The guide of integrated plant phenome databases
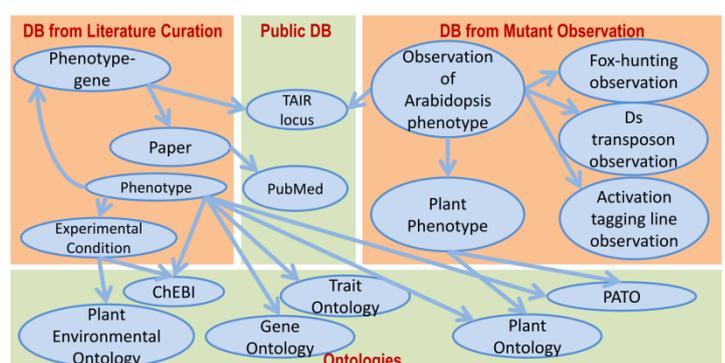
Contents

1. Introduction

A phenome represents the combination of various kinds of characteristics of organisms, expressed in response to genomic and environmental factors and observed by certain experimental methods. It is an important resource for future progress in functional genomics. To study relationships between phenotypes and other omic information collected from multiple databases, it is desirable to standardize phenotype terminology with common ontologies, and to standardize expression of relationships between phenomic data and other omics information.

We integrated phenome data of the *Arabidopsis* into our original database platform named RIKENBASE (http://database.riken.jp). It is based on Semantic Web technology, and supports some unique functions, including the PosMed high-speed semantic searching system, which facilitates bioinformatics analyses of integrated omics datasets.

2. Outline of database contents

RIKENBASE includes two major databases of Arabidopsis phenome. Items on these databases have links to various kind of omics and literature data in public databases, and users can easily collect related information of the focused topic (see figure).

Here the abstracts of these databases related to *Arabidopsis* phenome are shown. Some simple examples of searching bio-items using these contents are presented in the next section.

2.1 Phenome Integration of *Arabidopsis* mutants

- URL: http://database.riken.jp/db/arabidopsis_mutant_phenome
- Contents:
  - Updated phenome data of *Arabidopsis* mutants developed in RIKEN
    - Phenotypes of each mutant are shown in the standardized form with terms of Plant Ontology and PATO, such as "rosette leaf:round".
    - Genetic variation of each mutant

Remarks: RIKEN has contributed greatly to progress in the study of plant omics. Remarkable work includes the development of bioresources covering genomic and phenomic data for large-scale collections of *Arabidopsis thaliana* mutant lines, namely Ds transposon lines, activation tagging lines, and FOX-hunting lines. The original databases of these mutant lines were developed and opened separately by developers of each mutant line. RIKEN Phenome Integration of Arabidopsis Mutants at http://database.riken.jp/db/arabidopsis_mutant_phenome covers updated phenome data of these mutants. To cancel inconsistency of existing descriptions of phenotypes, data instances are expressed in principle with a simple format called Entity-Quality (EQ) formalism. Data standardization and construction of RDF data were achieved by the RIKENBASE development team.

2.2 The Database of Curated Plant Phenome

- URL: http://database.riken.jp/db/curated_plant_phenome
- Contents:
  - Standardized phenotypes of *Arabidopsis thaliana* collected via literature curation of published scientific papers.
  - Relationships between phenotypic and genetic variation (i.e. standardized phenotypes and AGI code).
  - Environmental factors in observing phenotypes.
  - Reference data of evidence papers for data mentioned above.
- Remarks:
  - We picked up over 820 relationships of phenotypic variation and genes by careful reading of published papers. Various types of phenotypes, such as morphological variation, stress response, seed productivity etc., are covered.
  - Each phenotype listed in the former step was carefully interpreted and standardized with ontology terms. We also took a serious view of environmental factors with which phenotypes were observed. They were also standardized and stored in the presented database. Ontologies used for standardization are listed in the following table.

Ontology	Covered field	Example	Reference
Plant Ontology (PO)	organ	flower	Cooper et al. 2013
	developmental stage	flowering	
Plant Trait Ontology (TO)	developmental stage	flowering time	Youens-Clark et al. 2011
The ontology of phenotypic qualities (PATO)	physical quality	increased length	Mungall et al. 2010
Gene Ontology (GO)	biological process	nitrogen fixation	Ashburner et al. 2000
	cellular component	chloroplast	
	molecular function	glycerate kinase activity	
Plant Environment Ontology (EO)	environmental condition	long day length regimen	Jaiswal et al. 2002
Chemical Entities of Biological Interest (ChEBI)	chemical compound	sugar	Hastings et al. 2013

3. Quick Start: How to search for bio items in plant phenome databases

Here shows some simple example to find notable bioitems such as genes and mutants using the presented databases.

3.1 Search genes involved in focused phenotypes

1. Access to the "Database of Curated Plant Phenome" (http://database.riken.jp/db/curated_plant_phenome). Click "Phenotype abnormality" under top menu bar.
2. Enter keyword (ex: "delayed flowering") in search box.

3. Click a search button, and a list of hit phenotypes will appear at the lower part of the page (See upper figure).

4. Click a link (Ex: "delayed whole plant flowering stage in environment of long day length regimen"), and a new table will appear which includes information about phenotype-gene interaction, derived from literature curation (See upper figure).

5. Click a "phenotype-gene" item (Ex: "delayed whole plant flowering stage in environment of long day length regimen for AT1G26830"), and detailed reference data will appears. It includes links to PubMed, with which users can access original description to check validity of search results.

3.2 Search bioresources (mutants) from focused phenotype

It is possible to search *Arabidopsis* mutants from keywords related to focused phenotype, by the following steps.

1. Access to "RIKEN Phenome Integration of *Arabidopsis* mutants" (http://database.riken.jp/db/arabidopsis_mutant_phenome).

2. Specify keywords using the menu at the center of the page. This menu allows to specify organ names or　growth stage. Users can see detailed categories by clicking "Show subcategories" button. For example, "flower:abnormal,shape" is selected in this text. Alternatively, users can enter the keywords directly in the search box at the top of this page.

3. The new page includes observation data of *Arabidopsis* mutant of which "observed phenotype" is "flower:abnormal,shape".

observation of Arabidopsis phenotype (RIKEN Phenome Integration of Arabidopsis mutants)
F06248_obs_5 (FOX hunting)   [Browse list of all records]

Original phenome observation data	F06248_obs_5 Hide Detail Search Related Information	
	hasPhenotypicData	flower:others Show Detail Search Related Information
	isObservationForGeneration	T1 Show Detail Search Related Information
	hasLineData	F06248 Show Detail Search Related Information
	observation of Arabidopsis phenotype ("Original phenome observation data" is F06248_obs_5)	F06248_obs_5 (FOX hunting) Search Related Information
observed phenotype	flower:abnormal,shape Hide Detail Search Related Information	
	quality in PATO	abnormal, shape
	entity in PO	flower Search Related Information
	observation of Arabidopsis phenotype ("observed phenotype" is flower:abnormal,sha...)	106 subjects Search Related Information
	FOX line ("observed phenotype" is flower:abnormal,sha...)	41 subjects Search Related Information
	Ds tagging line ("observed phenotype" is flower:abnormal,sha...)	16-0443-1, 15-3580-1, 15-2793-1 Show Detail Search Related Information

4. Click "Show Detail" button of "Original phenome observation data", and original information opened by developers of mutants. Furthermore, "Show Detail" of "observed phenotype" is linked to information of other mutant lines showing same phenotypes. In the presented example, users can find that 41 Fox line and 3 Ds tagging line mutants share a phenotype, "morphology of flower is abnormal".

The presented example shows that users effectively can find useful mutants for their research for the focused phenotype, by cross-database search of "Ds tagging", "Activation tagging" and "FOX-hunting" mutants.

4. Advanced search for improved bioinformatic studies

RIKENBASE integrates databases not only of phenome but of various kinds of omics.

4.1 Gene-phenotype search started from TAIR locus DB: Reasoning of unknown roles or functions of focused gene

Here we show an example of search using The Database of Plant Phenome and the Database of *A. thaliana* locus, by which users can deepen knowledge about biological roles or functions of focused *Arabidopsis* genes.

The first example is about genes related to "abscisic acid beta-glucosidase" as a keyword. Abscisic acid is a well-known plant hormone involved in drought stress, and abscisic acid beta-glucosidase has an important role for biosynthesis of abscisic acid. Information of related genes are covered by TAIR locus database, which is integrated in RIKENBASE as well as the Database of Curated Plant Phenome.

1. Access to the TAIR locus database (https://database.riken.jp/sw/en/TAIR10_Locus/crib102u0i/).
2. Enter "abscisic acid beta-glucosidase" as keyword into search box at the top of the page (See Figure below).

3. The result suggests that the gene AT1G52400 is involved in production of this enzyme (See upper figure).
4. Click the AT1G52400 link in the result list.

5. Phenome collected by literature curation is shown in the column labeled "phenotype-gene" (Figure below).

6. Click "Show detail" button.

7. A new table appears. Relationships between phenotypes and genes with reference data are listed in this table (Figure below).

phenotype-gene	annotation	has_mutation	outcome_paper
non-functional stomatal movement for AT1G52400	non-functional stomatal movement	AT1G52400	Lee Kwang Hee et al. 2006 Sep. Cell 126(6):1109-2...
premature during process named seed germination f...	premature during process named seed germination	AT1G52400	Lee Kwang Hee et al. 2006 Sep. Cell 126(6):1109-2...

Click texts in "outcome_paper" to view another table showing further detailed reference data including abstract. This table also includes a link to PubMed as "Original site", where users can easily check validity of shown phenotype-gene relationships based on original descriptions, figures etc.

This example shows that users can search a notable gene closely related to drought tolerance. It is also shown that the focused gene participates in stomatal movement. In the same manner, it is possible to find new research theme by focusing unknown biological role of genes.