

An Integrated Biotechnology Database of the National Agricultural Information Center in Korea

Chang Kug Kim, Dong Suk Park, Young Joo Seol, Jang Ho Hahn

Abstract—The National Agricultural Biotechnology Information Center (NABIC) plays a leading role in the biotechnology information database for agricultural plants in Korea. Since 2002, we have concentrated on functional genomics of major crops, building an integrated biotechnology database for agro-biotech information that focuses on bioinformatics of major agricultural resources such as rice, Chinese cabbage, and microorganisms. In the NABIC, integration-based biotechnology database provides useful information through a user-friendly web interface that allows analysis of genome infrastructure, multiple plants, microbial resources, and living modified organisms.

Keywords—biotechnology, database, genome information

I. INTRODUCTION

THE agricultural biotechnology information research has proceeded by bioinformatics center through various genome projects and specific databases have constructed to organize projects based on the long-linear model from DNA sequence to proteomics. The National Agricultural Biotechnology Information Center (NABIC) was established by the National Academy of Agricultural Science (NAAS), in 2002. The NABIC has constructed agricultural biology-based infrastructure and provided the biological information for agricultural research in Korea. Major functions are being focused on biotechnology development for agricultural bioinformatics and has provided integrated biotechnology database which consist of rice genome, Chinese cabbage (*Brassica rapa*) genome, microbes genome, rice mutants, allergen information, genetic resources, and agricultural bio-safety information database. In 2011, we constructed an agricultural biology-based infrastructure and developed a biological integrated-based model database. The NABIC provides the integrated biotechnology information database using by updated resources of various field through developing of several bioinformatics tools. The major functions are focused on biotechnology development for agricultural bioinformatics and provide a Web-based service to easily construct bioinformatics workflows.

C.K. Kim is with the National Academy of Agricultural Science (NAAS), Rural Development Administration (RDA), Suwon 441-707, Republic of Korea (phone: +82-31-299-1655; fax: +82-31-299-1657; e-mail: chang@korea.kr).

D.S. Park is with the NAAS, RDA, Suwon 441-707, Korea (phone: +82-31-299-1699; fax: +82-31-299-1657; e-mail: dspark@rda.go.kr).

Y.J. Seol is with the NAAS, RDA, Suwon 441-707, Korea (phone: +82-31-299-1634; fax: +82-31-299-1657; e-mail: yjseol@rda.go.kr).

J.H. Hahn is with the NAAS, RDA, Suwon 441-707, Korea (phone: +82-31-299-1620; fax: +82-31-299-1657; e-mail: jhhahn@rda.go.kr).

We hope that this database will contribute to the agricultural bioinformatics research field to extend the usefulness of breeding for new crops.

II. METHODOLOGY

A. Data collection

The information was collected from the Korean rice genome project from the NAAS (<http://www.naas.go.kr/>), Korean rice Ds-tagging lines project (<http://nabic.naas.go.kr/RDS/>), Chinese cabbage project (<http://www.brassica-rapa.org/BGP/>), the allergen project (<http://nabic.naas.go.kr/allergen/>), the genetic resources project (<http://www.genebank.go.kr/>), the BG21 project (<http://atis.rda.go.kr/>), and from universities and various institutes in Korea.

B. Database design

The integration-based database (<http://nabic.naas.go.kr/>) is designed to provide information on the genome of agricultural crops. This database has six major categories which consist of genome research, gene expression, rice mutant database, analysis tools, genome annotation, and other databases. It could be accessed by using a web-based graphical view and anonymous users can query and browse the data using the various functions. This database has developed a portable system able to handle very large genomes and associated requirements for sequence analysis. The platform was developed using MySQL, commonly available network protocols such as Hypertext Transfer Protocol and JAVA language. The logical and physical scheme of the database followed the standard principle of relational database by ERWin Data Modeler software (<http://www.ca.com>). This data was stored in an Oracle relational database management system (Oracle Database 10g, Redwood, CA, USA, <http://www.oracle.com/>).

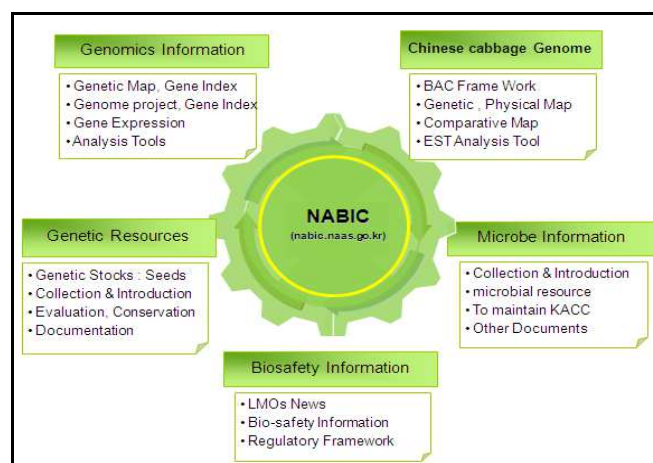


Fig. 1 Overview of relationships among five major groups in NABIC

C. Web service of database

The integrated biotechnology database provides not only simple text information on individual genome sequence but also analysis tables and genetic markers for annotation. It consists of multiple sub databases which provide information and analysis functions on wide bioinformatics fields. This database platform consists of multiple layer architecture which can be accessed using a web-based graphical interface and anonymous users can query and browse the data using various functions. To ensure software from various databases, service systems are developed by open standards protocol such as Simple Object Access Protocol (SOAP, <http://www.w3.org/TR/soap/>), an independent platform, commonly available network protocols as Hypertext Transfer Protocol (HTTP), and XML-based model.

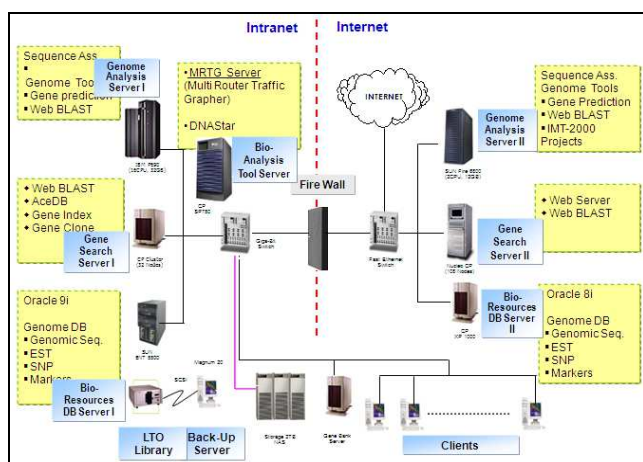


Fig. 2 Overview of the web-service network in NABIC. The database system architecture was designed using various open standard protocols and dual network

III. RESULTS

A. Genome information database

The Gnome information database consists of multiple sub systems such as genome project, gene expression, genetic map, genome annotation, and analysis tools. The genome project provides a bioinformatics framework to study biological function based on sequences of genomes of rice, Chinese cabbage, and microorganism. It is a source of annotation in genome sequence, physical map, sequence comparison, and gene prediction. The gene expression has presented an integrated web-based tool for automatic multistep analysis of gene expression data. The genetic map has three major functions which is possible to detect specific markers on each of the 12 chromosomes and the information is presented with the associated trait and a mapped locus on a genetic map. The analysis tools have seven major functions such as web blast search, gene finding, transcription factor analysis, repeat sequence searching, gene marker searching, and protein-structure analysis.

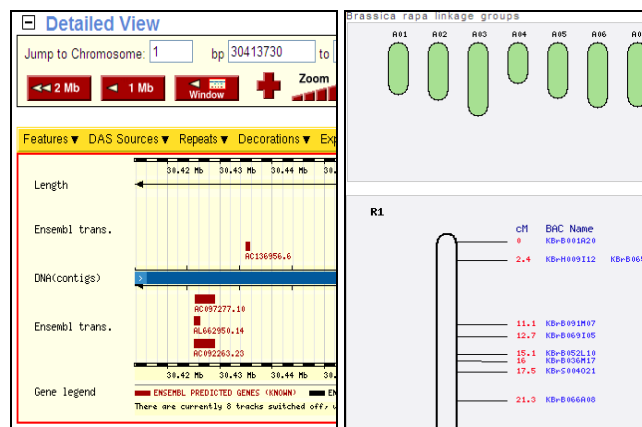


Fig. 3 Screenshot of genetic map in rice (left) and Chinese cabbage (right). This detailed view has a hyper-text to allow the user to select the features being displayed

B. The Korean rice Ds-tagging lines database

The Korean Rice Ds-tagging lines Database (KRDD) is designed to provide information about Ac/Ds insertion lines and activation tagging lines using japonica rice. This database has provided information on 18,158 Ds lines, which includes the ID, description, photo image, sequence information, and gene characteristics. The KRDD is visualized using a web-based graphical view, and anonymous users can query and browse the data using the search function. It has four major menus of web pages such as blast search, primer design tool to identify genotypes, phenotype for Ds lines, searching by identification name and phenotype characteristics.

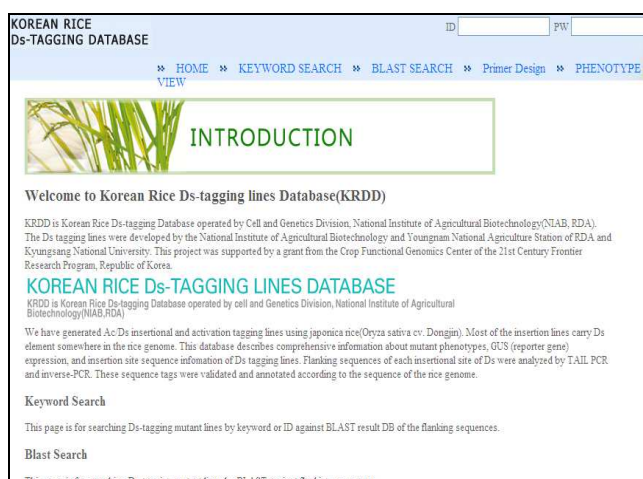


Fig. 4 Screenshot of the KRDD. This database shows comprehensive information about rice mutant for Ds tagging lines

C. Allergen database

The allergen database (<http://nabic.naas.go.kr/allergen/>) has developed a web-based system that will provide information about allergen in microbes, animals and plants. The database has three major parts and functions such as database list, allergen search, and allergenicity prediction. There are 2,434 allergen information readily available in the database such as on allergens in rice microbes, animals and plants [1]. Furthermore,

this database provides bioinformatics tools for allergenicity prediction. Users can search for specific allergens by various methods and can run tools for allergenicity prediction using three different methods.

D. National genebank

The national genebank is the conservation center for crop and microbial genetic resources in Korea. The genetic resources (<http://www.genebank.go.kr/>) database provides the information which collected from domestic and exotic plants, insects, and microbial species. This database has four major menus to search for information, such as an accession number, taxonomy, collecting information, and breeding information. In the database, there are four major menus to search for information, such as an accession number, taxonomy (scientific or common name), collecting information, and breeding information. It services a function to secure on the registration and distribution, and analysis platform to provide the information for various resources [2].

E. Korean agricultural bio-safety database

The Korean Agricultural Biosafety Information Center performs the review process for the approval of genetically engineered agricultural products. We provide to information in the international cooperation, assessment of environmental risks for national policy, guidelines for environmental risk assessment, and standard protocol for agricultural crops. The bio-safety (<http://biosafety.rda.go.kr/>) database provides living modified organisms (LMOs). This database available on this site covers genetically engineered plants intended for food that have completed required reviews for food, feed or planting use in the Korea. The center provides information on LMOs, legislation of GMOs, bio-safety examination, support the national policy, and assessment information of environmental risk.

IV. DISCUSSION AND FUTURE WORK

The NABIC was established in 2002 with the main objectives analyzing genome information of agricultural crops, and provides related services to professional genomic research institutes and societies [8]. The integration-based database provides information through a user-friendly Web-interface from searching gene sequences to genome metabolism infrastructure analysis. In addition, this database provides genome information which includes various genome projects, genetic map, gene numbers, genetic marker, gene index, metabolic pathway, systems biology, specific gene sequence, and functional genetic tables for annotation in agricultural crops. The integration biotechnology database provides information through a user-friendly web interface and can approach to agricultural biotechnology to extend the usefulness of breeding for agricultural plants. We have updated the additional information from international public sources and develop a useful tool for functional gene analysis. In the future, NAAS plans to develop an integrated biotechnology database

that combines genome and gene functional information in the major crops. In addition, we will provide an integration-based database and bioinformatics tools to solve complex biological problems with next-generation sequencing technologies.

ACKNOWLEDGEMENT

This work was conducted with support from the “Research Program for Agricultural Science & Technology Development (Project No. PJ006651)”, National Academy of Agricultural Science, Rural Development Administration, Republic of Korea.

REFERENCES

- [1] C.K. Kim, S.J. Kwon, and J.H. Hahn, “ A database for allergenic proteins and tools for allergenicity prediction,” *Bioinformation.*, vol. 3, Aug. 2009, pp. 344-345.
- [2] NAAS, NAAS Annual report 2010, National Academy of Agricultural Science, RDA, Korea, 2010.