An online database for genome information of agricultural plants

ChangKug Kim, DongSuk Park, YoungJoo Seol, UngHan Yoon, GangSeob Lee & JangHo Hahn*

Genomics Division, National Academy of Agricultural Science (NAAS), Suwon 441-707, Korea; JangHo Hahn – Email: jhhahn@rda.go.kr; *Corresponding author

Abstract:
The integration-based genome database provides useful information through a user-friendly web interface that allows analysis of comparative genome for agricultural plants. We have concentrated on the functional bioinformatics of major agricultural resources, such as rice, Chinese cabbage, rice mutant lines, and microorganisms. The major functions are focused on functional genome analysis, including genome projects, gene expression analysis, gene markers with genetic map, analysis tools for comparative genome structure, and genome annotation in agricultural plants.

Availability: The database is available for free at http://nabic.naas.go.kr/

Keywords: Genome information, Agricultural plants, Genome database

Background:
The genome information research of agricultural plants has performed through genome projects and integrated databases have constructed to organize projects based on the information-based database from simple sequence to specific metabolism [1]. In Korea, the National Agricultural Biotechnology Information Center (NABIC) has constructed agricultural biotechnology-based infrastructure and provided the integrated genome information for agricultural plants [2, 3]. In 2012, we developed a biological integrated-based genome database for agricultural plants. Our database is being focused on genome analysis tools development for agricultural bioinformatics, the number of which is now growing rapidly. In addition, this database has provided integrated genome information for comparative analysis of sequenced genomes which consist of various agricultural plants. We hope that this database will contribute to the genome research field to extend the usefulness for agricultural hybrid-plants breeding.
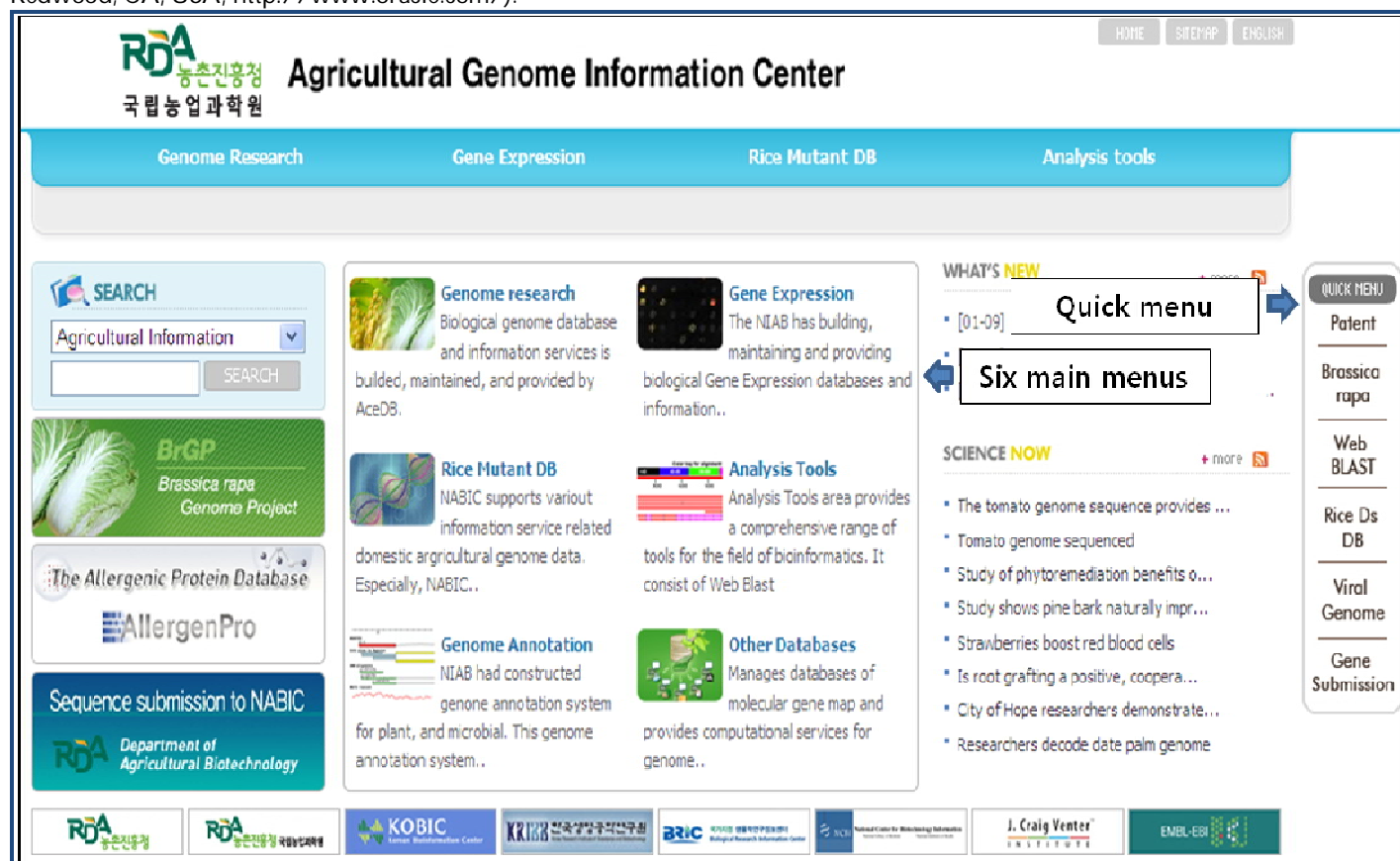
Methodology:
Data collection
The genome information was collected from the genome project from the National Academy of Agricultural Science [4] and the New Bio-Green 21 project [5]. In addition, genomic information was collected through international public institutes and collaborative institutes.

Database development:
This database is designed to provide information on the genome of agricultural plants and the platform consists of multiple layer architecture which can be accessed using a web-based graphical interface. It has six major categories, such as genome research, gene expression, rice mutant, analysis tools, genome annotation, and other sub-databases. The platform was developed using MYSQL, JAVA language, and commonly available network protocols as hypertext transfer protocol, and XML-based model. The collected data were stored in an Oracle

BIOINFORMATION

relational database management system (Oracle Database 10g, Redwood, CA, USA, http://www.oracle.com/).

Figure 1: The screenshot of the genome information database, which shows the six information categories for genomes of agricultural plants.

Implementation and Features:
This genome information database consists of multiple sub systems such as genome research, gene expression, rice mutant, analysis tools, genome annotation, and other sub-databases **(Figure 1)**. The genome research provides a bioinformatics framework to study biological function based on sequences of genomes of rice (*Oryza sativa*), Chinese cabbage (*Brassica rapa*), microbes, and viral genomes. In addition, it provides the functions which as genome annotation, sequence comparison with molecular markers, and gene prediction with genetic map. The gene expression provides a web-based tool for multistep analysis of gene expression data **[6]**. This menu has two major categories which are possible to detect specific markers and gene clone which mapped locus on a genetic map. It has related information on expressed sequence tags for 10 species: rice, wheat, maize, soybean, barley, Chinese cabbage, tomato, hot pepper, mushroom, and Arabidopsis **[7].** The rice mutant provides comprehensive information about Ds mutant phenotypes and insertion site sequence information of rice mutant lines. This menu has four major categories, such as Blast for mutant lines, search of phenotype, primer design tool to identify genotypes, and management menu. The analysis tools have eight major functions such as blast search of sequences, gene finding program, transcription/motif analysis, multiple repeat sequence, and genetic marker analysis. Finally, genome annotation provides specific genomic analysis through four different view panels. This menu shows relationships between the basic sequence and annotated gene information. The user can access information about individual genes along with functional annotations within the selected chromosome and can obtain new information using comparative genomics methods. In addition, this menu provides improved comparative annotation from more revised genome information based on progress made by Chinese cabbage genome project **[8]**.

Discussion and Future work:
The NABIC was established with the main objective of analyzing the genomic information of agricultural plants and provides related services to genomic institutes and universities. The integrated-based database consists of multiple subsystems and provides a bioinformatics framework to study biological function based on the genomic sequences of rice, Chinese cabbage, rice mutant lines, and microorganisms. This database provides information through a user-friendly Web-interface from searching gene sequences to genome metabolism infrastructure analysis. In the 2013, NABIC plans to develop an upgraded genome database that combines genome information and related metabolism pathway in the major agricultural plants. In addition, we will provide an upgraded database to solve specific biological pathways with next-generation sequencing information.

BIOINFORMATION

open access

References:
[1] Skinner ME *et al. Genome Res.* 2009 **19**: 1630 [PMID: 19570905]
[2] Kim CK *et al. Bioinformation.* 2011 **3**: 246 [PMID: 21887015]
[3] Kim CK *et al. Bioinformation.* 2009 **3**: 344 [PMID: 19707297]
[4] http://www.naas.go.kr/
[5] http://atis.rda.go.kr/
[6] Kim CK *et al. Bioinformation.* 2008 **3**: 61 [PMID: 19238232]
[7] Kim CK *et al. Bioinformation.* 2008 **2**: 344 [PMID: 18685722]
[8] http://www.brassica-rapa.org/

Edited by P Kangueane
Citation: Kim *et al. Bioinformation 8(21): 1059-1061 (2012)