NABIC SNP: an integrated database for SNP markers

Chang-Kug Kim, So-Youn Won, Young-Joo Seol & Tae-Ho Lee*

Genomics Division, National Academy of Agricultural Science (NAAS), Jeonju 54874, Korea; Tae-Ho Lee1 – Email: thlee0@korea.kr; *Corresponding author

Received June 14, 2015; Accepted July 15, 2015; Published July 31, 2015

Abstract:
The National Agricultural Biotechnology Information Center (NABIC) constructed a web-based database to provide information about 54,310 single nucleotide polymorphisms (SNPs) identified in the seven species in a high-throughput manner. The database consists of three major functional categories: SNP marker search, detailed information viewer and download of SNP sequence. The SNP annotation table provides detailed information such as ownership information, basic information, bio-entry information, reference, comments, features, and sequence data.

Availability: The database is available online for free at http://nabic.rda.go.kr/SNP

Keywords: NABIC, SNP database, SNP marker search, molecular marker, NGS

Background:
A single nucleotide polymorphism (SNP) is a DNA sequence variation that occurs at the level of one base-pair among individuals or between paired chromosomes [1]. It is highly abundant in the genomes and has been frequently utilized as a molecular marker in the diverse field. Particularly, its association with a specific trait such as development, resistance to abiotic or biotic stresses or disease symptoms has served as one of the most useful tools in the area of, for example, crop breeding or disease diagnosis etc. Recently, the progress in next-generation sequencing (NGS) technologies has tremendously accumulated sequence information from which great amounts of SNPs have been mined. Furthermore, NGS, SNP chip and high-resolution melting analysis can easily detect SNPs with the relatively lower cost and labor in both high- and low-throughput mode, which facilitates the usage of SNPs in the molecular, genetic and genomic study such as quantitative trait loci mapping, genotyping-by-sequencing and genome-wide association studies etc. With the explosion of SNP-related information, multiple databases for SNP markers have been constructed. The dbSNP is a public large-scale database of simple genetic variations in any species and is operated by the National Center for Biotechnology Information [2]. JSNP is a repository specializing in SNPs from the Japanese population and supports user interface facilitating database accessibility via keywords search such as polymorphisms, genes and drug-metabolizing enzymes, etc [3]. For the important crops such as rice, barley, wheat and Brassica, the AutoSNPdb provides both information on annotated SNPs using graphical interface and integrated pipeline to discover SNPs [4]. Especially for rice, the SNP-Seek system is a database for SNPs as well as genotype, phenotype and variety information about 3,000 varieties [5], while the HapRice is a repository for haplotype SNPs and provides web-based tools for finding polymorphic SNPs and designing primers to develop cleaved amplified polymorphic sequence markers at any SNP [6]. Considering the significance of SNP database and its application, the National Agricultural Biotechnology Information Center (NABIC) previously provided a function for SNP discovery across the entire rice genome and SNP information such as marker name, EST number, gene definition and general marker information [7]. Recently, it has
expanding its archives to the other species, which will be discussed below.

**Methodology:**

**Data collection and development:**

The information on SNPs was collectively acquired from the research projects supported by Next-Generation BioGreen 21 Program (http://www.tagc.or.kr/) of Rural Development Administration. Initially, SNPs were discovered by pair-wisely aligning expressed sequence tags (ESTs) to genomic sequences and identifying polymorphic sequences between the two. In addition, they were collected from high density microarray platforms and next generation whole-genome sequencing data. A total of 54,310 potential SNPs were detected from non-redundant genes for cattle (*Bos taurus*, 12,000 records), swine (*Sus scrofa*, 609 records), oriental melon (*Cucumis melo*, 41,651 records), grape (*Vitis flexuosa*, 14 records), rootstock for grape (*Vitis vinifera*, 10 records), mung bean (*Vigna radiate*, 24 records) and small sea snails (*Hemifusus tuba*, 2 records). Using the collected SNP markers, a database was constructed using Hypertext Markup Language (HTML), Java and Hypertext Preprocessor (PHP). Its platform for the web server was developed using MySQL for PHP environment and database handling.

![Figure 1](image-url)  

**Figure 1:** A snapshot of the marker search result. On the top image, the oriental melon (*Cucumis melo*) is selected for keyword search. The bottom table shows the list of SNPs and their brief information that are present in the oriental melon.
Implementation and features:
The NABIC SNP database features three major functions: marker search, detailed information viewer, and download of SNP sequence (Figure 1). Through marker search, users can easily access specific SNPs information among 54,310 SNP markers in seven species. Specifically, this database supports keyword search consisting of seven identification categories (i.e., ID, source, species, type, content, pub date, and ownership). For example, if ‘Cucumis melo’ is entered as a query in the species of marker search, a summarized table is generated, as shown in Figure 1. Clicking on ID shows the detailed information of SNPs, such as ownership information, basic information, bio-entry information, reference, comments, features and sequence data (Figure 2).

Utility, caveats and future developments:
The NABIC SNP database provides detailed information on SNPs in seven species and has three major utility features (detailed information viewer, SNP search and download function). Using specific SNP markers, any new structures of the chromosomes and positional function of gene are easily...
tracked. The NABIC SNP database was planned to integrate the detailed information on SNPs and provide them through a user-friendly platform with marker search options online. In the future, NABIC will be further improved by combining more information including SNPs and other genetic markers into its database storage and expanding or refining keyword options for marker search.

Acknowledgement:
This study was conducted with support from the Research Program for Agricultural Science & Technology Development (Project No. PJ010112) of the National Academy of Agricultural Science, Rural Development Administration, Republic of Korea.

References:
[1] Mammadov J et al. Int J Plant Genomics. 2012 2012: 728398 [PMID: 23316221]
[2] Acland A et al. Nucleic Acids Res. 2014 42: D7 [PMID: 24259429]
[3] Hirakawa M et al. Nucleic Acids Res. 2002 30: 158 [PMID: 11752280]
[4] Duran C et al. Nucleic Acids Res. 2009 37: D951 [PMID: 18854357]
[5] Alexandrov N et al. Nucleic Acids Res. 2015 43: D1023 [PMID: 25429973]
[6] Yonemaru J et al. Plant Cell Physiol. 2014 55: e9 [PMID: 24334415]
[7] Kim C et al. Bioinformation 2009 4: 269 [PMID: 20975922]

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License.