ARTICLE

Genomics & Informatics
Vol. 9(2) 85-88, June 2011

DOI: 10.5808/GI.2011.9.2.85
pISSN 1598-866X eISSN 2234-0742

An Information-based Model for an Interactive Web Service with Agricultural Biotechnology

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

Genomics Division, National Academy of Agricultural Science (NAAS), Suwon 441-707, Korea

Abstract

The National Agricultural Biotechnology Information Center (NABIC) constructed an agricultural biology-based infrastructure and developed a biological information-based database. The major functions of the NABIC are focused on biotechnological developments for agricultural bioinformatics and providing a web-based service to construct bioinformatics workflows easily, such as protein function prediction and genome systems biology programs. The NABIC has concentrated on the functional genomics of major crops, building an integrated biotechnology database for agro-biotech information that focuses on the proteomics of major agricultural resources, such as rice, Chinese cabbage, rice Ds-tagging lines, and microorganisms.

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

Introduction

The genomic information from humans to microorganisms is rapidly increasing in the 21st century. Today, technological advances in the fields of genomic sequencing and protein structure have led to the development of genomic, proteomic, microarray, and functional genomic data (Ann, 2008). Biological databases have been established to optimize scientific exploitation of the explosion of data within bioinformatics. Bioinformatics databases and tools provide analyzed results to understand the basic principles of molecular interactions and systemic functional behaviors of organisms (Russ, 2007). With the continuous elaboration of bioinformatics tools around the world, various databases have been constructed at bioinformatics centers with gene expression, genetic marker, microarray gene family, protein prediction, and functional genetic information (Kim *et al.*, 2010). The National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) provides analysis and retrieval resources for the data in GenBank and other biological data that are made available through the NCBI website (Sayers *et al.*, 2008). GRAMENE (http://www.gramene.org/) provides genomic information for *Oryza sativa* using a genomic browser (Youens-Clark *et al.*, 2008).

In Korea, biological information-based models have been developed with a knowledge-based approach for functional gene prediction, gene data-mining using microarray, and protein-protein interaction networks (Kim *et al.*, 2008). The Korean Bioinformation Center (http://www.kobic.re.kr/) is a national research center for bioinformatics that plays a key role in various areas, such as genomics, proteomics, systems biology, and personalized medicine. The Biology Research Information Center (http://bric.postech.ac.kr/) and Biotech Policy Research Center (http://www.bioin.or.kr/) support the government in setting up biotechnology information and all of the relevant information provided in the portal site. The NABIC (http://nabic.naas.go.kr/) has constructed an agricultural biology-based infrastructure and has provided comprehensive agricultural biological research information. Major functions are focused on biotechnological developments for agricultural bioinformatics.

Methods

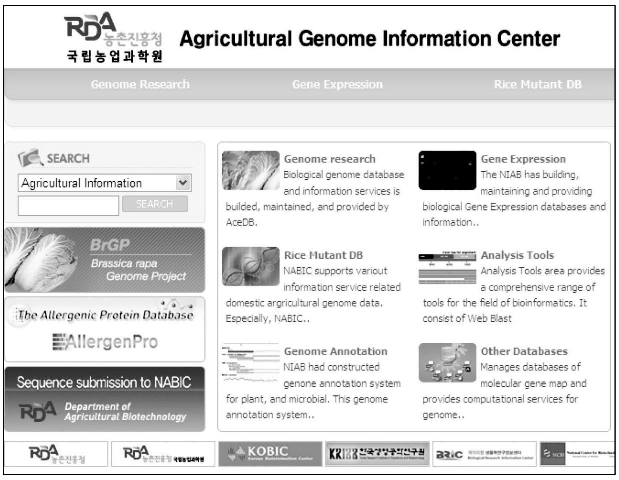
Data collection

The biotechnological information on agricultural crops was collected from the Korean rice genome project from the National Academy of Agricultural Science (NAAS, http://www.naas.go.kr/), the Chinese cabbage project (http://www.brassica-rapa.org/BGP/), the microorganism project (http://kacc.rda.go.kr/), the BG21 project (http://atis.rda.go.kr/), and universities and various institutes in Korea. In addition, genomic information was accumulated and collected through several collaborative institutes and public international institutes.

Model design

The integrated biotechnology database is designed to provide information on the genomes of agricultural crops. This database (http://nabic.naas.go.kr/) has six major

*Corresponding author: E-mail jhhahn@rda.go.kr
Tel +82-31-299-1620, Fax +82-31-299-1657
Accepted 20 May 2011


RDA 농촌진흥청
국립농업과학원
Agricultural Genome Information Center
Genome Research
Gene Expression
Rice Mutant DB
SEARCH
Agricultural Information
SEARCH
BrGP
Brassica rapa Genome Project
The Allergenic Protein Database
AllergenPro
Sequence submission to NABIC
RDA Department of Agricultural Biotechnology
RDA 농촌진흥청
RDA 농촌진흥청 국립농업과학원
KOBIC
KRIBB 한국생명공학연구원
BRIC
NCBI

Genome research
Biological genome database and information services is builded, maintained, and provided by AceDB.

Gene Expression
The NIAB has building, maintaining and providing biological Gene Expression databases and information..

Rice Mutant DB
NABIC supports variout information service related domestic argricultural genome data. Especially, NABIC..

Analysis Tools
Analysis Tools area provides a comprehensive range of tools for the field of bioinformatics. It consist of Web Blast

Genome Annotation
NIAB had constructed genone annotation system for plant, and microbial. This genome annotation system..

Other Databases
Manages databases of molecular gene map and provides computational services for genome..

Fig. 1. The homepage screenshot of the Agricultural Genome Information Center, which shows information categories for various genomes of agricultural crops.

Table 1. Details of the genomic database for Chinese cabbage (*Brassica rapa*), rice, and Arabidopsis

Dataset	Chinese cabbage	Rice	Arabidopsis
Contig	1,725	3,360	1,619
Gene	–	50,717	30,853
Marker	716	9,587	2,135
SNP	–	72,304	–
EST	15,265	171,805	–
Exon	25,750	193,605	166,873
Total	43,456	501,378	201,480

categories: genome research, gene expression, rice mutant DB, analysis tools, genome annotation, and other databases (Fig. 1). The platform was developed using MYSQL and commonly available network protocols, such as Hypertext Transfer Protocol and JAVA, and data were stored in an Oracle relational database management system (Oracle Database 10g, Redwood, CA, USA, http://www.oracle.com/).

Results

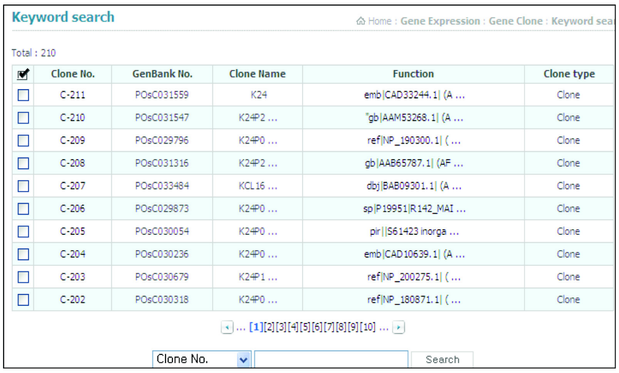
The NABIC information-based database model

The NABIC was established by the NAAS in December 2002. NABIC has provided resources for bioinformatics by developing several bioinformatics tools and constructing 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 integrated database provides genome information from the website (http://nabic.naas.go.kr/) and includes not only simple textual information on individual genome sequences but also analysis tables and genetic markers for annotation. The database provides information and analysis functions on bioinformatics fields: genome, gene expression, microarray, gene cloning, gene index, genetic maps, and genetic markers.

Analysis system for information-based model

The information-based model consists of multiple subsystems: genomic research, gene expression, rice mutant DB, analysis tools, and other databases. This model database provides a bioinformatics framework to study biological function based on the genomic sequences of rice, Chinese cabbage, rice Ds-tagging lines, and microorganisms. The information-based model is a source of annotation for genomic sequences, physical maps, sequence comparison, and gene prediction. The model has developed a portable system that is able to handle very large genomes and the associated requirements for sequence analysis. To achieve scalability and consistency of annotation, we developed a system based on a relational database for rice, Chinese cabbage, and microbes. The analysis tools provide a comprehensive range of tools that are applicable to the bioinformatics field, including Web Blast, Gene Prediction, promoter/TF Analysis, Repeat Sequence Analysis, SNP Analysis, and Protein Structure Analysis. They are subdivided into menu categories for convenience and are provided through various software packages, such as WebBlast, FPC, Glimer, and Phrap. To advance genomic research, we constructed an integrated genomic browser database for sequence analysis of the rice (*Oryza sativa*), Arabidopsis (*Arabidopsis thaliana*), and Chinese cabbage (*Brassica rapa*) genomes. The genomic browser database provides annotated genomic information from 43,456, 501,378, and 201,480 records mapped to Chinese cabbage, rice, and Arabidopsis, respectively (Table 1).

The genomic browser provides specific genomic analysis through four different view panels. This browser shows relationships between the genomic sequence and annotated data. Consisting of four viewable panels that are accessible by clicking, the user can access information about individual genes along with functional annotations within the entire chromosome (Kim *et al.*, 2009). The chromosome panel depicts the chromosome banding region for selection, and the overview panel shows the locations of markers and genes. The detailed view panel shows genomic sequence features and genes, as predicted by the FGENESH (http://nabic.naas.go.kr) analysis program. The base pair view panel exhibits the correlation between the ancestry of in-



Total : 210

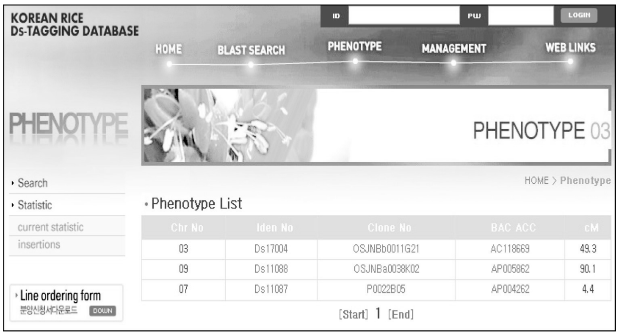
☑	Clone No.	GenBank No.	Clone Name	Function	Clone type
□	C-211	POsC031559	K24	emb|CAD33244.1| (A ...	Clone
□	C-210	POsC031547	K24P2 ...	"gb|AAM53268.1| (A ...	Clone
□	C-209	POsC029796	K24P0 ...	ref|NP_190300.1| (...	Clone
□	C-208	POsC031316	K24P2 ...	gb|AAB65787.1| (AF ...	Clone
□	C-207	POsC033484	KCL16 ...	dbj|BAB09301.1| (A ...	Clone
□	C-206	POsC029873	K24P0 ...	sp|P19951|R142_MAI ...	Clone
□	C-205	POsC030054	K24P0 ...	pir||S61423 inorga ...	Clone
□	C-204	POsC030236	K24P0 ...	emb|CAD10639.1| (A ...	Clone
□	C-203	POsC030679	K24P1 ...	ref|NP_200275.1| (...	Clone
□	C-202	POsC030318	K24P0 ...	ref|NP_180871.1| (...	Clone

Fig. 2. A snapshot of the gene expression database. The view page shows individual windows, searched by clone number.

dividuals and the common variability of pair-wise linkage. In the comparative genome analysis between the Arabidopsis and Brassica rapa genomes, users can obtain new genetic information resulting from comparative genomics methods and identify missing regions within a single genome.

The gene expression database (http://nabic.naas.go.kr/agic/expression/) provides an integrated web-based tool for automatic multistep analysis of gene expression data. It uses bioinformatics tools to compare and evaluate gene expression data originating from treatment with newly developed gene expression systems. The current version contains related information on 34,000 expressed sequence tags for 10 species: rice, wheat, maize, soybean, barley, Chinese cabbage, tomato, hot pepper, mushroom, and Arabidopsis (Kim *et al*., 2008). Both methods allow a more general approach for discovering the underlying structures and patterns in gene expression data than previous methods and can be a valuable tool for revealing the complexity and fine structure of plants (Fig. 2).

The rice mutant database (http://nabic.naas.go.kr/RDS/) provides comprehensive information about mutant phenotypes and insertion site sequence information of Ds-tagging lines, which generated 115,000 Ac/Ds insertional mutation lines using japonica rice (Kim *et al*., 2008). The rice mutant DB has five major menus of web pages: a Blast Search for mutant lines, Blast from rice Ds-tagging mutant lines, a primer design tool to identify genotypes, a phenotype menu for Ds lines, and management information for Ds lines (Fig. 3).



Phenotype List

Chr No	Iden No	Clone No	BAC ACC	cM
03	Ds17004	OSJNBb0011G21	AC118669	49.3
09	Ds11088	OSJNBa0038K02	AP005862	90.1
07	Ds11087	P0022B05	AP004262	4.4

Fig. 3. The homepage of the Korean rice Ds-tagging lines database. This database shows comprehensive information about mutant phenotypes and sequence information of Ds-tagging lines. The phenotype view page shows the results of search for a DS line.

Discussion

A challenge in the postgenomic era is complete genomic information, such as proteomics, metabolism, and systems biology. Bioinformatics centers have been increasing the capacity of biotechnological research by improving the performance of their services. In the future, centers will provide services to construct workflows and pipelines easily, combining two or more instructions to solve complex biological problems, such as protein function prediction, systems biology, genomic annotation, gene pathway, and microarray analysis. Therefore, bioinformatics centers must develop an integrated network system to aid the navigation of genomics, system biology, metabolism, and proteomics tasks. The NABIC was established in 2002 with the main objective of analyzing the genomic information of agricultural crops and provides related services to professional genomic research institutes and societies (NAAS, 2010). The integrated biotechnology database provides genomic information, including genome projects, gene identification numbers, genetic markers, genetic information, specific gene sequences, and genetic information tables for annotation in agricultural crops. In addition, we constructed a fundamental system to analyze massive sequencing data using next-generation sequencing technologies and have contributed to the application of this informatics approach to agricultural biotechnology to extend the usefulness of breeding new crops. In the future, NABIC will provide a web service to construct bioinformatics workflows and pipelines easily, combining two or more instructions to solve complex biological tasks.

Acknowledgments

This study was conducted with support from the Research Program for Agricultural Science & Technology Development (Project No. PJ006651 and PJ006758), National Academy of Agricultural Science, Rural Development Administration, Republic of Korea.

References

Ann, F.B. (2008). Bioinformatics, Genomics, and Proteomics: Getting the Big Picture (Biotechnology in the 21st Century). *Bioinformatics* 9, 94-95.

Kim, C.K., Baek, H.J., Park, H.J., Kim, Y.H., Seol, Y.J., Hahn, J.H., and Lee, G.S. (2010). The Activity and Web Service of Bioinformatics Center in Europe. *The Journal of the Korean society of international agriculture* 22, 1-7.

Kim, C.K., Choi, J.W., Park, D.S., Kang, M.J., and Seol, Y.J. (2008). PlantGI: a database for searching gene indices in agricultural plants developed at NIAB, Korea. *Bioinformation* 2, 344-345.

Kim, C.K., Han, J.H., Shin, Y.H., Park, S.H., Yun, D.W., Ahn, B.O., Kim, D.H., Park, B.S., and Hahn, J.H. (2009). A genome browser database for rice (Oryza sativa) and Chinese cabbage (Brassica rapa). *Afr. J. Biotechnol.* 8, 5253-5259.

Kim, C.K., Kim, J.A., Kim, M.S., Yun, D.W., and Kim, Y.H. (2008). The Recent Database Construction and Web Service of Bioinformatics Center in Japan. *Korean Soc. Int. Agri.* 20, 272-277.

Kim, C.K., Lee, M.C., Ahn, B.O., Yun, D.W., Yoon, U.H., Suh, S.C., Eun, M.Y., and Hahn, J.H. (2008). KRDD: Korean Rice Ds-tagging lines database for rice (Oryza sativa L.). *Genomics & Informatics* 6, 64-67.

NAAS. (2010). NAAS Annual report 2010, National Academy of Agricultural Science (NAAS), RDA, Korea.

Russ, B. (2007). Current progress in bioinformatics. *Briefings in bioinformatics* 8, 277-278.

Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Federhen, S., Feolo, M., Fingerman, I.M., Geer, L.Y., Helmberg, W., Kapustin, Y., Landsman, D., Lipman, D.J., Lu, Z., Madden, T.L., Madej, T., Maglott, D.R., Marchler-Bauer, A., Miller, V., Mizrachi, I., Ostell, J., Panchenko, A., Phan, L., Pruitt, K.D., Schuler, G.D., Sequeira, E., Sherry, S.T., Shumway, M., Sirotkin, K., Slotta, D., Souvorov, A., Starchenko, G., Tatusova, T.A., Wagner, L., Wang, Y., Wilbur, W.J., Yaschenko, E., and Ye, J. (2011). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* 39, D38-51.

Youens-Clark, K., Buckler, E., Casstevens, T., Chen, C., Declerck, G., Derwent, P., Dharmawardhana, P., Jaiswal, P., Kersey, P., Karthikeyan, A.S., Lu, J., McCouch, S.R., Ren, L., Spooner, W., Stein, J.C., Thomason, J., Wei, S., and Ware, D. (2011). Gramene database in 2010: updates and extensions. *Nucleic Acids Res.* 39, D1085-1094.