An integrated database for the enhanced identification of silkworm gene resources

ChangKug Kim[1], KeeYoung Kim[2], DongSuk Park[1], YoungJoo Seol[1], JangHo Hahn[1], SooChul Park[1], PilDon Kang[2,*]

[1]Genomics Division, National Academy of Agricultural Science (NAAS), Suwon 441-707, Korea, [2]Sericultural & Apicultural Materials Division, National Academy of Agricultural Science (NAAS), Korea. PilDon Kang-Email id: kangpd@rda.go.kr. *corresponding author

Abstract:
The National Academy of Agricultural Science (NAAS) has developed a web-based database to provide characterization information in silkworm. The silkworm database has four major function menus: variety searching, characterization viewing, general information and photo gallery. It provides 321 silkworm varieties characterization information for six different regions namely, Korean, Japanese, Chinese, European, Tropical and non-classified group. Additionally, the database provides 1,132 photo images regarding life cycle of various silkworm varieties. A specific characterization information table provides accession number, variety, strain and larval marking, blood color, cocoon color, cocoon shape, egg colors, remarks and image table provides photos which consist of shape and color in the different stages of larval, egg and cocoon stages.

Availability: The database is available for free at http://www.naas.go.kr/silkworm/english/

Keywords: silkworm variety, gene resources, silkworm races, silkworm database

Background:
The silkworm is one of the most economically important, beneficial insects in many developing countries owing to for silk production and source of biomaterials. With advances in genomic technologies, there is a rapid increase in silkworm-related data including genome sequences and genetic markers. The SilkDB database is a web based repository for the curation, integration and study of silkworm genetic and genomic data [1]. KAIKObase is an integrated silkworm genome database and data mining tool with 4 map browsers, 1 gene viewer, and 2 independent databases [2]. The Central Sericultural Germplasm Resources Centre (CSGRC) provides 1,120 mulberry and 432 silkworm germplasm resources [3]. The WildSilkbase provides the database is provided with 57,113 ESTs which are clustered and assembled into 4,019 contigs and 10,019 singletons derived from 3 major wild silk-moth [4]. The SilkSatDb is a relational database of microsatellites extracted from the available expressed sequence tags and whole genome shotgun sequences of the silk-moth [5]. The National Academy of Agricultural Science (NAAS) has developed a database system to provide integrated information about various gene resources [6-8] and has launched the silkworm gene resources project [9]. Towards this effort, we developed a web-based database to provide characterization and images information in silkworm.

Methodology:
Dataset:
The silkworm varieties collected from the silkworm gene resources project of NAAS. As a result of this project, we have investigated the characterization of silkworms as variety, strain and larval marking including shape and color in the different stage images. The silkworm database provides 321 silkworm variety information for six different regions namely, Korean (4 varieties), Japanese (90 varieties), Chinese (96 varieties), European (32 varieties), Tropical (9 varieties) and non-classified (90 varieties). The 1,132 photo images of shape and color information collected from the life cycle of various silkworm varieties. Additionally, the database was distinguished four different race types, namely univoltine (32 races), bivoltine (188 races), multivoltine (11 races) and others (90 races).

Figure 1: A snap shot of the specific characterization information table through silkworm database. The characterization and image information table show sample (Hansung banmun) by clicking a specific identification number (No. 143).

Database development:
Using the collected silkworm varieties characterization and image information, we have developed a silkworm gene resources database with four main features: variety searching, characterization viewing, general information and photo gallery. The 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 silkworm database **[10]** has four major function menus: variety searching, characterization viewing, general information and photo gallery. The variety searching viewer allows that users can access the information of 321 silkworm gene resources. Characterization viewer provides silkworm gene resources characterization information including images in the life cycle such as larval, egg and cocoon stages. General information menu provides a table of detailed information such as description of silkworm, brushing, rearing of young silkworms, rearing of advanced silkworms, mounting, cocoon harvesting and cocoon sorting. The photo gallery menu provides 1,132 images regarding shape and color in the life cycle of various silkworm varieties. All images of database provide an enlargement photograph by clicking a specific image. When using category search or clicking identification number, users can view a table of detailed characterization table and image table of three life cycle stages in silkworm. A specific characterization information table provides accession number, variety, strain and larval marking, blood color, cocoon color, cocoon shape, egg colors, remarks and image table provides photos which consist of shape and color in the different stages of larval, egg and cocoon stages **(Figure 1).**

Future work:
The silkworm gene resources database consists of four major functional categories: variety searching, characterization viewing, general information and photo gallery. The database provides of 321 silkworm gene resources and 1,132 images regarding life cycle of various silkworm varieties. Users are assisted in tracing any new variety and utilization on basic such as silkworm breeding using specific characterization information. In the future, we plans to develop an integrated silkworm database not only commercial silkworm genes but also wild silkworms and other rare insect species.

References:
[1] J Duan *et al. Nucleic Acids Res.* **38**: D453-6 (2010) [PMID: 19793867]
[2] M Shimomura *et al. BMC Genomics*. **10**: 486 (2009) [PMID: 19843344]
[3] http://www.silkgermplasm.com/
[4] KP Arunkumar *et al. BMC Genomics.* **9**:338 (2008) [PMID: 18637161]
[5] MD Prasad *et al. Nucleic Acids Res.***33**: D403-6 (2005) [PMID: 15608226]
[6] CK Kim *et al. Bioinformation*. **2**(8): 344-5 (2008) [PMID: 18685722]
[7] CK Kim *et al. Bioinformation*. **3**(2):61-62 (2008) [PMID: 19238232]
[8] CK Kim *et al. Bioinformation.* **3**(8):344-345 (2009) [PMID: 19707297]
[9] http://www.naas.go.kr/silkworm/
[10] http://www.naas.go.kr/silkworm/english/