Development of a Platform for Functional Genomics under the Generation Challenge Programme (GCP)

Masaru Takeya1  
katu@affrc.go.jp

Kohji Suzuki2  
qxzsuzuk@bank.dna.affrc.go.jp

Koji Doi1  
kdoi@nias.affrc.go.jp

Shoshi Kikuchi1  
skikuchi@nias.affrc.go.jp

Richard Bruskiewich3  
r.bruskiewich@cgiar.org

1 National Institute of Agrobiological Sciences, 2-1-2 Kannondai Tsukuba, Ibaraki 305-8602, Japan
2 Hitachi Software Engineering Co., Ltd, 4-12-7, Higashishinagawa, Shinagawa-ku, Tokyo 140-0002, Japan
3 International Rice Research Institute, DAPO Box 7777, Metro Manila, Philippines

Keywords: domain model, database, microarray, cis-element, web services

1 Introduction

The Generation Challenge Programme (GCP) [7] is an international consortium of the Consultative Group on International Agricultural Research (CGIAR) centers, five advanced research institutes (ARIs), and a number of national research systems (NARS) of developing countries. By using advances in molecular biology and harnessing the rich global stocks of crop genetic resources, the GCP aims to contribute to a new generation of plants that will meet the needs of resource-poor farmers. Drought tolerance was chosen as a long-term case study in the GCP.

One objective of GCP is to create “an integrated platform” of biological data source and data analysis tools for researchers and breeders. A comprehensive common scientific domain model has been developed to describe the structure and relationships among the data. We have been developing the web services applications of BioMOBY to exchange functional genomics data as XML. The applications in the GCP are described by using Java language. The source codes of applications are posted in CropForge [6].

2 Method and Results

2.1 Domain Model

The GCP domain model [1] is documented in Unified Modeling Language (UML). Computable versions of the UML model are published in the GCP “Pantheon” middleware project in CropForge. The model consists of a generic metadata model and specific subdomains. This metadata model defines the general concept of a system “Entity” with “Identifier” and “Feature” components. The specific subdomains denote germplasm, passport, phenotype, genotype, mapping, geographic information system (GIS), and functional genomics. The functional genomics subdomain involves gene expression and proteomics components.

2.2 Gene Expression Database

The microarray database RED [3] and the full-length cDNA database KOME [2] were initially developed for rice. A crop gene expression database is being developed using the open-source database MAXD [5] for genomic expression data. MAXD has been customized for documenting crop study information using MIAME/Plant [4]. The customized MAXD will be installed in GCP institutes for the exchange of crop gene expression data.

Adapter programs are being developed for handling the data in RED and KOME. The data are exchanged to the BioMOBY data type in class modules. Linkage of the developed web services enables researchers to identify gene expression datasets of corresponding reporter genes from interested Gene Ontology term. The schema of the relationship between application and data source is shown in Figure 1.
2.3 Data Mining Tools

We have established a pipeline system for promoter analysis. The pipeline system is constructed to list cis-element candidates particularly corresponding to user-defined gene list, using information on the upstream sequences of genes. The frequency of appearance of motifs in the upstream sequences of the listed genes is counted and then compared with those of all the genes in the KOME database. The system consists of several small perl scripts. One of these called MEME, which is an open software for motif searches. Another script is used to perform association rule analyses to evaluate the specificity of the cis-element candidates listed by MEME and filter these candidates by their list index values. Those motifs showing good values (list > 1.0) are listed as the final cis-element candidates.

3 Discussion

The functional genomics platform successfully gives researchers efficient data transfer in GCP. Our next steps will be to create and provide ontology for the crop data sources in the domain model and to improve and extend the web services and data mining tools.

References


[5] [http://bioinf.man.ac.uk/microarray/maxd/](http://bioinf.man.ac.uk/microarray/maxd/)

[6] [https://cropforge.org/](https://cropforge.org/)

[7] [http://www.generationcp.org](http://www.generationcp.org)