RIKEN Plant Metabolome MetaDatabase: an integrated plant metabolome data repository based on the semantic web

Atsushi Fukushima[1], Mikiko Takahashi[1], Nozomu Sakurai[3,2], Toshiaki Tokimatsu[3], Hideki Nagasaki[2], Hideki Hirakawa[2], Takeshi Ara[4], Masanori Arita[3,1], and Norio Kobayashi[5,1]

[1] RIKEN Center for Sustainable Resource Science, RIKEN
1-7-22 Suehiro,Yokohama, Kanagawa, 230-0045, Japan
{atsushi.fukushima, mikiko.takahashi}@riken.jp
[2] Kazusa DNA Research Institute
2-6-7 Kazusa-kamatari, Kisarazu, Chiba, 292-0818, Japan
{sakurai, nagasaki, hh}@kazusa.or.jp
[3] DDBJ Center, National Institute of Genetics
1111 Yata, Mishima, Shizuoka, 411-8540, Japan
{sakurai, tokimatu, arita}@nig.ac.jp
[4] Graduate School of Agriculture, Kyoto University
Gokasho, Uji, Kyoto, 611-0011, Japan
ara@kais.kyoto-u.ac.jp
[5] Head Office for Inrormation Systems and Cybersecurity (ISC), RIKEN,
2-1 Hirosawa, Wako, Saitama, 351-0198 Japan
norio.kobayashi@riken.jp

Abstract. We developed the RIKEN Plant Metabolome MetaDatabase (RIKEN PMM), which stores gas chromatography-mass spectrometry-based (i.e. GC-MS-based) metabolite profiling data of plants together with their detailed experimental metadata, including sampling and experimental procedures. Our metadata are described using the Resource Description Framework (RDF) and standardised vocabularies, such as the Metabolomics Standardisation Initiative Application Ontology (MSI-AO), which are to be integrated with various life and biomedical science data on the World Wide Web. The RIKEN PMM implements intuitive and interactive operations for plant metabolome data, including raw data, mass spectra and metabolite annotations. The RIKNE PMM is suitable not only for scientists who are interested in metabolomic phenotypes but also for researchers who would like to investigate plant metabolomic approaches.

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

1 Introduction

Semantic web technologies based on the Resource Description Framework (RDF) are powerful tools for realising the integration of distributed global data. We

have developed a semantic web-based system called the RIKEN MetaDatabase[1], which encourages researchers to participate in the RDF-based distributed global data integration. Metabolome data provide new opportunities to gain a deeper understanding of plant metabolism relevant to both, plant and human health benefits. Although some major public repositories for general metabolomics data have been launched over the past decade, the data sharing aspect is hardly complete, especially for improving reanalysis, reusability and reproducibility. In this study, we have developed the RIKEN Plant Metabolome MetaDatabase (RIKEN PMM)[2], which stores gas chromatography-mass spectrometry-based (i.e. GC-MS-based) metabolite profiling data of plants together with their detailed experimental metadata.

2 Implementation and design

We have developed a novel ontology, called Plant Metabolomics Ontology, to describe our metabolome data. This ontology takes over the DNA Data Bank of Japan (DDBJ) data structure, including the BioSample database, and is extended with additional classes, such as experimental condition and data analysis. These additional classes describe the concepts of metabolomics for GC-MS and statistical data analysis. The metabolomic part of the Plant Metabolomics Ontology is designed to realise metadata interoperability according to the recommendation of the Metabolomics Standards Initiative (MSI[3]). The RIKEN PMM is implemented on top of the RIKEN MetaDatabase, which provides a biologist-friendly graphical user interface (commonly known as GUI), including tabular and card forms that show classes and instances of RDF datasets simultaneously as well as functions using a SPARQL endpoint as the application programming interface.

3 Discussion and future studies

We have developed a semantic web-based repository database known as the RIKEN PMM that enables efficient sharing, spreading and retrieving of plant metabolome data using native RDF technologies. Furthermore, we develop an easy-to-use spreadsheet software to rapidly generate RDF data for the RIKEN PMM. Our framework can provide all the metadata required for reanalysing and reusing metabolomics data, and will contribute to the development of other general-purpose metabolomics repository. This framework also realises data integration among different disciplines. The data sharing aspect discussed in this study will pave way for findable, reproducible and reusable metabolomics data as well as the robust interpretation of plant metabolomics data.

Acknowledgements
This research is supported by the National Bioscience Database Center (NBDC), Japan Science and Technology Agency (JST) and JSPS KAKENHI Grant Numbers 17K00434, 17H03621 and 17K00424.

[1] `http://metadb.riken.jp/`
[2] `http://metabobank.riken.jp/`
[3] `http://www.metabolomics-msi.org/`