

Full Length Research Paper

KACC: An identification and characterization for microbial resources in Korea

ChangKug Kim¹, YoungAh Jeon², GyuTaek Cho³, SoonWoo Kwon², JangHo Hahn¹ and SeungBeom Hong^{2*}

¹Bioinformatics Division, National Institute of Agricultural Biotechnology (NIAB), Suwon 441-707, Korea.

²Korean Agricultural Culture Collection (KACC), NIAB, Suwon 441-707, Korea.

³Genetic Resources Div., NIAB, Suwon 441-707, Korea.

Accepted 16 September, 2008

Korean Agricultural Culture Collection (KACC) is an authorized organizer and the official depository for microbial resources in Korea. The KACC has developed a web-based database system to provide integrated information about microbial resources. It includes not only simple text information on individual microbe but also morphological images and DNA sequence data for the strain. The KACC now provides the characterization information which maintains 7,433 cultures of microorganisms including 2,687 strains of bacteria, 3535 fungi, 476 actinomycetes, 64 yeasts, and 671 others (mushrooms, gene clones, etc).

Key words: KACC, microorganism database, strain information.

INTRODUCTION

A wide variety of microorganisms are closely associated with our daily life and their maintenance and preservation are very important. In microorganisms, two databases based on various fields have been developed, namely; (i) the IMG 2.3 database which contains 729 bacterial, 46 archaeal genomes, 50 eukaryotic genomes, 1661 viruses and 402 plasmids information (Victor et al., 2007), and (ii) the VirGen database which contains 559 complete genomes in addition to 287 putative genomes of viruses of animals and plants (Kulkarni-Kale et al., 2004). To increase the efficient management for microorganisms, the various integrated systems have been constructed; (i) the GEM (<http://www.gem.re.kr/>) provides an integrated portal web site in a microbial genomic area which includes genome sequence information and functional data, (ii) the NMPDR (<http://www.nmpdr.org/>) provides annotations in an environment for comparative analysis of genomes and biological subsystems which includes 30 archaeal, 565 bacterial, and 29 eukaryal genomes (McNeil et al., 2007), (iii) the VIMSS (<http://vimss.lbl.gov/>) supports an integrated and multi-institutional program to understand

the ability of bacteria and other microorganisms to respond to and survive external stresses, (iv) the MGD (<http://mbgd.genome.ad.jp/>) is a comprehensive platform for microbial comparative genomics and now contains over 300 published genomes (Ikuo, 2007). The users add their own genome sequences to MGD for the purpose of identifying orthologs among the new and the existing genomes, (v) the CBS Fungal Biodiversity Center (<http://www.cbs.knaw.nl/>) maintains over 50,000 strains of micro-organisms, representing a large percentage of the species in the fungal kingdom that have been cultured to date, (vi) the DSMZ microbial collection (<http://www.dsmz.de/microorganisms/>) contains over 15,000 cultures representing some 6,900 species, (vii) the NRRL Culture Collection (<http://nrnl.ncaur.usda.gov/>) is the largest of its kind in the world, with over 85,000 accessions of bacteria and fungi and includes the ARS Patent Culture Collection.

Korean Agricultural Culture Collection (KACC) performs an authorized organizer and official depository for microbial resources in Korea. The KACC maintains the microbial resources including bacteria, actinomycetes, yeasts and filamentous fungi, and has developed a web-based database system to provide integrated information about microbial resources.

*Corresponding author: E-mail sbhong@rda.go.kr.

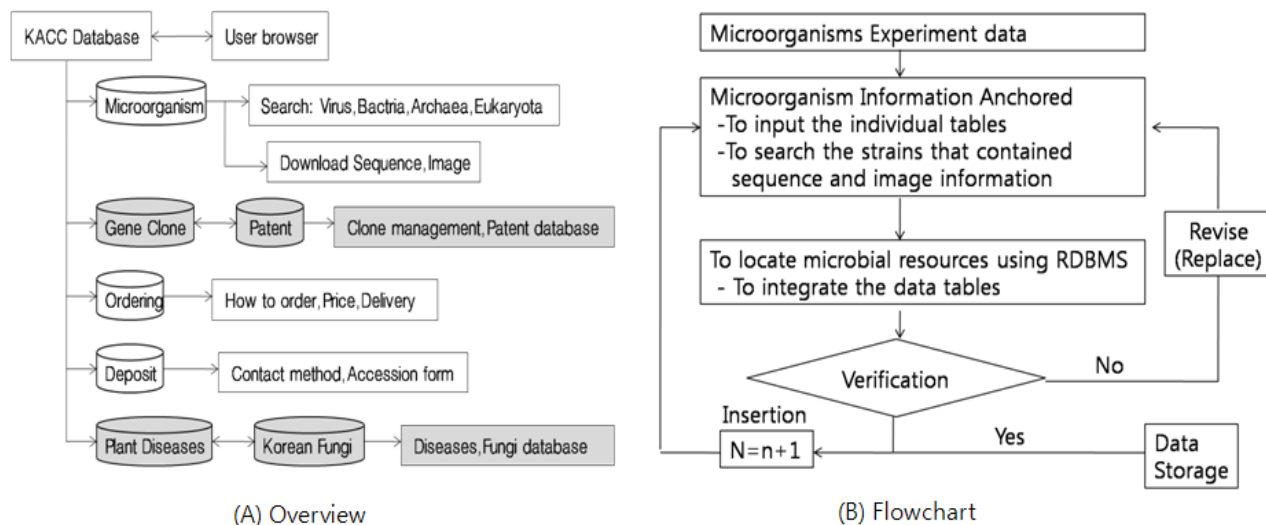


Figure 1. Database design and system flowchart. (A) Overview of KACC database design concepts. Grey objects are supported by only Korean language. (B) System flowchart of registration microorganism information. Finally, data collected from users are stored in the KACC.

METHODOLOGY

Dataset

The information of microbial resources was collected from microorganism project (<http://kacc.rda.go.kr/>), Bio-Green 21 project (<http://biogreen21.rda.go.kr/>), university, and various institutes in Korea. In most instances, the strains were mainly isolated from Korean agricultural environments including crop fields, forest, soil, green house, etc., but some type of strains were shared with other culture collections. Also, it was accumulated and maintained through several international collaborative institutes, such as Centraalbureau voor Schimmelcultures (CBS) in Nederland, Deutsche Sammlung von Microorganismen und Zellkulturen GmbH (DSMZ) in Germany, and ARS Culture Collection in USA.

Database design

The KACC is designed to provide information on Korean microbial resources. The overview of KACC database and simplified view for registration microorganism is shown in Figure 1, and the stored data are visualized using a web-based graphical view. It has five major categories; (i) searching of microorganism by genus and species name, download sequence and image, (ii) gene clone and patent strains management, (iii) ordering and distributing process for research request, (iv) deposit process for patent strains and various biological materials, and (v) plant diseases and Korean fungi database.

Development

Using the collected information, we constructed a microbial resources database system. The platform is developed by MYSQL and JAVA language, and data were stored in Oracle relational database management system (RDBMS). Logical and physical schema of database followed the standard principle of relational database by ERWin Data Modeler software (<http://www.ca.com>) and data were distributed into numerous tables as to have hierarchical relationships according to database design.

Database content

The KACC now maintains 7,433 cultures of microorganisms including 2,687 strains of bacteria, 3535 fungi, 476 actinomycetes, 64 yeasts, and 671 others (mushrooms, gene clones, etc). In 2007, 1,363 strain of new microbial resources were registered which include 647 isolates of bacteria, 617 isolates of fungi (including Basidiomycetes), 95 actinomycetes, and 4 gene clones. In most instances, microorganisms are industrial bacteria as a *Lactobacillus*, *Bacillus*, *Burkholderia*, and contain a rare fungi as *Cerocospora*, *Septoria*, *Ramularia*. The phenotypic and genotypic properties of all these isolates were identified and registered into the KACC. Table 1 shows the number of registered microbial resources according to year.

Implementation and features

The KACC provides strain information from this web site (<http://kacc.rda.go.kr>). It includes not only simple text information on individual microbe but also morphological images and DNA sequence data for the strain. Users can search different classification fields by using the corresponding web form in KACC. There were two field search methods: (i) Keyword search: Users can search by partial or complete genus and species name among classification (Bacteria, Fungi) or categories (scientific name, KACC number, synonyms, sources, characteristics, and other collection number). To choose classification and categories menu, users must specify one sub-menu from the pull-down menu. (ii) Alphabetical order search; strains are alphabetically arranged according to the scientific names. The strain designations listings serve as links to the corresponding individual data pages. Users may search for specific data using by combining these different search parameters and more complex searches can be performed. The search result consists of well-annotated two tables (Figure 2); (i) One general information table, which is shown in the upper part of data page, contains records on Scientific name, Lineage, Characteristics, Literature, Biosafety level, and Strains. Any comments related to the data are provided in the 'Literature' field. (ii) The other KACC information table, which is shown in the lower part of data page, has a records about KACC number, Type, Other collection number,

DOWNLOAD KACC STRAIN SEQUENCE
You can download DNA sequences by using a taxon name.

SEARCH FOR

VIRUS

BACTERIA

Bacteria[295]

Proteobacte

Firmicutes[1

Deinococcu

Bacteroidete

Actinobacter

[total result : 8]

LINEAGE(full) : Bacteria ; Proteobacteria ; Betaproteobacteria

CHECK	ACCESSION No	KACC No	Scientific Name	DEFINITION	DOWN
<input type="checkbox"/>	KA004586	11182	Burkholderia glumae	Partial 16S rRNA gene sequence	[Down]

(A)

(B)

KaccNo	Scientific Name	Definition
11182	Burkholderia glumae	Partial 16S rRNA gene sequence

(C)

LINEAGE(full) : Eukaryota ; Fungi/Metazoa group ; Fungi ; Ascomycota ; mitosporic Ascomycota ;

IMG	KACC No	Scientific Name	TITLE	VIEW
	43022	Spirosphaera sp.	Colony	[View]
	43022	Spirosphaera sp.	Conidia	[View]
	43136	Spirosphaera sp.	Colony	[View]
	43136	Spirosphaera sp.	Conidia	[View]
	43189	Spirosphaera sp.	Colony	[View]

(D)

Figure 3. Web interfaces to the KACC microorganism resources. The individual view page shows samples of analysis result. (A) Users can search different taxon name or historical classification fields by using the corresponding web form. (B) The list of element report page with detailed information. (C) The window provides information to download specific DNA sequence. (D) The list of graphical view element report page with detailed information.

ACKNOWLEDGEMENT

This is to acknowledge the support from NIAB project (No: 05-4-11-28-3) in coming up with this database.

REFERENCES

- Ikuo U (2007). MBGD: a platform for microbial comparative genomics based on the automated construction of orthologous groups, Nucl. Acids Res. 35: D343-D346.
- Kulkarni-Kale U, Bhosle S, Manjari GS, Kolaskar AS (2004). VirGen: a comprehensive viral genome resource, Nucl. Acids Res. 32: D289-D292.

- McNeil LK, Reich C, Aziz RK, Bartels D, Cohoon M, Disz T, Edwards RA, Gerdes S, Hwang K (2007). The National Microbial Pathogen Database Resource (NMPDR): a genomics platform based on subsystem annotation. Nucleic Acids Res. 35: D347-353
- Victor M, Markowitz, Ernest S, Krishna P, Yuri G, Ken Chu, I-Min A, Chen, Inna D, Iain A, Athanasios L, Konstantinos M, Natalia NI, Nikos CK (2007). The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions, Nucl. Acids Res. 36: D528-33.