Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Web Database**

PlantGI: a database for searching gene indices in agricultural plants developed at NIAB, Korea

Chang Kug Kim[1], Ji Weon Choi[2], DongSuk Park[1], Man Jung Kang[1], Young-Joo Seol[1], Do Yoon Hyun[3] and Jang Ho Hahn[1,*]

[1]Bioinformatics Division, National Institute of Agricultural Biotechnology (NIAB), Suwon 441-707, Korea; [2]Postharvest Technology Div., National Horticultural Research Institute (NHRI), Suwon 440-706, Korea; Genetic Resources Div., NIAB, Suwon 441-707, Korea;
JangHo Hahn* - E-mail: jhhahn@rda.go.kr; * Corresponding Author

received April 02, 2008; revised April 22, 2008; accepted April 28, 2008; published May 27, 2008

Abstract:
The Plant Gene Index (PlantGI) database is developed as a web-based search system with search capabilities for keywords to provide information on gene indices specifically for agricultural plants. The database contains specific Gene Index information for ten agricultural species, namely, rice, Chinese cabbage, wheat, maize, soybean, barley, mushroom, Arabidopsis, hot pepper and tomato. PlantGI differs from other Gene Index databases in being specific to agricultural plant species and thus complements services from similar other developments. The database includes options for interactive mining of EST CONTIGS and assembled EST data for user specific keyword queries. The current version of PlantGI contains a total of 34,000 EST CONTIGS data for rice (8488 records), wheat (8560 records), maize (4570 records), soybean (3726 records), barley (3417 records), Chinese cabbage (3602 records), tomato (1236 records), hot pepper (998 records), mushroom (130 records) and Arabidopsis (8 records).

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

Keywords: gene index; EST CONTIGS; identifier EST; database

Background:
The gene indices use available expressed sequence tag (EST), tentative consensus (TC), expressed transcript (ET) and gene sequences along with a reference genome. The EST data in the public databases provide an important resource for comparative and functional genomics studies **[1]**. Gene index is constructed for species organisms by first clustering, then assembling EST and annotated gene sequences. The TGI gene indices provide such an analysis for humans, animal, fungi, plant and use the available EST/gene identifier along with the 34 species genomes **[2, 3]**. The importance of EST assembly and clustering has been well established as evidenced by databases such as XGI **[4]** and TGI database **[3]**. The mining of these datasets is an important component of gene discovery and expression profiling **[5]**. The Gene Ontology (GO) is one of the most important and well-used ontology within the bioinformatics community and is dynamically controlled to describe molecular function, process/location for a protein **[6]**.

The Chinese cabbage (*Brassica rapa*) and rice (*Oryza sativa*) are most important vegetables in Korea and in northeast Asia. The National Institute of Agricultural Biotechnology (NIAB) was established in 1997 to complete rice genome using the cultivar japonica with the participation of the International Rice Genome Sequencing Project (IRGSP) consortium **[7]**. We obtained over 120,000 ESTs of Chinese cabbage (*B. rapa ssp. pekinensis*) derived from 26 different cDNA libraries described elsewhere **[8]**. We used similar base data for different agricultural species genomes to construct PlantGI at NIAB. Here, we describe the construction and utility of PlantGI.

Methodology:
Dataset:
The EST data was collected for Chinese cabbage and rice genome project at NIAB (127,144 EST). We then used the EST dataset at NCBI [11] for species specific Gene Index development.

Development:
Using the collected EST sequence, we constructed a database for EST CONTIGS and an EST assemble map viewer for CONTIG viewing. The EST CONTIG was assembled to EST sequences using the Fragment Assembly program **[12].** The developed used is MYSQL in JAVA application with Oracle RDBMS. Thus, the database is developed using MYSQL **[9]**, JAVA **[10]** and Oracle relational database management system (RDBMS).

Database content:
PlantGI database provides 34,000 EST CONTIGS information for 10 species namely, rice (8488 records), wheat (8560 records), maize (4570 records), soybean (3726 records), barley (3417 records), Chinese cabbage (3602 records), tomato (1236 records), hot pepper (998 records), mushroom (130 records) and Arabidopsis (8 records).

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Web Database**

Database design

The PlantGI is designed to provide information on gene indices in agricultural plants. The databases consist of ten species-specific menus that include Rice (*Oryza sativa*), Chinese cabbage (*Brassica rapa*), Wheat (*Triticum aestivum*), Maize (*Zea mays*), Soybean (*Glycine max*), Barley (*Hordeum vulgare*), Mushroom (*Pleurotus ostreatus*), Arabidopsis (*Arabidopsis thaliana*), Hot Pepper (*Capsicum annuum*) and Tomato (*Lycopersicon esculentum*). It is a web application for interactive mining of EST contig and assembled individual EST data. It is possible to view individual EST sequence data and GenBank records as well as expandable nodes that correspond to EST members for CONTIGS. Figure 1 shows four major menus, namely (i) a web blast search, (ii) identifier searching, (iii) keyword searching and, (iv) gene ontology analysis. Search result shows assembled EST map of CONTIG, individual EST sequences and an expression summary of ESTs within each species.

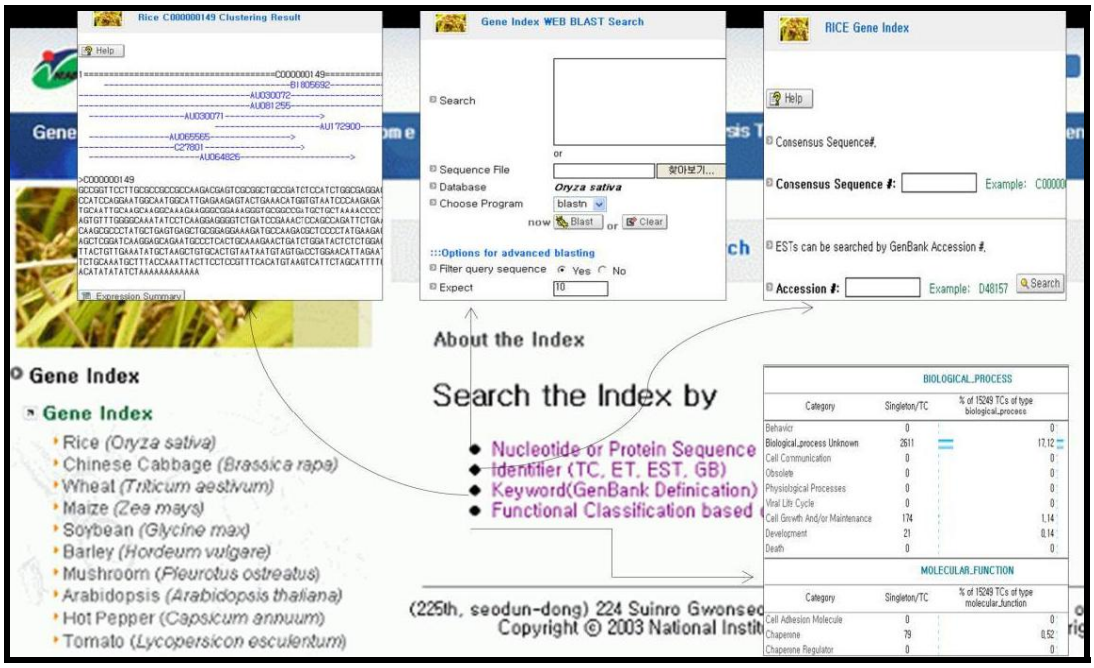

Figure 1: A web snapshot for PlantGI database is shown. The view page shows individual windows; blast search, searching by identifier, EST contig number and GO.

Database usage:

The user can access PlantGI through the web browser using internet. The database is visualized using a web-based graphical view and anonymous users can query and browse the data using the search function. We propose to update data from various public resources on a quarterly basis and hence develop a useful tool for gene index analysis for agricultural plants.

Acknowledgement:

PlantGI database was supported by NIAB research project (No: 2007139062200001602, the construction of agricultural biotechnology information management system)

References:

[01] F. Liang *et al.*, *Nucleic Acids Res.*, 28: 3657 (2000) [PMID: 10982889]
[02] J. Quackenbush *et al.*, *Nucleic Acids Res.*, 28: 141 (2000) [PMID: 14681408]
[03] http://www.tigr.org/tdb/tgi
[04] http://www.ncgr.org/xgi
[05] H. Yecheng *et al.*, *Bioinformatics*, 21: 669 (2005) [PMID: 15374864]
[06] E. Camon *et al.*, *Genome Res.*, 13: 662 (2003) [PMID: 12654719]
[07] http://rgp.dna.affrc.go.jp/IRGSP/
[08] http://www.brassica-rapa.org/BGP/NC_brgp.jsp
[09] http://www.mysql.com
[10] http://www.java.com/ko/
[11] http://www.ncbi.nlm.nih.gov/
[12] http://sems.niab.go.kr/

Edited by P. Kangueane

Citation: Kim *et al.*, Bioinformation 2(8): 344-345 (2008)