

# An integrated database to enhance the identification of SNP markers for rice

ChangKug Kim<sup>1</sup>, UngHan Yoon<sup>1</sup>, GangSeob Lee<sup>1</sup>, SungHan Park<sup>1</sup>, Young-Joo Seol<sup>1</sup>, HwanKi Lee<sup>2</sup> and JangHo Hahn<sup>1,\*</sup>

<sup>1</sup>Genomics Division, National Academy of Agricultural Science (NAAS), Suwon 441-707, Korea; <sup>2</sup>Technology Information Division, Rural Development Administration (RDA), Korea; ChangKug Kim – E-mail: jhhahn@rda.go.kr; \*Corresponding Author

Received December 01, 2009; Accepted December 15, 2009; Published December 31, 2009

**Abstract:**

The National Academy of Agricultural Science (NAAS) has developed a web-based marker database to provide information about SNP markers in rice. The database consists of three major functional categories: map viewing, marker searching and gene annotation. It provides 12,829 SNP markers information including gene location information on 12 chromosomes in rice. The annotation of SNP marker provides information such as marker name, EST number, gene definition and general marker information. Users are assisted in tracing any new structures of the chromosomes and gene positional functions using specific SNP markers.

**Availability:** The database is available for free at <http://nabic.niab.go.kr/SNP/>

**Keywords:** SNP marker, genetic marker, rice database, annotation

**Background:**

Rice is one of the most important crops in Korea and in the Northeast Asia. To reveal the genomic relationship among genetic markers has become a necessary step. In rice, several genetic marker database based on various genetic markers have been constructed for the purpose of displaying positional relationships among mapped markers in each chromosome [1]. A Single Nucleotide Polymorphism (SNP) is a DNA sequence variation occurring when a single nucleotide in the genome differs between members of a species. The SNP marker may be considered the ultimate genetic marker as they represent the finest resolution of a DNA sequence and SNP marker is the most frequently found DNA sequence variations [2]. The development of high-throughput methods for the detection of SNPs has led to a revolution in their use as molecular markers within DNA sequence data [3]. To increase the importance of SNP markers, various databases have been constructed. The SNP Database Network in Japan [4] provides an integrated menu for SNP

searching and marker annotation. The dbSNP [5] provides a function for SNP discovery across the entire plant genome. The AutoSNPdb [6] provides an annotated single nucleotide polymorphism database for the important crops rice, barley and Brassica. The OryzaSNP [7] has generated a large fraction of the SNPs present in cultivated rice through 20 rice genomes. The National Academy of Agricultural Science (NAAS) of Rural Development Administration (RDA) developed a web-based database system to provide information about SNP and QTL markers in rice. The genetic marker database provides 7,227 SNP markers on each of 12 chromosomes by using genetic map in rice [8]. We have released a major update that comprises various new features for annotation including localization of the gene (RSN) information of SNP marker. General improvements to accuracy of gene location information are included along with novel functionality that allows users to compare the different method for varying sets of SNP markers.

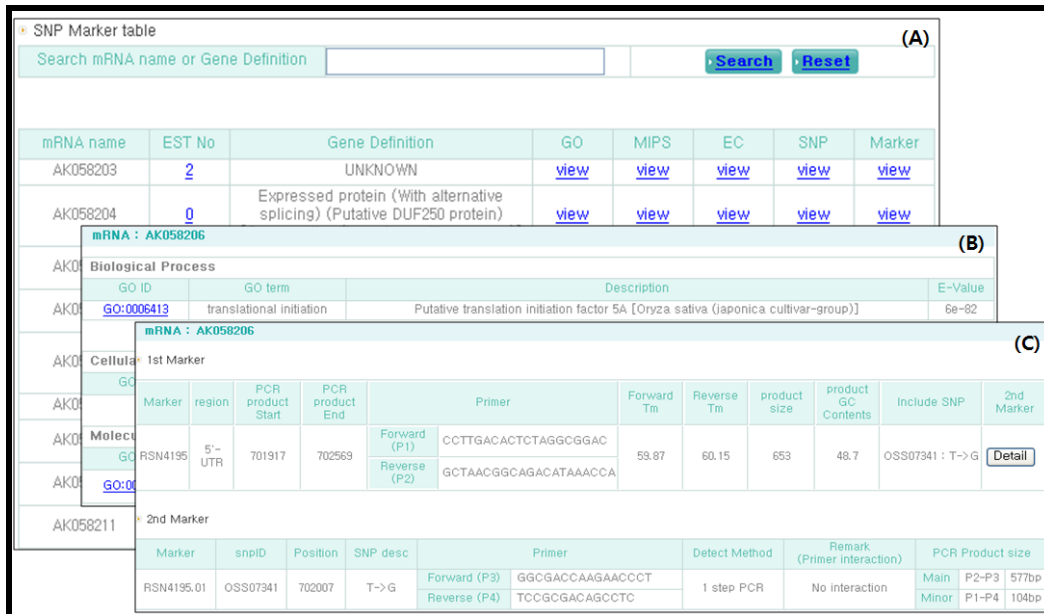


Figure 1. A snap shot of the marker search results table through marker database (A). The individual view table shows samples of search results such as the (B) Gene Ontology (GO) information table (ex: AK058206) by clicking a specific accession number in the 'GO' menu and the (C) general information view table with related marker information (ex: AK058206) by clicking a 'Marker' menu.

**Methodology:**

**Dataset**

The SNP markers were collected from the SNP project (*Oryza sativa*, *Nipponbare*: Japonica and China93-11: Indica) of NAAS [9]. As a result of this project, we have investigated SNPs for genetic markers in rice genome. A total of 12,829 potential SNPs in 3,356 genes were detected from 22,423 non-redundant genes which were generated by pair-wise alignment between ESTs and genomic sequences of rice. In order to construct a database of SNP markers, we analyzed the variation in gene sequences between japonica and indica rice using EST sequence data of 12,829 potential SNPs.

**Database development**

Using the collected SNP markers, a marker database was constructed to distinguish SNP markers. The database platform was developed using MYSQL and JAVA languages. The data was stored in Oracle relational database management system (RDBMS). The logical and physical schema of the database followed the standard principles of relational database by ERWin Data Modeler software.

**Implementation and features:**

The SNP marker database [9] features three major functions: map viewing, marker searching and gene annotation. Map viewer provides SNP markers information including sequence and location information on chromosome using a genetic map. The marker viewer allows that users can access the information of 12,829 SNP markers. Gene annotation menu provides a table of detailed annotation information. When using keyword search or inputting identification number, users can view a table of detailed annotation information through GO, MIPS, EC, SNP and Marker view tables. A specific annotation table provides information such as marker name, EST number, gene definition and general marker information (Figure 1).

**Future work:**

The SNP marker database of NAAS consists of three major functional categories: map viewing, marker searching and gene annotation. The database provides 12,829 SNP markers including gene location information on each of 12 chromosomes in rice. Users are assisted in tracing any new structures of the chromosomes and gene positional functions using specific SNP markers. In the future, NAAS plans to develop an integrated marker database that combines an SNP marker database and the other genetic marker database for rice.

**Acknowledgements:**

This work was supported by the National Academy of Agricultural Science and the Biogreen 21 Program of Rural Development Administration (Project number: 200901FHT020711430).

**References:**

- [1] S Katsumi *et al.*, *Nucleic Acids Research* **28**: 97 (2000) [PMID: 10592192]
- [2] C Duran, *et al.*, *Nucleic Acids Res.* **37**:D951 (2008) [PMID: 18854357]
- [3] D Savage *et al.*, *Nucleic Acids Res.* **33**:W493 (2003) [PMID: 15980519]
- [4] <http://snpnet.jst.go.jp/>
- [5] <http://www.ncbi.nlm.nih.gov/projects/SNP/>
- [6] <http://autosnpdb.qfab.org.au/>
- [7] <http://www.oryzasnp.org/>
- [8] Chang *et al.*, *Bioinformatics.* **3**(2):61 (2008) [PMID: 19238232]
- [9] <http://nabic.niab.go.kr/SNP/>

**Edited by P. Kanguane**

**Citation: Kim *et al.***, *Bioinformatics* 4(6): 269-270 (2009)

**License statement:** This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.