

PDSB: Public Domain Software in Biology

G. P. s. Raghava

Institute of Microbial Technology, Chandigarh, India

Abstract

Public Domain Software in Biology (PDSB) is a collection of biological computer programs available in the public domain that have been created and maintained at IMTECH (Institute of Microbial Technology, Chandigarh, India). The collection consists of more than

400 computer programs of biological interest. A category has been assigned to each software package in the collection based upon its primary function. The aim of PDSB is to create a central repository of free software in the field of bioscience in order to serve the scientific community. A web server has been developed to search and download software from PDSB (imtech.chd.nic.in/pdsb). The collection is also accessible via anonymous FTP (imtech.chd.nic.in/pub) and by email (freeware@imtech.chd.nic.in). The server also helps the au

thors by providing a mechanism to conveniently distribute their software to a wide range of users. Software can be submitted via a form on the PDSB web site.

Introduction

Numerous software packages have been developed over the years to solve a variety of biological problems. These packages enable a user to perform various tasks in biology, such as sequence alignment, protein structure prediction, gene design, etc. (1). There are a number of public-domain computer programs available in the field of biology, which are available free of charge for academic users. These programs have been developed by individuals (academics, researchers, students, and others) who wish to provide a resource for the scientific community. In order to utilize the full potential of these resources, it is important to manage this fast-growing body of software (2).

Recognizing the importance of these public-domain software packages, the following archival sites have been created: 1) EMBL File Serve, 2) IUBIO archive and, 3) BioCatalog. The EMBL file server maintains free software and shareware in molecular biology, where one can submit software via e-mail and can access software packages via e-mail or FTP (2). The IUBIO archive is located at

Indiana University and has a large collection of molecular biology software (3). BioCatalog is a database of information on software of interest in the fields of molecular biology and genetics (4). There are a number of other sites that are collecting and maintaining public-domain software in various areas, including the medical sciences (5,6).

In order to manage and distribute public-domain software of biological interest, we have created a collection, or repository, called Public Domain Software in Biology (PDSB). This is a comprehensive database that contains a variety of software packages in the field of biology and information about these packages. A userfriendly interface has been developed that allows users to access data or download software from PDSB. It is created in PostgreSQL, a free RDBMS protocol for academic users. The database stores the following information about a program: 1) name of program, 2) category of program (based upon its function), 3) operating system, 4) main function, 5) description of program, 6) reference, 7) author of program, 8) hardware requirements, 9) software requirements, and 10) original FTP/HTTP site.

All software in the PDSB data base is classified into categories. The classification of software is based upon function. This classification helps the user in searching

for desired programs in an appropriate category. Besides the obvious advantage to the user community, the PDSB server also offers benefit to the authors of the software. This site provides a convenient method to rapidly distribute a program to a wide range of users.

Classification of Software

In order to assist users in finding software of interest, we categorize all the software in the database based upon function. It is often difficult to assign a piece of software to a single category because it can have multiple functions. Because of this, we assign software to primary and secondary categories. A software package can have one primary category and up to four secondary categories. A list of categories and their descriptions are given in Table 1.

Accessing or Downloading Software from PDSB

Web Server (HTTP Server)

A dynamic web server has been developed to access the database and repository online over the Internet (imtech.chd.nic.in/pdsb). Users can access information from this server either by search query or by category browsing. Search query allows a user to search the PDSB database by submitting a query to the database on any field (program name, function, description, keyword, etc.). It provides a number of options such as: 1) case-sensitive or insensitive search, 2) search only in a category or in all categories, 3) select the field to display the result, and 4) select the field on which the result is to be sorted. Category browsing allows a user to search the site for a desired program by browsing a

category. Selection of a category results in the display of all the programs in that category. This strategy is very helpful when a user does not know which program will be useful for their work. By visiting a desired category, users can easily identify programs that may be suitable for their needs.

Anonymous FTP Server

File transfer protocol (FTP) is an Internet service designed for file transfer. In order to provide FTP service from our server, we have created an anonymous account on our machine. This allows any user to download files from our server, at either

imtech.chd.nic.in or www.imtech.res.in.

E-mail Server

An e-mail server has been developed that allows for searching and downloading of software from the PDSB server via e-mail. Send an e-mail message to free-ware@imtech.chd.nic.in with the command "help" in the body message for detailed information on our e-mail server.

Distribution of Data on Magnetic Media

The aim of the PDSB database and repository is to distribute

Table 1: Category or directory used to classify the software packages in PDSB database repository.

| Directory/Category | Description |
|--------------------|---|
| 3DSD | 3-D Structure Determination (NMR CD and x-rays) |
| BRM CB | Biosensor Related Methods |
| COM | Classical Botany |
| CZM | Classification Of Molecules |
| DSD | Classical Zoological Methods |
| EMRM | Drug Scanning and Design |
| FPM | Electron Microscopic Related Methods |
| GDRM | Finger Printing Methods |
| GEM | Gene Design Redesign and Mapping |
| GSA | Gel Electrophoresis Methods |
| IRD IRM | Genome Sequence Analysis |
| MBD | Immunology Related Databases |
| MMBM | Immunology Related Methods |
| MRM | Miscellaneous Biological Databases |
| MSPM | Miscellaneous Molecular Biological Method |
| NDSS | Microbiology Related Methods |
| NSA | Miscellaneous Structure Prediction Methods Nucleotide |
| PDSS | Databases Searching and Submission Nucleotide |
| PLB | Sequence Analysis |
| PRM | Protein Databases Searching and Submission Protein |
| PSA | Ligand Binding |
| PSPT | Pattern Recognition Methods |
| PSS SFR | Protein Sequence Analysis |
| OTH | Protein Structure Prediction Techniques |
| | Protein Secondary Structure |
| | Sequence/Structure Function Relationship |
| | Other |

software to a wide range of users, including those users who do not have Internet connectivity or whose transfer speed is very slow. In order to provide off-line service to these users, we distribute software from the PDSB database via magnetic media. Users can send a request to obtain data on CDROM, DAT, or floppy-disk. We will then send the software by post. A catalog of the software in the repository is available on-line.

How to Submit a Software Package to the PDSB Database and Repository

The PDSB server allows authors to submit their public domain software packages for inclusion in the repository. In order to submit a computer program to the PDSB repository, a user needs to fill out a simple form (imtech.chd.nic.in/pdsb/submit.html). The submitter needs to provide full information about their computer program. The PDSB server will then assign a unique code for the program. Authors can upload their program from local disk to the PDSB site or they may send it by e-mail to pdsb@imtech.chd.nic.in. We encourage authors of programs to submit their program to the PDWSB database. Authors are requested to provide comprehensive documentation or on-line help and preferably the source code of their program.

Mirror Sites

Besides creating and maintaining biological software, we are also maintaining a mirror of sites that maintain public-domain soft

ware. We are maintaining mirrors of two sites of biological software and two sites of general-purpose software. The following is a description of these sites.

Biological Software

1. EBI Site: Mirror site: <ftp://www.imtech.res.in/pub/mirror/sites/ebi/>
Original site: <ftp://ftp.ebi.ac.uk/pub/software/>
2. IUBID Archive: Mirror site: <ftp://www.imtech.res.in/pub/mirror/sites/ebi/>
Original site: <ftp://www.imtech.res.in/pub/mirror/sites/iubio/>

General Software

1. GNU Site: Mirror site: <http://www.imtech.res.in:8084/>
Original site: <http://www.gnu.org/>
2. PostgreSQL site: Mirror site: www.imtech.res.in:8080/ Original site: www.postgresql.org/

Discussion

PDSB Server is intended to be a central repository for public domain software in the field of bioscience. Most of these software packages were collected from various sites that maintain public domain software (2,3). Our intention is to collect only freeware and not shareware or commercial software. If there is any software in our database that is not freeware, please write to us and we will remove it from the database and repository.

All of the software packages in

our software repository belongs to their respective authors. We are only assisting in their management and distribution by providing a distribution platform. We would greatly appreciate the submission of new programs directly from authors, in order to maintain high quality of software. For any queries, comments, or suggestions, please contact the author of this article at raghava@imtech.res.in.

Acknowledgements

The author is grateful to Dr. Amit Ghosh (Director, IMTECH), and Dr. J. R. Arora, (adviser, DBT, New Delhi), for their encouragement, support, and help in creation of this database. The author is also grateful to Ms. Tranum Kaur for her assistance in the collection of software.

References

1. Butler, B. (1994). *Curro Opinion Biotech.* 5, 19-23.
2. Fuchs, R. (1990). *Bioinformatics* 6, 120-121.
3. Gilbert, D. (2000). *Meth. Mol. Biol.* 132, 149-84.
4. Rodriguez-Tome, P. (1998). *Bioinformatics* 14, 469-70.
5. Longnecker, S.M. (1988). *Hosp. Pharm* 752, 748-50.
6. Pincirol, F., Combi c., Pozzi G. (1994). *Comput. Biol. Med.* 24, 243-54.

Address for Correspondence

Dr. G. P. S. Raghava
Scientist, Bioinformatics Centre
Institute of Microbial Technology
Sector 39A, Chandigarh, India
Phone: +91-172-690557
Fax: +91-172-690632
Email: raghava@imtech.res.in Web:
imtech.res.in/raghava