

# WinBioDBs: A Windows-based Integrated Program for Manipulating Major Biological Databases

Hyeweon Nam<sup>1</sup>, Jin-Ho Lee<sup>2</sup> and Kiejung Park<sup>1,3\*</sup>

<sup>1</sup>Information Technology Institute, SmallSoft Co. Ltd., Daejeon 305-343, Korea, <sup>2</sup>Department of Food Science and Biotechnology, Kyungsoo University, Busan 608-736, Korea, <sup>3</sup>PNU SuperComputing Center, Pusan National University, Busan 609-735, Korea

## Abstract

We have developed WinBioDBs with Windows interfaces, which include importing modules and searching interfaces for 10 major public databases such as GenBank, PIR, SwissProt, Pathway, EPD, ENZYME, REBASE, Prosite, Blocks, and Pfam. User databases can be constructed with searching results of queries and their entries can be edited. The program is a stand-alone database searching program on Windows PC. Database update features are supported by importing raw database files and indexing after downloading them. Users can adjust their own searching environments and report format and construct their own projects consisting of a combination of a local databases. WinBioDBs are implemented with VC++ and its database is based on MySQL.

**Availability:** The detailed features of this system are available from <http://www.smallsoft.co.kr/product/WinBioDBs/index.html>.

**Keywords:** bio-database, GenBank, local database, public database

## Introduction

Many biological databases are supported in commercial softwares or via web interfaces. Entrez and SRS, which are provided with web interfaces, are known as the most popular and powerful programs for this purpose. While users usually need to save searched results, most programs do not support the features.

We have developed a Windows program, WinBioDBs, which provides searching 10 major databases, GenBank

(Benson, 2004), PIR (Barker, 2000), SwissProt (Bairoch, 2004), KEGG Pathway (Kanehisa, 2002), EPD (Schmid, 2006), Enzyme (Bairch, 2000), REBASE (Roberts, 2007), Prosite (Hulo, 2004), BLOCKS (Henikoff, 2000), Pfam (Bateman, 2004) and supports local database features for saving and searching the searched results. Users can import and update the major databases with Windows interfaces and can manipulate their own local databases with searched results.

## Features and Results

The detailed features of WinBioDBs are as follows.

- Windows program
- supporting the search of 10 major biological databases (Table 1) including large databases like GenBank
- searching and formatting the results
- local databases for the searched results
- editing and searching local databases
- customization of search conditions for each database
- customization of output fields for each database
- project features - multiple local databases
- updating public DB
- backup of local databases

The database schema of WinBioDBs is a little complicated (Fig. 1) to support the complicated features (Fig. 2). Database information table is a meta table and each of 10 major public databases constructs a public DB table and a private DB table.

## Discussion

WinBioDB is a Windows program with diverse and

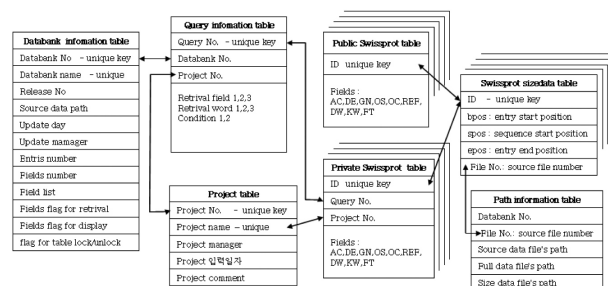


Fig. 1. WinBioDBs database schema

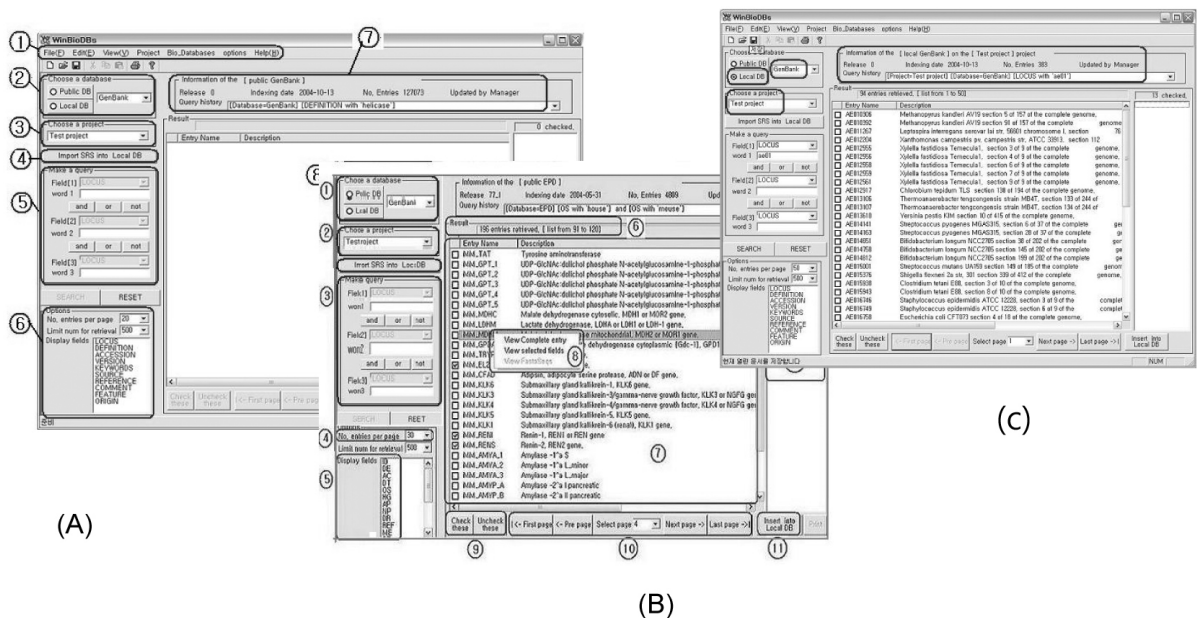
\*Corresponding author: E-mail [kjpark63@gmail.com](mailto:kjpark63@gmail.com)

Tel +82-17-535-5245, Fax +82-42-864-2524

Accepted 23 August 2009

**Table 1.** 10 major public databases in WinBioDBs

	Database names	Entries	URLs
1	GENBANK	25,604,113	<a href="http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html">http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html</a>
2	PIR	239,888	<a href="http://pir.georgetown.edu/">http://pir.georgetown.edu/</a>
3	SWISSPROT	144,731	<a href="http://www.expasy.org/sprot/sprot-top.html">http://www.expasy.org/sprot/sprot-top.html</a>
4	PATHWAY	11,889	<a href="http://www.genome.ad.jp">http://www.genome.ad.jp</a>
5	EPD	4,810	<a href="http://www.epd.isb-sib.ch">http://www.epd.isb-sib.ch</a>
6	ENZYME	4,261	<a href="http://au.expasy.org/enzyme">http://au.expasy.org/enzyme</a>
7	REBASE	4,378	<a href="http://www.hgmp.mrc.ac.uk/Bioinformatics/Databases/rebase-help.html">http://www.hgmp.mrc.ac.uk/Bioinformatics/Databases/rebase-help.html</a>
8	PROSITE	1,641	<a href="http://www.expasy.org/prosite/">http://www.expasy.org/prosite/</a>
9	Blocks	655	<a href="http://blocks.fhcrc.org/">http://blocks.fhcrc.org/</a>
10	Pfam	—	<a href="http://www.sanger.ac.uk/Software/Pfam">http://www.sanger.ac.uk/Software/Pfam</a>



**Fig. 2.** WinBioDBs interfaces. (A) main interface (B) public database searching (C) local database searching. (\*) the numbers indicates sub-sections of interfaces.

unique features as a PC-based program. With more practical requirements through user tests, it will be upgraded to a more useful database program. It could include more public databases and could be integrated with analysis programs.

As WinBioDBs use MySQL as its database engine, the server version programs can be easily constructed by just changing the database server while using its Windows interfaces.

**References**

Barker, W.C., Garavelli, J.S., Huang, H., McGarvey, P.B., Orcutt, B.C., Srinivasarao, G.Y., Xiao, C., Yeh, L.S., Ledley, R.S., Janda, J.F., Pfeiffer, F., Mewes, H.W.,

Tsugita, A., and Wu, C. (2000). The protein information resource (PIR). *Nucl. Acids Res.* 28, 41-44.  
 Bairoch, A., Boeckmann, B., Ferro, S., and Gasteiger, E. (2004). Swiss-Prot: Juggling between evolution and stability. *Brief. Bioinform.* 5, 39-55.  
 Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., Studholme, D.J., Yeats, C., and Eddy, S.R. (2004). The Pfam Protein Families Database. *Nucl. Acids Res.* 32, D138-D141.  
 Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2004). GenBank: update. *Nucl. Acids Res.* 32, 23-26.  
 Bairoch, A. (2000). The ENZYME database in 2000. *Nucl. Acids Res.* 28, 304-305.  
 Hulo, N., Sigrist, C.J., Le Saux, V., Langendijk-Genevaux,

- P.S., Bordoli, L., Gattiker, A., De Castro, E., Bucher, P., and Bairoch, A. (2004). Recent improvements to the PROSITE database. *Nucl. Acids Res.* 32, D134-D137.
- Henikoff, J.G., Greene, E.A., Pietrokovski, S., and Henikoff, S. (2000). Increased coverage of protein families with the blocks database servers. *Nucl. Acids Res.* 28, 228-230.
- Kanehisa, M., Goto, S., Kawashima, S., and Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucl. Acids Res.* 30, 42-46.
- Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2007). REBASE-enzymes and genes for DNA restriction and modification. *Nucl. Acids Res.* 35, D269-D270.
- Schmid, C.D., Perier, R., Praz, V., and Bucher, P. (2006). EPD in its twentieth year: towards complete promoter coverage of selected model organisms. *Nucl. Acids Res.* 34, D82-D85.
- <http://www.mysql.com>.