

NABIC Microarray: an integrated database of high-throughput data for gene expression profiles

Jae-Hee Lee¹, Sang-Ho Kang¹, Jong-Yeol Lee², Chang-Kug Kim^{1*}

¹Genomics Division, National Academy of Agricultural Science (NAAS), RDA, Jeonju 560-500, Republic of Korea; ²Metabolic Engineering Division, National Institute of Agricultural Sciences, Jeonju 54874, Korea; Chang-Kug Kim - Email: chang@korea.kr; Phone:+82-63-238-4555; Fax +82-63-238-4554; *Corresponding author

Received November 12, 2015; Accepted November 15, 2015; Published November 30, 2015

Abstract:

The National Agricultural Biotechnology Information Center (NABIC) in Korea constructed a web-based database to provide information about gene expression profiles identified in the microorganism, plants, and animals. The deposited archive of NABIC microarray database consists of metadata spreadsheet, matrix spreadsheet, and raw data files. It provides three major functions such as microarray search, viewer and download option of raw data. An information table of five fields (i.e., ownership, basic, series, samples, and protocols) shows the specific description of data for selected DNA microarray.

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

Keywords: DNA microarray, gene expression profile, NABIC microarray

Background:

The massive volumes of gene expression data have greatly expanded our knowledge of genetic mechanisms. Microarray and other gene expression platforms can provide a vast amount of information about transcriptional products of biological samples under a variety of conditions [1]. Many relational databases developed for a repository containing gene expression data has been linked to biological annotations for the genes on the array [2-3]. The Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) repository leads to the archiving as a hub for microarray data deposit and retrieval [4]. ArrayExpress (<https://www.ebi.ac.uk/arrayexpress/>) is a public database which has two major goals not only to serve as an archive providing access to microarray data supporting publications but also to build a knowledge base of gene expression profiles [5]. Expression Atlas (<https://www.ebi.ac.uk/gxa/>) database consists of selected high-quality microarray and RNA-sequencing experiments from ArrayExpress that have been annotated and processed using standardized microarray [6]. According to accumulation of vast amounts of microarray data in public databases, a user is

possible to retrieve, integrate, and compare microarray results from many datasets [7-8].

Methodology:

Data collection and development

The microarray information on the DNA chip was collected from the National Institute of Agricultural Sciences (NAS, <http://www.naas.go.kr/>), Next-Generation BG 21 Program (BG21, <http://atis.rda.go.kr/>), and National Post-genome Project (<http://nabic.rda.go.kr/nagp/>). A total of 66 DNA microarray profiles were collected for Arabidopsis (*Arabidopsis thaliana*, 2 records), Chinese cabbage (*Brassica rapa*, 18 records), rice (*Oryza sativa*, 27 records), potato (*Solanum tuberosum*, 2 records), bacteria (*Cyanobacterium*, 3 records), chickens (*Gallus gallus*, 13 records), and human (*Homo sapiens*, 1 records). This database was developed as a web-based system to enable searches for agricultural DNA microarray data, and it is the official management database for government-funded biotechnology research projects in Korea. We receive data submissions for microarray, and perform data quality checks, storage, and management. This database has been developed

using the BioSQL schema to construct a standard database covering public and private platforms, which are derived from NCBI/GEO platform (<http://www.ncbi.nlm.nih.gov/geo/>). Its

platform was developed using MySQL Enterprise, and Red Hat Enterprise Linux system.

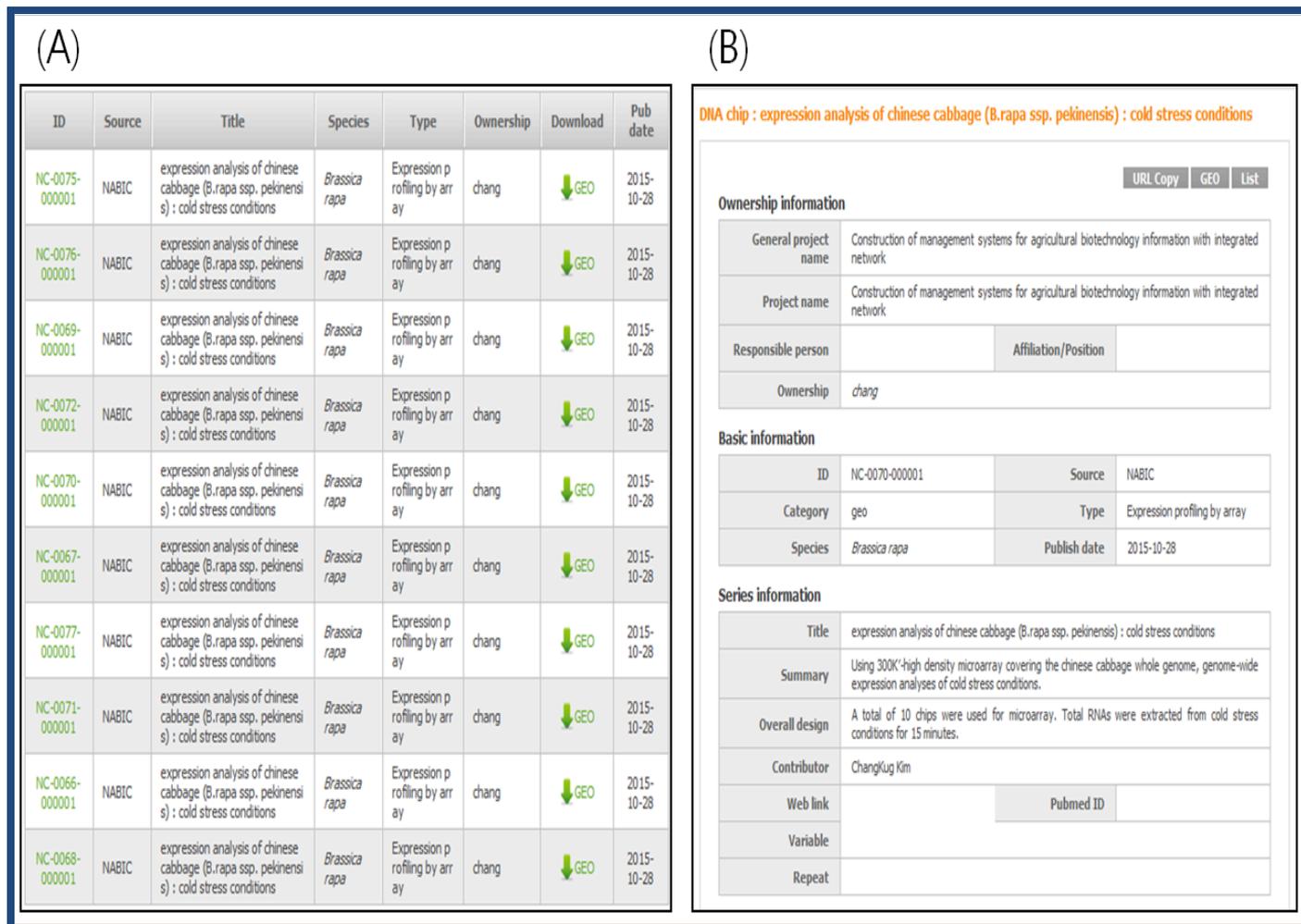


Figure 1: A snap shot of the DNA microarray search result: **A)** The Chinese cabbage (*Brassica rapa*) selected for keyword search and general table shows the list of microarrays with their brief information; **B)** A screenshot shows detailed information of a particular microarray. The tables show each field table of DNA microarray ID NC-0070-000001 in the Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*).

Implementation and features

This database features three major functions: search, viewer, and download. Using the various search options, users can easily access specific gene expression profiles information among 66 DNA microarrays in seven species. Specifically, a user can identify microarray profile datasets of interest by entering keywords in seven identification categories (i.e., identification number, source, registered species, data type, experimental array content, publication date, and ownership). For example, if '*Brassica rapa*' is entered as a query in the species search option menu, a summarized table is generated, as shown in **Figure 1A**. Also, user can receive the raw data file by clicking on download menu. Clicking on ID shows the detailed field information of selected microarray, such as ownership, basic, series, samples, and protocols field (**Figure 1B**). The series field consists of unique title, summary of objectives, and design of experiment. The samples field was further categorized according to experimental variables such as unique title of this sample, source name (i.e., biological material), organism, characteristics tag, type of molecule, label, description,

platform, and raw data file. The protocols field shows the following description: extract material, label, hybridization, scanning and image acquisition, data processing, and value definition.

Utility, caveats and future developments

The NABIC microarray database in Korea provides detailed information on gene expression profiles in seven species and has three major utility features such as DNA microarray search, detailed field information viewer, and download of raw data file. The deposited archive of NABIC microarray database is a flexible spreadsheet-based submission format useful for batch deposit of experiments. Metadata spreadsheet refers to descriptive protocols for the overall experiment and matrix table is a spreadsheet containing the normalized values. In future, we have further planned to provide them through a user-friendly platform to integrate the detailed information with both DNA microarrays and RNA-sequencing expression data.

Acknowledgement:

This study was conducted with support from the research program for agricultural science and technology development (Project No. PJ010112) of the NAAS and the Next-Generation BioGreen 21 Program (SSAC, Grant No PJ011650), RDA, Republic of Korea.

References:

- [1] Couture O *et al.* *Mamm Genome*. 2009 **20**: 768 [PMID: 19936830]

- [2] Vollrath AL *et al.* *BMC Bioinformatics*. 2009 **10**: 280 [PMID: 19732451]
[3] Zhu Y *et al.*, *BMC Bioinformatics* 2008, **9**: 46 [PMID: 18218103]
[4] Barrett T *et al.* *Nucleic Acids Res.* 2005 **33**: D562 [PMID: 15608262]
[5] Brazma A *et al.* *Methods Enzymol.* 2006 **411**: 370 [PMID: 16939801]
[6] Petryszak R *et al.* *Nucleic Acids Res.* 2014 **42**: D296 [PMID: 24304889]
[7] Ramasamy A *et al.* *PLoS Med.* 2008 **5**: e184 [PMID: 18767902]
[8] Kim C *et al.* *Bioinformatics*. 2012 **8**: 1059 [PMID: 23275706]

Edited by P Kanguane

Citation: Lee *et al.* *Bioinformatics* 11(11): 509-511 (2015)

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.