Original Research Paper

BioInfoKnowledgeBase: Comprehensive Information Resource for Bioinformatics Tools

[1]**Ratna Prabha**, [1]**D.P. Singh and** [2]**Anil Rai**

[1]*ICAR-National Bureau of Agriculturally Important Microorganisms,*
Indian Council of Agricultural Research, Kushmaur, Maunath Bhanjan 275103 India
[2]*Center for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute,*
Indian Council of Agricultural Research, Library Avenue, New Delhi 110012 India

Abstract: Bioinformatics is revolutionizing biological sciences and facilitating every area of biology. Right from the start of any biological project, information regarding availability of resources or databases appropriate for the analysis of data repository are becoming essentially required. Biological scientists use to browse across the web for searching tools/softwares and to identify most appropriate one for analysis purpose. Time of any project can be saved and efforts will go in the right direction if the most appropriate analysis tools are selected. With the idea of putting maximally used data resources at a single place, we developed BioInfoKnowledgeBase (BIKB), a comprehensive information resource on softwares and databases applicable for various biological analyses. This database provides information of more than 1000 bioinformatics softwares grouped in six major biological themes and categorized in thirty-one different categories along with information of ~100 biological databases and public repositories. BioInfoKnowledgeBase is online accessible at http://webapp.cabgrid.res.in/BIKB/edb_home.html.

Keywords: Bioinformatics, Information Resources, Softwares, Tools, Online Analysis

Introduction

In the past few years, bioinformatics has emerged as young discipline in the information technology and facilitated biologists across the globe for storage, management and analysis of biological data. This further helped in search and retrieval of information about various databases and softwares available over internet and allowed them for the exchange of information for various kind of analyses (Hagen, 2000; Zhang *et al*., 2010). Bioinformatics is the science of integration, management and analysis of data in genomics, proteomics and in numerous other areas of biological sciences (Tripathi, 2000; Ouzounis, 2002). With the upcoming biological experiments and projects during the last several years, there is an increase in number of bioinformatics tools and databases and this is still progressing at a fast pace (Bansal, 2005; Gilbert, 2004). Selection of the most relevant one or even those most frequently used is a tough task. In general, researchers prefer various web search engines e.g., Google, Yahoo etc. for accomplishing this task, though these search engines provide only limited information in this context (Neerincx and Leunissen, 2005). Furthermore, most of the biological data interpretation requires a series of tools/software rather than a single one, e.g., genome sequencing projects require softwares for quality check, trimming, base calling, assembly, gene prediction etc. alongwith databases for final sequence submission or similarity searching (Searls, 2000; Kanehisa and Bork, 2003; Singh *et al*., 2012). This is why there is need to provide a platform from where information about multiple tools, softwares, methods and databases can be accessed easily along with the information about their possible applications.

Across the internet resources, information about biological softwares can be grouped into four general categories, resource sites, web lists, different forums and groups and lastly publications. The available tools are generally confined to a specific field of biology and most importantly including tools largely as a list hardly categorized by analysis needs and even categorized in broad categories only (Gilbert, 2004). Some of the existing important information resource for bioinformatics tools and softwares are NCBI Biocomputing Resources at

NCBI (http://www.ncbi.nlm.nih.gov/), European Bioinformatics Institute (EBI) (http://www.ebi.ac.uk./, Rodriguez-Tomé *et al*., 1996; Lopez, 2008), The Online Bioinformatics Resources Collection (http://www.hsls.pitt.edu/guides/genetics/obrc; Chen *et al*., 2007), Bioinformatics Links Directory (http://bioinformatics.ca/links_directory/, Brazas *et al*., 2011), ExPASy (http://www.expasy.org/, Artimo *et al*., 2012) and IntoPub (Jung *et al*., 2011). Among these, the Bioinformatics Links Directory represents the most comprehensive resource although it contains tools in broad categories and misses categories like metagenome analysis and information about tools generally used for various analysis along with details of databases (specifically microbial databases).

Here, we present a database BioInfoKnowledgeBase (BIKB) which was developed with a mandate to facilitate researchers with comprehensive information about numerous tools, softwares, methods, databases and their possible applications at a single place. This database provides information about variety of tools used in different kinds of biological analyses and provides details about numerous databases/public repositories on biological information. It covers varied aspects of bioinformatics along with a snapshot of computation involved in it. It also provides list of tools categorized as per the tasks accomplished and facilitates options to researchers so that they can choose any particular tool depending on their choice of analysis and usage. BIKB is developed in an easily accessible way, includes tools for numerous biological areas and also includes databases on various aspects like general, nucleotide, protein etc along with databases on bacteria, cyanobacteria and fungi.

Materials and Methods

Source of Information

Information about bioinformatics tools and softwares for analysis was searched and identified from the internet. Webserver issue of Nucleic Acid Research served a major source. Once tools were collected, they were grouped as per the biological fields they cover and then subdivided as per their analytical categories (Table 1).
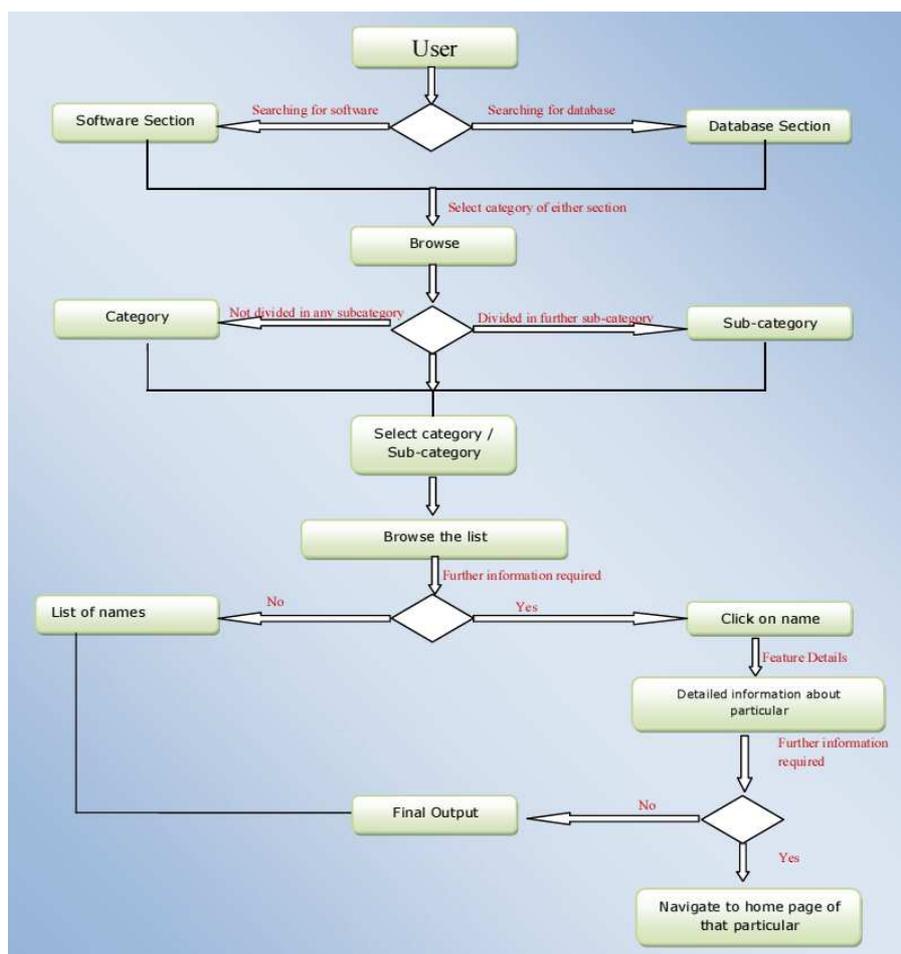

Fig. 1. Information flow in BioInfoKnowledgeBase (BIKB)

Table 1. List of the tools/databases (categorized as per their functions) at BioInfoKnowledgeBase (BIKB)

S. No.	Category	Sub-category	No. of softwares/ databases
1. Softwares			
1.1	Genome Annotation	Genome annotation (general tools)	29
		Gene Prediction	64
		Genome Assembly	20
		Ontology Analysis	05
1.2	Genome Analysis	Genome Analysis (general tools)	16
		Metagenome Analysis	19
		Genome Alignment	19
		Genome Browser	13
1.3	Gene Analysis	Gene Analysis (general tools)	11
		Gene Regulation	92
		Transcription Element Analysis	64
		Splicing	16
		Expression Analysis	203
		Pathway Analysis	07
1.4	Nucleotide Analysis	Nucleotide Analysis (general tools)	50
		Motif identification	10
		Nucleotide Alignment	24
		Genome Alignment	19
		Probe and Primer designing	29
		Restriction Enzyme Mapping	16
		SNP identification	19
		Structure and Sequence Detection	91
1.5	Protein Analysis	Protein Sequence Analysis (general tools)	03
		Secondary Structure and Transmembrane Segment Prediction	09
		Super Secondary Structure Prediction	19
		3D-motif Analysis	10
		Tertiary Structure and Fold Prediction	32
		MHC peptide prediction	14
		Modelling and Simulations	20
		Structure Visualization	09
		Functional Pattern Recognition	04
1.6	Evolutionary Analysis	--	55
2. Databases			
2.1	General Databases	--	25
2.2	Nucleotide Databases	Sequence Submission and Retrieval System	09
		Genome Sequence Databases	16
2.3	Protein Databases	Protein sequence and structure Databases	13
		Specialized Databases for Proteins	12
2.4	Bacterial Databases	--	09
2.5	Cyanobacterial Databases	--	09
2.6	Fungal Databases	--	10

Database Development and Architecture

Database was designed using WAMP technology where Windows served as the platform (http://windows.microsoft.com/) and Apache was used for server (http://www.apache.org/). Database designing was performed using MySQL relational database management system (http://www.mysql.com/). For designing of web interface HTML and PHP scripting languages were used where HTML served the purpose of static webpage generation and PHP for dynamic webpage generation. Information flow in BIKB is endowed with a glimpse on the database functioning and information extraction (Fig. 1).

Results

Web Interface of the BIKB Database

BioInfoKnowledgeBase is hosted at the server of National Agricultural Biocomputing Portal and can be accessed via web-link http://webapp.cabgrid.res.in/BIKB/edb_home.html. This user friendly database contains information of more than 1000 bioinformatics softwares of various aspects such as genome annotation, gene regulation, expression analysis, protein structure prediction, modeling and metagenome analysis etc. It also includes information of ~100 databases of various areas. BIKB covers softwares

required for varied bioinformatics analyses and is practically applicable for the purpose of information extraction and data analysis for persons working in the area of biosciences and bioinformatics and is best suitable for the beginners.

Database contains six main sections 'Home', 'About us', 'Softwares', 'Databases', 'Search' and 'Team'. Section 'Home' is the main page of this database that provides brief information about database. Section 'About us' includes the information about the institute and the project under which this database was developed. Users can access information about numerous softwares required for various kinds of biological analysis from section 'Softwares' (Fig. 2). BIKB also provides glimpse of different categories of databases in the section 'Databases'. Section

'Search' facilitates users to search for software or database as per their need. 'Team' section provides brief information about the group who developed this database. Sidebar contains two major sections 'Software' and 'Databases' under which many subsections are available from which users can directly navigate to the page containing information about software/database of specific category. 'Software' section includes six different kind of analysis as sub-sections like Genome Annotation, Genome Analysis, Gene Analysis, Nucleotide Analysis, Protein Analysis and Phylogenetic Analysis whereas 'Databases' section includes General Databases, Nucleotide Databases, Protein Databases, Bacterial Databases, Cyanobacterial Databases and Fungal Databases as sub-sections (Table 1).
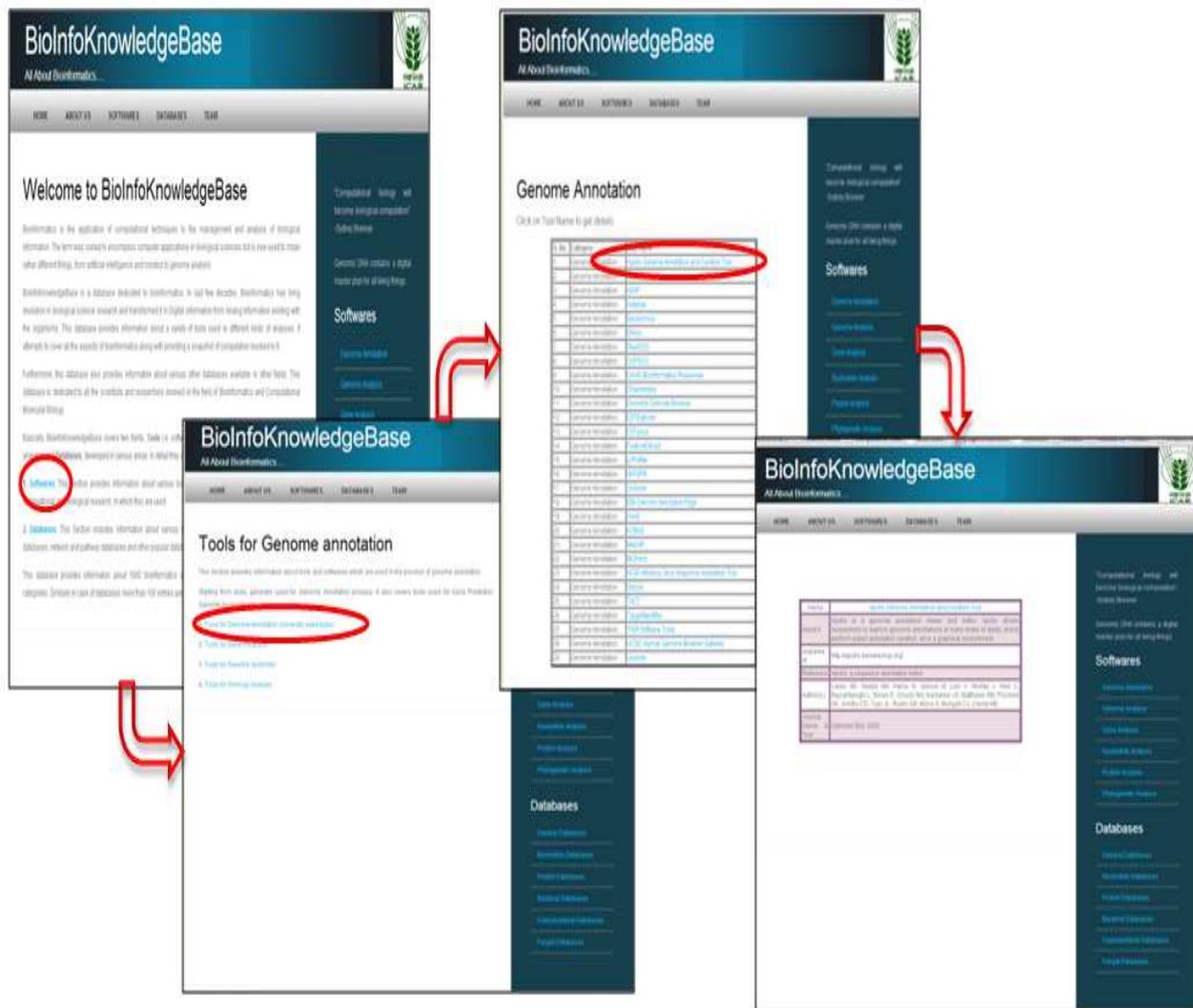

Fig. 2. Snapshots showing navigation in BioInfoKnowledgeBase (BIKB) during extraction of information for any particular software (red circles include the hyperlinks while red arrows represents navigation)

Overall BIKB serves as information resource for softwares of six major biological areas i.e., Genome Annotation, Genome Analysis, Gene Analysis, Nucleotide Analysis, Protein Analysis and Evolutionary Analysis which are further sub divided in 31 different categories, except evolutionary analysis (Genome Annotation: 4 sub-categories; Genome Analysis: 4 sub-categories; Gene Analysis: 6 sub-categories; Nucleotide Analysis: 8 sub-categories; Protein Analysis: 9 sub-categories) (Table 1). By a single click users can directly access list of tools available for any of the above mentioned biological categories.

Conclusion

With the advancements in computational capacity and technological availability, the number of bioinformatics resources are increasing with a great pace. Availability of various kinds of bioinformatics tools/softwares with different algorithm and application can create a major problem for researchers requiring appropriate tools for bioinformatics analysis and interpretation of massive dataflow flooding across the globe. Among the users of such biological resources are numerous young researchers who have started beginning of their career in the field of molecular biology, life science and bioinformatics. Considering the need of a suitable platform hosting information about the softwares/tools employed for biological data analysis, we developed BioInfoKnowledgeBase (BIKB). This information resource comprehensively incorporates details of various databases and help researchers that are either in the initial phase of their analysis or those who need to compare different softwares available for similar comparative interpretation. It will bestow researchers with a comprehensive list of softwares and their details and ease their research and will assist biologists in better interpretation of their work by allowing them to choose the most appropriate tools for their research and will eventually save their time and efforts in the analysis and interpretation of biological data.

Acknowledgement

Author's Contributions

RP and DPS designed and developed the database. AR helps in online implementation of database. RP and DPS compiled manuscript and it was revised by AR.

Conflict of Interest

Authors declared no conflict of interest.

References

Artimo, P., M. Jonnalagedda, K. Arnold, D. Baratin and G. Csardi *et al*., 2012. ExPASy: SIB bioinformatics resource portal. Nucl. Acids Res., 40: W597-W603. DOI: 10.1093/nar/gks400

Bansal, A., 2005. Bioinformatics in microbial biotechnology-a mini review. Microbial Cell Factories, 4: 19-19. DOI: 10.1186/1475-2859-4-19

Brazas, M.D., D.S. Yim, J.T. Yamada and B.F. Ouellette, 2011. The 2011 Bioinformatics Links Directory update: more resources, tools and databases and features to empower the bioinformatics community. Nucl. Acids Res., 39: W3-7. DOI: 10.1093/nar/gkr514

Chen, Y.B., A. Chattopadhyay, P. Bergen, C. Gadd and N. Tannery, 2007. The online bioinformatics resources collection at the university of pittsburgh health sciences library system-a one-stop gateway to online bioinformatics databases and software tools. Nucl. Acids Res., 35: D780-D785. DOI: 10.1093/nar/gkl781

Gilbert, D., 2004. Bioinformatics software resources. Brief. Bioinform., 5: 300-304. DOI: 10.1093/bib/5.3.300

Hagen, J.B., 2000. The origins of bioinformatics. Nat. Rev. Genet., 1: 231-236. DOI: 10.1038/35042090

Jung, D., J.H. Kim, S. Lee and B. Lee, 2011. IntoPub: A directory server for bioinformatics tools and databases. IBC, 3: 1-4. DOI: 10.4051/ibc.2011.3.3.0012

Kanehisa, M. and P. Bork, 2003. Bioinformatics in the post-sequence era. Nat. Genet., 33: 305-310. DOI: 10.1038/ng1109

Lopez, R., 2008. Biological data resources at the EMBL-EBI. Rev. Colomb. Biotecnol., 10: 120-128.

Neerincx, P.B. and J.A. Leunissen, 2005. Evolution of web services in bioinformatics. Brief. Bioinform., 6: 178-188. DOI: 10.1093/bib/6.2.178

Ouzounis, C., 2002. Bioinformatics and the theoretical foundations of molecular biology. Bioinformatics, 18: 377-378. DOI: 10.1093/bioinformatics/18.3.377

Rodriguez-Tomé, P., P.J. Stoehr, G.N. Cameron and T.P. Flores, 1996. The European Bioinformatics Institute (EBI) databases. Nucl. Acids Res. 24: 6-12. DOI: 10.1093/nar/24.1.6

Searls, D.B., 2000. Bioinformatics tools for whole genomes. Ann. Rev. Genom. Hum. Genet., 1: 251-279. DOI: 10.1146/annurev.genom.1.1.251

Singh, D., R. Prabha, A. Rai and D. Arora, 2012. Bioinformatics-assisted microbiological research: Tasks, developments and upcoming challenges. Am. J. Bioinform., 1: 10-19.

Tripathi, K.K. 2000. Bioinformatics: The foundation of present and future biotechnology. Curr. Sci., 79: 570-575.

Zhang, W., F. Li and L. Nie, 2010. Integrating multiple 'omics' analysis for microbial biology: Application and methodologies. Microbiology, 156: 287-301. DOI: 10.1099/mic.0.034793-0