

## **Bioinformatics software resources**

### **Abstract**

This review looks at Internet archives, repositories and lists for obtaining popular and useful biology and bioinformatics software. Resources include collections of free software, services for collaborative development of new programs, software news media and catalogs of links to bioinformatics software and web tools. Problems with such resources arise from needs for continued curator effort to collect and update these, combined with less than optimal community support, funding and collaboration. Despite some problems, the available software repositories provide needed public access to many tools that are a foundation for analyses in bioscience research efforts.

### **Keywords**

bioinformatics software archives, web hyperlink catalogs, bioinformatics news

### **Introduction**

When the time comes to analyze results of a new experiment, bioscientists now supplement their tool set of spreadsheet, commercial bioinformatics and statistics programs, and word processor with bioinformatics tools located through the web. This is essential for many biology fields including sequence analysis, bio-data management, phylogenetics, gene expression, and proteomics. One can save time and reach more reliable conclusions by choosing the most appropriate analysis tools. Free bioinformatics tools are widely available, but it is not always easy to find the relevant ones, or even those that were available a few years ago. Web searches via Google, Yahoo and similar general search systems can miss the best of biology's unique and focused resources. This review highlights some of biology's current clearing-houses of informatics tools.

Internet resources that one can use to find suitable biology software fall into four general groupings: resource sites (archives, bioinformatics service organizations); web lists and catalogs pointing to other sources; news and discussion groups where information on new and updated software are announced; and publications. Table 1 summarizes a selection of these resources, which is by no means comprehensive. These were selected to represent some of the more comprehensive and/or actively managed resources for biology software.

Archives or repositories maintain collections of software from several authors. These serve both as long-term libraries of useful tools, and a reference for new software authors who want to avoid re-inventing wheels. Bioinformatics programs, including older ones especially with source code, are widely read and used as reference works and building blocks by other bioinformaticians in development of new software. Long-term archives provide continuing access when existence of authors' web or ftp sites is often limited to a few years<sup>5</sup>.

### **Resources**

#### **Archives and web tools**

When one wants to run software on a workstation, stock a bioinformatics service center with tools, or refer to related programs when developing new software, one needs to browse the web for programs. Large collections of programs are available through BioWeb at Pasteur, BioPortal at Weizmann Institute, European Bioinformatics Institute (EBI), IUBio Archive at Indiana University,

and RFSB at IMT India. The project repositories at Bioinformatics.org and SourceForge.net also offer bio-software with source code, many in active development.

IUBio Archive maintained by this author is one such long-term archive, serving public biology and bioinformatics software since 1989. It houses over 500 software titles, many of which have been added in recent years. The addition of EPrints.org<sup>1</sup> self-archiving web database for promoting author-contributed software will help expand this archive with minimal support. The Repository of Free Software in Biology (RFSB) at the Institute of Microbial Technology, India, collects and archives free, academic biology software<sup>4</sup>. This resource also offers long-term access to many popular biology programs.

Among those resources offering bioinformatics web tools, some of the more comprehensive include EBI, BioWeb Pasteur, and Canadian Bioinformatics Resource. There are a number of common bioinformatics analyses one can perform at these sites, including BLAST and sequence analyses, primer tools, and phylogenetics tree construction. EMBOSS sequence analysis package and SRS bio-database access are among the widely useful web tools available at these and other resource sites.

### **Web lists and catalogs**

There are numerous web lists of bioinformatics resources, with many aimed at the biologist looking for software. These are successors to last decade's popular bioinformatics lists by Keith Robison<sup>2</sup> and Pedro Coutinho<sup>3</sup>. Some of these, such as Bioinformatics.net, include discussion forums on use of biology software. These are useful for biologists, as well as bioinformatics engineers looking for tools related to their work, or to be used at service centers. Many of these share a similar organization by functional categories, with many of the same links. It is useful to compare these for their different editorial perspectives, e.g. genomics/molecular biology or proteomics/biochemistry, as well as effort to update and remove obsolete links. General resources such as Google, Amazon's Alexa, and Open Directory Project at Mozilla.org include biology and bioinformatics categories in their directories. These directories are populated by robots or from submissions; they tend to lack the comprehensiveness of biologist-maintained lists.

Bioinformatics.ca provides a curated list of links that are well organized in categories, with main sections that include human genome and model organisms, sequences, gene expression, education and computer related resources. Most or all of these include useful editorial comments on the content and value of the linked resources, making this list especially useful in learning about resources. The GenomeWeb at MRC, UK offers a similar very useful catalog of links with editorial abstracts. An interesting function at Bioinformatics.ca is provided by an XML standard for web news called RSS, for sharing bioinformatics links. This allows customers and other web sites to have computable access to this catalog. For instance, you can use an RSS program to notify you of additions and changes to this catalog.

The BioNetbook project at Pasteur Institute provides an example of resource lists that are searchable by several bioinformatics criteria: Biological Domain, e.g. sequence analysis or structural biology, Resource type, e.g. database or online analysis tools, and Organism. This biology-focused search engine proves especially useful in finding that tool or resource most relevant to one's research. This project also has implemented link maintenance by using semi-automatic scanning of Internet news and resources (robot-like) to update the catalog. A similar project is BioHunt, which uses Internet robot technology to search and update molecular biology resources. BioHunt maintains current entries (it shows update times of this review month for several searches), making it especially useful to find new or updated tools that one has heard of, but lacks curated cataloging of these to make it easy to find by subject matter.

**Table 1. Sources for biology software**

Source	Web URL	Notes
<i>Selected bioinformatics resource sites</i>		
Bioinformatics.org	<a href="http://bioinformatics.org/">http://bioinformatics.org/</a>	bioinformatics developer repository
BioPortal, Weizmann Institute	<a href="http://biportal.weizmann.ac.il/">http://biportal.weizmann.ac.il/</a>	software and data archive, web tools
BioWeb, Pasteur Institute	<a href="http://bioweb.pasteur.fr/">http://bioweb.pasteur.fr/</a>	software and data archive, web tools
Canadian Bioinformatics Resource (CBR)	<a href="http://cbr-rbc.nrc-cnrc.gc.ca/">http://cbr-rbc.nrc-cnrc.gc.ca/</a>	web tools
European Bioinformatics Institute	<a href="http://www.ebi.ac.uk/">http://www.ebi.ac.uk/</a>	web tools, data and software archive, catalog
IUBio Archive, Indiana University	<a href="http://iubio.bio.indiana.edu/">http://iubio.bio.indiana.edu/</a>	software and data archive, catalog, web tools
RFSB, Bioinformatics Centre, IMT	<a href="http://imtech.res.in/pdsb/">http://imtech.res.in/pdsb/</a>	software archive
SourceForge.net	<a href="http://sourceforge.net/">http://sourceforge.net/</a>	general developer repository with bioinformatics section
<i>Selected web lists and catalogs</i>		
BioHunt, ExPASy	<a href="http://www.expasy.org/BioHunt/">http://www.expasy.org/BioHunt/</a>	automatic robot updated list
Bioinformatics.ca	<a href="http://www.bioinformatics.ca/-links_directory">http://www.bioinformatics.ca/-links_directory</a>	curated list
Bioinformatics.net	<a href="http://www.bioinformatics.net/">http://www.bioinformatics.net/</a>	curated list
Bioinformatik.de	<a href="http://www.bioinformatik.de/">http://www.bioinformatik.de/</a>	curated list
BioNetbook, Pasteur Institute	<a href="http://www.pasteur.fr/recherche/-BNB/bnb-en.html">http://www.pasteur.fr/recherche/-BNB/bnb-en.html</a>	semi-automatic updates
CSM Molecular Biology Resource, SDSC	<a href="http://restools.sdsc.edu/">http://restools.sdsc.edu/</a>	curated list
GenomeWeb	<a href="http://www.hgmp.mrc.ac.uk/-GenomeWeb/">http://www.hgmp.mrc.ac.uk/-GenomeWeb/</a>	curated list
Open Directory Project	<a href="http://dmoz.org/Science/-Biology/Bioinformatics/">http://dmoz.org/Science/-Biology/Bioinformatics/</a>	submitted links
SouthWest Biotechnology and Informatics Center	<a href="http://www.swbic.org/">http://www.swbic.org/</a>	curated list
<i>News and discussion groups</i>		
BIOSCI/Bionet	<a href="http://www.bio.net/">http://www.bio.net/</a>	biologist and bioinformatics focus
Bioinformatics.org	<a href="http://bioinformatics.org/">http://bioinformatics.org/</a>	bioinformatics focus
Bioinformatics.net	<a href="http://bioinformatics.net/">http://bioinformatics.net/</a>	biologist focus
<i>Selected bioinformatics publications for software tools</i>		
BioInform	<a href="http://www.bioinform.com/">http://www.bioinform.com/</a>	bioinformatics news briefs
Bioinformatics	<a href="http://bioinformatics.-oupjournals.org/">http://bioinformatics.-oupjournals.org/</a>	bioinformatics focus, see application notes for biologists
BMC Bioinformatics	<a href="http://www.biomedcentral.com/-bmcbioinformatics/">http://www.biomedcentral.com/-bmcbioinformatics/</a>	bioinformatics focus
Briefings in Bioinformatics	<a href="http://www.henrystewart.com/-journals/bib/index.html">http://www.henrystewart.com/-journals/bib/index.html</a>	biologist & bioinformatics focus

Bioinformatics.net is a catalog of online biology resources, specializing in bioinformatics tools. Its focus is toward the needs of molecular biologists and life science professionals, more than for bioinformaticians, and includes discussion and help forums on the use of software and bioscience topics. Jonathan Rees, who developed this resource, also curates biology lists in the Open Directory Project. This service is supported in part by advertising, as are others reviewed here, one of the limited options available to maintain such services. Bioinformatik.de offers a similar directory style collection of curated bioinformatics and biology resource links. The CMS molecular biology resource is an extensive catalog of biology resources, including software tools. The SouthWest Biotechnology Center also maintains a useful catalog covering a broad range of biology resources.

Bioinformatics.org and SourceForge.net are resources that support software developers and bioinformatics engineers, but also are useful also to biologists looking for tools. Open-source software development in bioinformatics and other fields is being invigorated through agencies such as these. The number of active, widely used and valuable bioinformatics projects at these services is growing, including Generic Model Organism Database, Gene Ontology, GeneX Gene Expression Database, and Staden Package for sequence analysis. These agencies allow for software archiving, but the primary attractions to software developers are infrastructure and tools that enable collaborative software development. A historical archive or catalog service of bioinformatics software is limited, and maintenance of software releases is left to developers using this service.

## **News and publications**

Sources for announcements of software include the fifteen-year-old BIOSCI/Bionet public forums of [bionet.software](http://bionet.software), [bionet.software.www](http://bionet.software.www), [bionet.biology.computational](http://bionet.biology.computational) and [bionet.announce](http://bionet.announce) news groups. The Bionet groups have been overshadowed in recent years by web discussion forums, such as those at [Bioinformatics.org](http://Bioinformatics.org) and [Bioinformatics.net](http://Bioinformatics.net). They have a core value by carrying openly accessible and widely distributed biology software news and discussion, including through web portals such as [www.bio.net](http://www.bio.net) and Google groups. Other sources include the growing number of bioinformatics paper and electronic publications, and web list/forum sites devoted to biology and bioinformatics. Distinctions made between biologist and bioinformatics focus are fuzzy, but suggest the tendency of resources to address needs of those with one or the other field of primary training.

This journal Briefings in Bioinformatics covers a range of new software and bioinformatics projects, generally with a focus that spans the interests of bioscientists and bioinformaticians. The journal *Bioinformatics*, originally *CABios* in the 1980s, is the longest running publication for original papers in this field. It includes a useful section of short application notes, where many new programs are announced. The weekly newsletter *BioInform* has a short section of recent new releases and updates to software and database projects in biology, along with in-depth news articles, often focused on biotechnology and pharmaceutical industry issues. *BMC Bioinformatics* is an electronic journal that requires no subscriber fees. It publishes a range of original reports on new bioinformatics software, with open access for their redistribution.

## **Discussion**

There are numerous, well-developed bioinformatics resources that bioscientists can use to find the best available research tools. All of those discussed here have particular values; the interested scientist will find time well spent browsing through some of these to find a particular tool. The most useful are those that have a high level of editorial comment and curation by biologists; these also tend to become out-of-date as funding for maintenance declines. The *BioHunt* and *BioNetbook* projects show that robot automation can add significant currency and reduce curation effort to maintaining bioinformatics catalogs and search services.

Aside from the publications listed in Table 1, these resources are non-commercial, usually maintained by bioscientists and bioinformaticians, and are generally supported by an institution. A mixture of government support and growing biotechnology industry advertising supplements these. Although several of these have been in existence for many years, they share with a larger group of such resources an uncertain future. As Wren<sup>5</sup> noted in his survey of the longevity of access to URLs published in Medline abstracts (many of these for bioinformatics-related publications), 20% or more disappear in less than a decade. With a yearly loss in access, older software URLs are likely to be unavailable. Often bioinformatics tools remain useful beyond the support an individual author or group can provide, especially those designed for specific problems that don't attract new development. Many of these are used by other bioinformaticians as reference works for development of new software.

One major problem in maintaining software archives is collecting the software. All of the noted biology collections have been the effort of a few archivists, rather than author contributions. For several years, the journal *Bioinformatics* recommended IUBio Archive as a repository for published software, but this did not lead to any notable increase in author contributions. The issue of long-term software archiving is one that might be better handled with an institutional library paradigm such as maintains books, journals and other science reference material. The current archives are dependent on enthusiasm and dedication of those individuals who see value in contributing their time in the face of limited institutional and agency support.

Bioinformatics is maturing as an academic discipline, and the field's journals are catching up to other electronic media for timely release of news and software announcements, in addition to becoming the preferred route of authors for such announcements. Merging into one venue these functions of software archiving, news and discussion, along with author publication reports, would make much sense for future bioinformatics software publication. This may be forth-coming, though a growing influence of traditional academic publishing modes, dispersion of these functions among several institutions, and limited funding, all must be addressed by any project to improve biology software publication and access.

**Acknowledgments.** This work is supported in part by NSF grant 0090782 and NIH grant 1R01HG002733-01 to D. Gilbert.

*Don Gilbert  
Biology Department  
Indiana University,  
Bloomington, Indiana 47405 USA  
Tel: +1 812 855 0587  
E-mail: gilbertd@indiana.edu*

## **References**

1. EPrints.org, Self-Archiving and Open Archives organization. URL: <http://www.eprints.org/>
2. Robison, K. 1993. WWW Virtual Library for Biosciences. URL: [http://web.archive.org/web/\\*/http://golgi.harvard.edu/biopages.html](http://web.archive.org/web/*/http://golgi.harvard.edu/biopages.html) (See also <http://mcb.harvard.edu/BioLinks.html>, <http://vlib.org/Biosciences.html>)
3. Coutinho, Pedro, 1995. Pedro's Biomolecular Research Tools, URL: [http://www.public.iastate.edu/~pedro/research\\_tools.html](http://www.public.iastate.edu/~pedro/research_tools.html)
4. Raghava, G. P. S. (2001) PDSB: public domain software in biology. *Biotech Software and Internet Report* 2:154-156
5. Wren, J.D., 2004. 404 not found: the stability and persistence of URLs published in MEDLINE. *Bioinformatics*, Vol. 20:668-672. DOI: 10.1093/bioinformatics/btg465