦史 (理研CSRS)
- 時松敏明 (DBCLS)

- 大野圭一朗 (UCSD)
- 瀬々潤 (AIST)

- NBDC/DBCLS biohackathon