MetaboBank - メタボロームリポジトリの設計と構想
理研植物メタボロームメタデータベースの開発

○有田正規[1)]、金谷重彦[2)]、櫻井望[1)]、平川英樹[3)]、福島敦史[4)]

1）情報・システム研究機構 国立遺伝学研究所 生命情報・DDBJセンター　2）奈良先端科学技術大学院大学 情報科学領域
3）かずさDNA研究所　4）理化学研究所 環境資源科学研究センター

Keywords: Plant metabolism, Metabolomics, GC-MS, and Data sharing.

ABSTRACT

これまでメタボロームデータの公共リポジトリとして、欧州のMetaboLights、米国の Metabolomics Workbenchが運用されてきた。しかし、いまだメタボロームデータの再解析は困難であり、特に他オミックス分野との橋渡しは未完である。メタボロームデータの適切な解釈には、細胞、組織、生物種を超えて物質が循環する経路を示したマップ、"メタ代謝マップ"の構築が必要だろう。また、世界中の研究者に使いやすいマススペクトルおよびメタボロームデータを提供するには適切なメタデータが必要となる。このため、本プロジェクトでは、恒久的なメタボロームデータ・リポジトリ（MetaboBank）を設計、開発している。現在は初期データとして、かずさDNA研究所および理化学研究所環境資源科学研究センターが保有するデータを、**理研植物メタボロームメタデータベース**（**http://metabobank.riken.jp**/）へ登録、公開している。メタデータの整理は、ウェブ国際標準規格に沿ったRDF形式*（国際メタボロミクス学会の標準MSIに準拠）でおこなった。

* RIKEN Plant Metabolomics Metadata (RPMM)
→ https://github.com/afukushima/rpmm-metadata

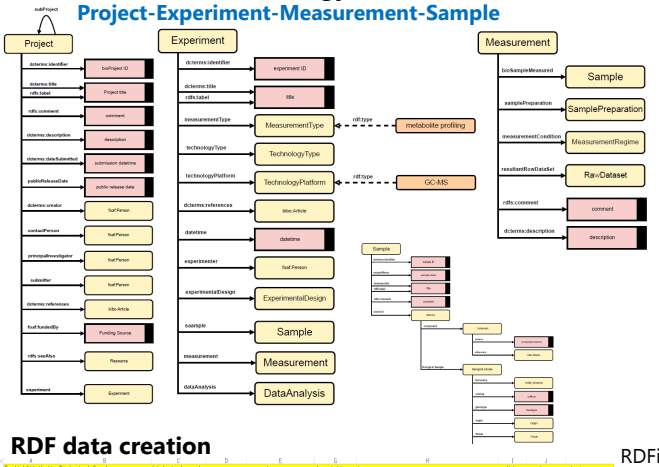
GitHub
afukushima/rpmm-metadata
Contribute to afukushima/rpmm-metadata development by creating an account on GitHub.

RIKEN Plant Metabolome MetaDatabase
http://metabobank.riken.jp/

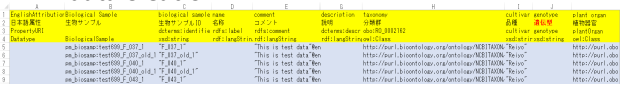
DATABASE DESCRIPTION

Plant Metabolomics Ontology

Project-Experiment-Measurement-Sample



RDF data creation

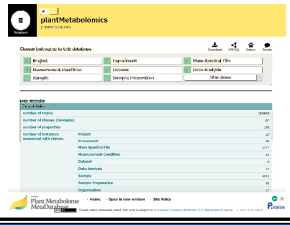


RDFising

An example of spreadsheet describing RIKEN PMM's "BiologicalSample" class.

RIKEN PMM

- is implemented on top of the RIKEN MetaDatabase [2]
- provides intuitive and interactive operations for plant metabolome data, including raw data, mass spectra and metabolite annotations
- is compliant with **the Metabolomics Standards Initiative (MSI)** [3]



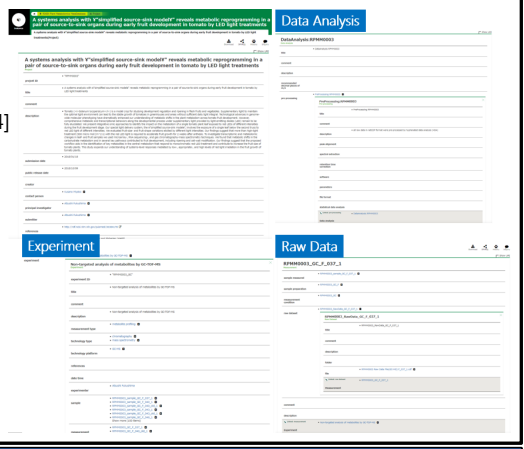


Fig. 1. A screenshot compilation of several RIKEN PMM data display by card forms. All metabolite data (*.netCDF format) are also downloadable from MetaboLights [4] (accession no. **MTBLS699**).

DISCUSSION AND CONCLUSION

RIKEN PMM contains cross-species metabolome (meta)data measured by GC-MS metabolomic approaches. It implements intuitive and interactive operations for plant metabolome data, including raw data (netCDF format files), mass spectra (NIST MSP format), and metabolite annotations. As a test data, we present >50 curated studies, which are publicly visible. These studies encompass several experimental and analytical protocols for >**5k biological samples with raw data files** and span over 6 different types of organism including *A. thaliana*, rice and tomato. We also present several hundreds of mass spectra of the metabolites identified in these studies.

Unlike other general or specific metabolome databases, all metadata in the RIKEN PMM, including Project, Experiment, Measurement and Sample, are expressed with controlled ontologies and vocabularies and are available in RDF format, resulting in machine-readable data exchange. It also provides a standardized API to view data as a SPARQL endpoint. This can facilitate rapid linking of our metabolome data to a large variety of other omics data, such as genomes, transcriptomes and proteomes. Data re-analysis is very important for systematic metabolite annotation for extracting novel information on metabolomic accumulation patterns in response to various genetic and/or environmental factors. Therefore, RIKEN PMM is beneficial not only for scientists who are interested in metabolomic phenotypes but also for researchers who would like to investigate plant metabolomic approaches.

FUTURE STUDIES

In the future, we plan to develop:

- An easy-to-use spreadsheet software that rapidly generates RDF data used in RIKEN PMM
- Easy-to-understand online help, instruction and tutorial sections

REFERENCES

1. Fukushima A and Kusano M, Front Plant Sci 4:73 (2013)
2. Kobayashi N et al., IJSWIS 14:140-164 (2018)
3. Sumner LW et al. Metabolomics 3:211-221 (2007)
4. Haug et al. NAR 41:D781-786 (2013)

ACKNOWLEDGEMENTS
This research was supported by the National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST)