Molecular marker database for efficient use in agricultural breeding programs

Chang-Kug Kim[1], Gang-Seob Lee[2], Ji-Su Mo[1], Seon-Hwa Bae[1], Tae-Ho Lee[1]*

[1]Genomics Division, National Academy of Agricultural Science (NAAS), Jeonju 54874, Korea; [2]Biosafety Division, NAAS, Jeonju 54874, Korea; Tae-Ho Lee - Email: thlee0@korea.kr; *Corresponding author

Abstract:
The National Agricultural Biotechnology Information Center (NABIC) constructed a web-based molecular marker database to provide information about 7,847 sequence-tagged site (STS) markers identified in the 11 species using a next generation sequencing (NGS) technologies. The database consists of three major functional categories: keyword search, detailed viewer and download function. The molecular marker annotation table provides detailed information such as ownership information, basic information, and STS-related characterization information.

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

Keywords: molecular marker, sequence-tagged site marker

Background:
A molecular marker is a fragment of DNA that is associated with a certain location within the genome. In general, molecular markers are used in molecular biology to identify a particular sequence in a pool of unknown DNA. Molecular markers are widely used in genetic studies of agricultural plants. Advancement in the next generation sequencing (NGS) technologies integrated with tools like association mapping studies. The NGS method is far more powerful than any existing in generating DNA markers [1]. Of currently available markers are based on the polymerase chain reaction (PCR) such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), sequence-tagged site (STS), cleaved amplified polymorphic sequences (CAPS), microsatellites or simple sequence repeat (SSR), and single nucleotide polymorphism (SNP) [2]. A STS marker is a short DNA sequence that has a single occurrence in the genome and whose location and base sequence are known. The STS marker can be easily detected using specific primers, and they are useful for developing physical map of a genome [3]. Molecular markers have been increasingly used in genetic studies of crop species for their applicability in breeding programs. These markers have been developed and used for the improvement of agricultural species, disease resistance, food quality and

environmental stress traits have been considered [4]. Considering the significance of molecular marker database, the National Agricultural Biotechnology Information Center (NABIC) previously provided a function for RSN marker discovery across the rice and Chinese cabbage genome [5]. Recently, we have released a major update that comprises various new features for annotation including 11 species STS molecular markers analysis information. The molecular marker database has expanded its archives to the other species, which will be discussed below.

Methodology:
Data collection
The information on molecular markers was collectively acquired from the research projects supported by Next-Generation BioGreen 21 Program (http://www. tagc.or.kr/) and post-genome project (http://atis.rda. go.kr/) of Rural Development Administration. A total of 7,847 molecular markers were collected from non-redundant genes for rice (Oryza sativa, 7418 records), small sea snails (Hemifusus tuba, 12 records), chili pepper (Capsicum chinense, 116 records), bell pepper (Capsicum annuum, 3 records), cattle (Bos taurus, 57 records), pear (Pyrus pyrifolia, 26 records), grass (Miscanthus, 201 records), spider mite (Tetranychus urticae, 2 records), wild

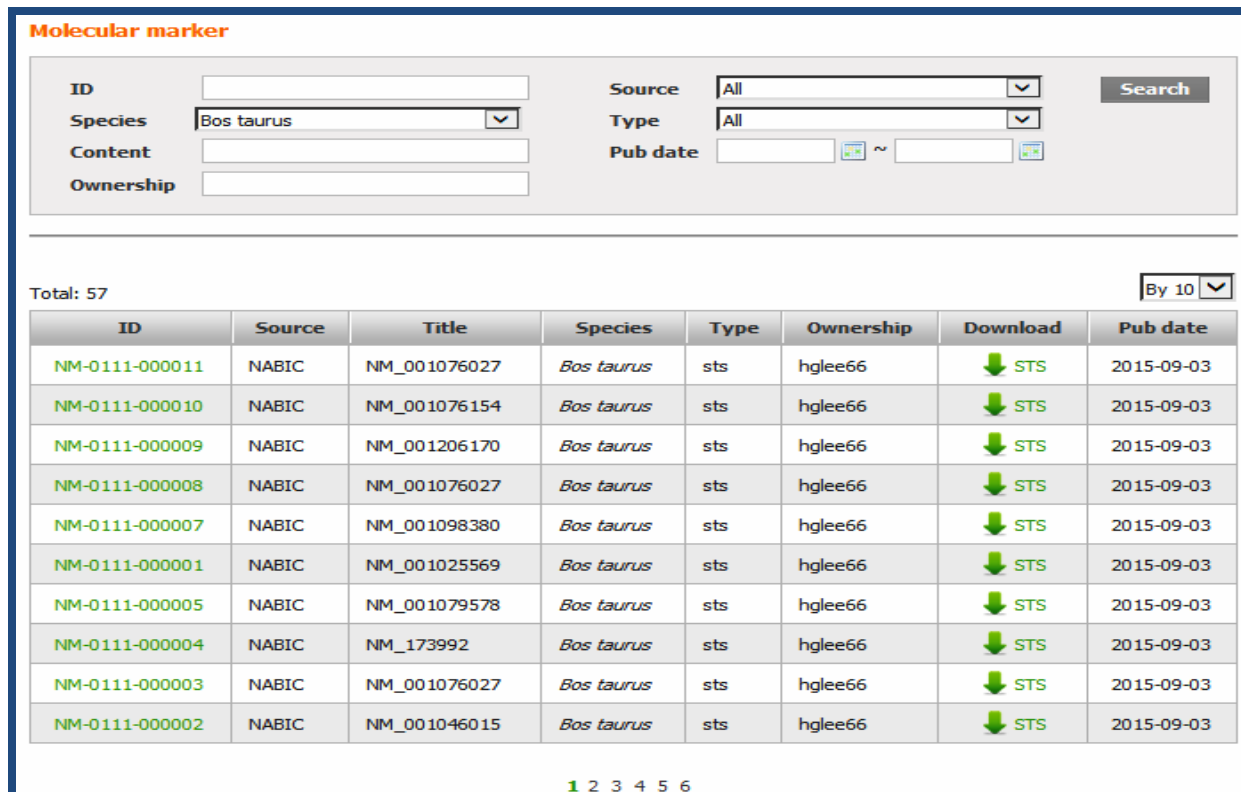soybean (Glycine soja, 5 records), soybean (Glycine max, 5 records), and Chinese cabbage (Brassica rapa, 2 records).

Figure 1: A snap shot of the molecular marker search result. On the top image, the cattle (Bos taurus) is selected for keyword search. The bottom table shows the list of molecular markers and their brief information that are present in the Bos taurus.
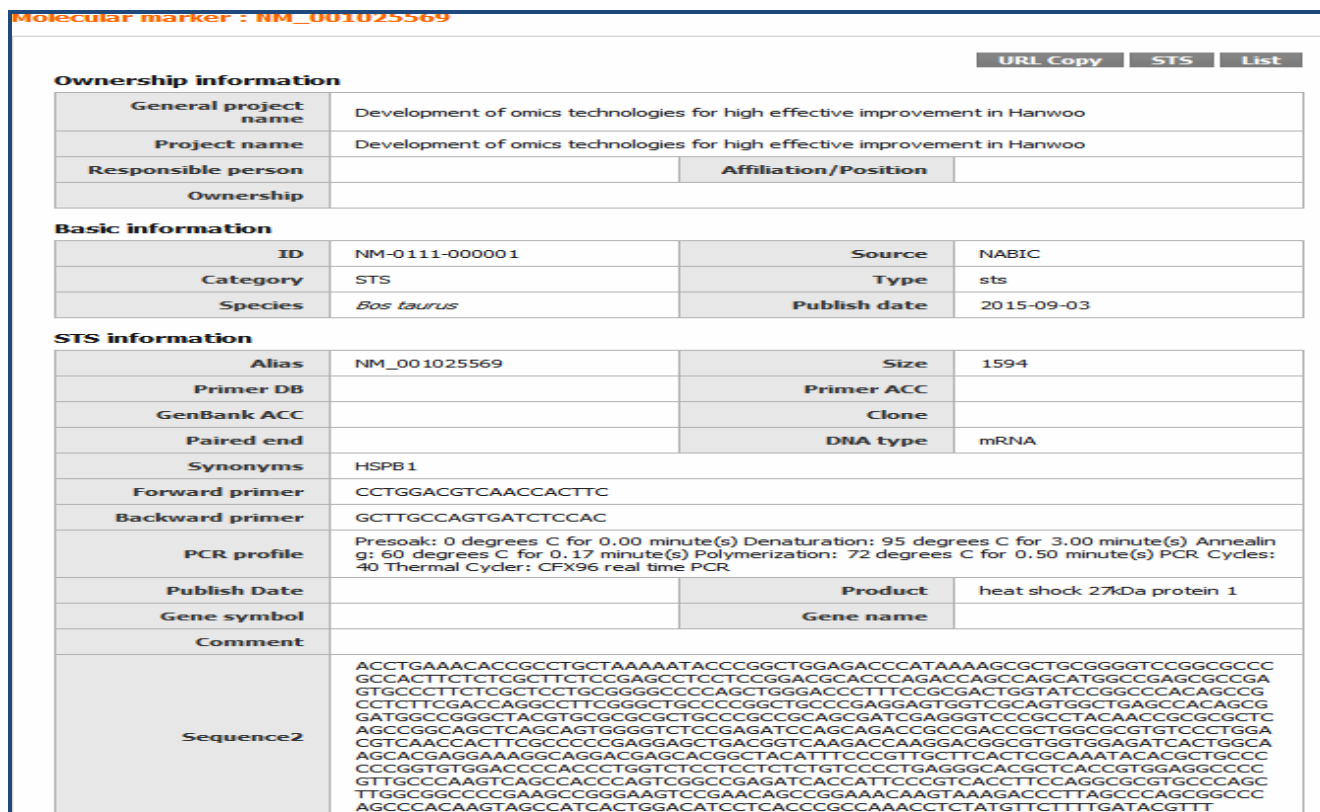

Figure 2: A screenshot showing detailed information of a particular STS marker. The tables show the annotated information for STS molecular marker NM-001025569 in the cattle (Bos taurus).

BIOINFORMATION

Database design

Using the collected molecular markers, a database was constructed using BioSQL schema to covering public and private platforms. Its platform for the web server was developed using MySQL for PHP environment and database handling. We developed a web-based system to enable searches for agricultural markers and to provide various services. The molecular marker database can be accessed freely through NABIC portals (http://nabic.rda.go.kr/Molecularmarker) without any license fees. No prior registration is required.

Implementation and Features:

The STS marker database features three major functions such as marker search, detailed viewer, and data download **(Figure 1).** Through marker search, users can easily access specific molecular markers information among 7,847 molecular markers in 11 species. Specifically, this database supports various keyword search methods consisting of seven identification categories such as ID, source, pecies, type, content, publication date, and ownership. For example, if 'Bos taurus' 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 molecular markers, such as ownership information, basic information, and sequence-tagged site (STS) data **(Figure 2)**.

Utility, caveats and future developments

The molecular marker database provides detailed information on STS molecular markers of agricultural species and has three major utility features (marker search, detailed viewer, and download function). Using molecular marker database, user can easily searched for tracking of new structures of the chromosomes and positional function of genes. In the future, NABIC continually provides related services to professional genomic research institutes and societies. We plan to integrate the detailed information on molecular markers for agricultural breeding and provide them through a user-friendly platform with various multi-search options online.

Acknowledgement:

References:

[1] Yang H *et al. BMC Genomics*. 2012 **13**: 318 [PMID: 22805587]
[2] Perry DJ *et al. Heredity*. 1999 **83**: 239 [PMID: 10504420]
[3] Ward T *et al Hum Mol Genet*. 1993 **2**: 1097 [PMID: 8401489]
[4] Kage U *et al. Crit Rev Biotechnol*. 2015 **16**: 1 [PMID: 26171816]
[5] Kim Ck *et al. Bioinformation*. 2013 **9**: 887 [PMID: 24250118]

Edited by P Kangueane

Citation: **Kim** *et al.* Bioinformation 11(9): 444-446 (2015)