The integrated web service and genome database for agricultural plants with biotechnology information

ChangKug Kim, DongSuk Park, YoungJoo Seol, 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 National Agricultural Biotechnology Information Center (NABIC) constructed an agricultural biology-based infrastructure and developed a Web based relational database for agricultural plants with biotechnology information. The NABIC has concentrated on functional genomics of major agricultural plants, building an integrated biotechnology database for agro-biotech information that focuses on genomics of major agricultural resources. This genome database provides annotated genome information from 1,039,823 records mapped to rice, Arabidopsis, and Chinese cabbage.

Keywords: genome database, biotechnology, web service, biotechnology information.

Background:
The genome information from humans to microorganisms is rapidly increasing in the 21st century. An integrated genome database provides a natural index for molecular biology and information for understanding biological data. To increase the importance of genome database, the various databases and browsers have been constructed [1]. The GRAMENE (http://www.gramene.org/) browses assembled genomes for Oryza sativa, using genomes browser [2]. The National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) provides analysis and retrieval resources for the data in GenBank, and also other biological data, made available through the NCBI Web site [3]. The *Brassica rapa* Genome Project (http://www.brassica-rapa.org/) provides the genome information for Chinese cabbage *Brassica rapa* using genome map browser [4]. In Korea, Web-based genomic databases have been developed with a knowledge-based approach for functional gene annotation, gene data mining, and genomic sequencing [5]. The NABIC (http://nabic.naas.go.kr/) has constructed agricultural biology-based infrastructure and provided comprehensive database with agricultural biotechnology. Major functions are focused on biotechnology development for agricultural plants with biotechnology information [6].

Methodology:
Data collection:
The biotechnology information on agricultural crops was collected from the International Rice Genome Sequencing Program (http://rgp.dna.affrc.go.jp/IRGSP/), the Korean rice genome project from the National Academy of Agricultural Science (http://www.naas.go.kr), the

Chinese cabbage project (http://www.brassica-rapa.org/BGP/), The Arabidopsis Information Resource (http://www.arabidopsis.org/), and from universities and various institutes in Korea. In addition, genome information was accumulated and collected through several collaborative institutes and public international institutes.

Database design:
The integrated biotechnology database is designed to provide information on the genome of agricultural crops. The database has two major categories. 'Over view' panel allows the user to identify the genome of interest. 'Detailed view' panel is designed for the detailed annotated information at the chromosome. The platform was developed using MYSQL, commonly available network protocols such as Hypertext Transfer Protocol and JAVA language, and the data was stored in an Oracle relational database management system (Oracle Database 10g, Redwood, CA, USA, http://www.oracle.com/).

Results and Discussion:
The NABIC has provided resources for bioinformatics through the development of several bioinformatics tools and the construction of integrated genome information systems. It can be accessed using a web-based graphical interface, and anonymous users can query and browse the data using various functions. The Web based relational database provides genome information from the web site (http://ensembl.naas.go.kr/Rice/) and includes not only simple text information on individual genome sequences, but also analysis tables and genetic information for annotation.
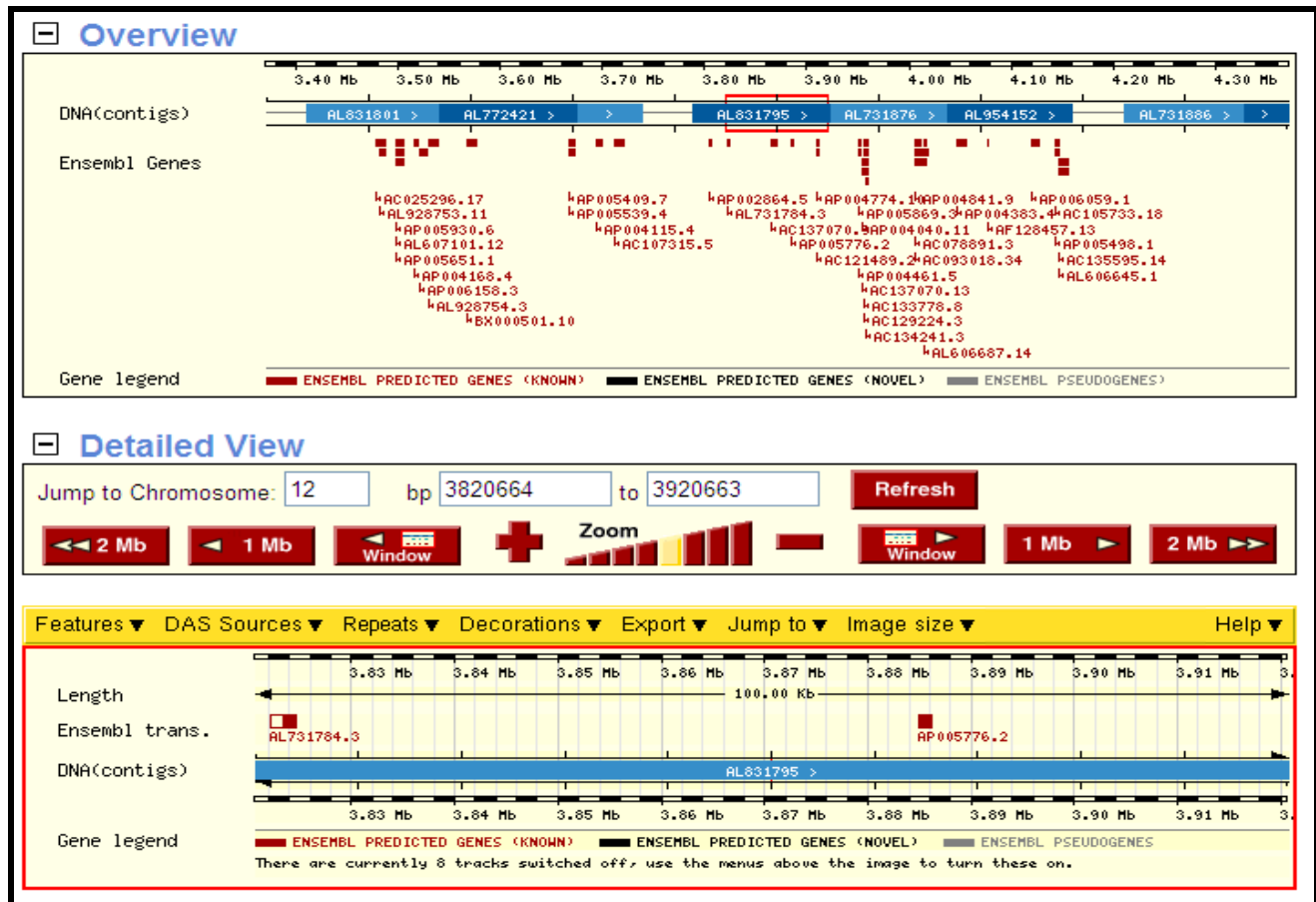
BIOINFORMATION

Figure 1: Screenshot of the two information view panel for rice. This window has a zoom function and pull-down menus to allow users to select the features displayed. Floating menus (not shown) allow users to access linked windows with additional information.

Genome browser for Web based relational database:
The genome database provides a bioinformatics framework to study biological function based on genomic sequences of rice, Arabidopsis, and Chinese cabbage. The database is a source of annotation for genome sequences, physical maps, sequence comparison, and gene prediction. The database has developed a portable system, capable of handling very large genomes and associated requirements for sequence analysis. To achieve scalability and consistency of annotation, we developed a browser based on a relational database for rice, Chinese cabbage, and microbes. To advance genome research, we constructed an integrated genome browser database for sequence analysis of rice (*Oryza sativa*), Arabidopsis (*Arabidopsis thaliana*), and Chinese cabbage (*Brassica rapa*) genomes. The genome browser database provides annotated genome information from 803,607; 201,419; and 34,797 records mapped to rice, Arabidopsis, and Chinese cabbage, respectively.

The rice browser provides specific genome analysis through two different view panels (**Figure 1**). This overview panel (accessible by clicking) shows the location of markers and genes. The user can access information about individual genes along with functional annotation within the entire chromosome. The detailed view panel shows genomic sequence features, and the chromosome bar depicts the chromosome banding region for selection. In the comparative genome analysis between Arabidopsis and *Brassica rapa* genomes, users can obtain new gene information resulting from comparative genomics methods and identify missing regions within a single genome.

The NABIC was established in 2002 with the main objective of analyzing genome information of agricultural crops, and provides related services to professional genomic research institutes and societies. The Web based relational database provides genome information including specific gene sequences, genome projects, gene identification numbers, gene location, and genetic information tables for annotation in agricultural plants. In the future, NABIC will provide a web service to easily construct bioinformatics workflows and pipelines combining two or more instructions to solve complex biological tasks such as protein function prediction, genome annotation and system biology.

References:
[1] Barrett T *et al. Nucleic Acids Res.* 2011 **39**: D1005 [PMID: 21097893]
[2] Youens-Clark K *et al. Nucleic Acids Res.* 2011 **39**: D1085 [PMID: 21076153]
[3] Sayers EW *et al. Nucleic Acids Res.* 2011 **39**: D38 [PMID: 21097890]
[4] Lee SC *et al. Mol Cells.* 2008 **26**(6): 595 [PMID: 18797175]
[5] Kim CK *et al. Bioinformation* 2008 **3**(2): 61 [PMID: 19238232]
[6] Kim CK *et al. Bioinformation* 2009 **4**(6):269 [PMID: 20975922]

Edited by P Kangueane
Citation: Kim *et al. Bioinformation 6(6): 246-247 (2011)