Paper 602: Unit III: B.Sc 3rd Year Botany Major (6th Semester)

Searching Biological data using Search Engine:

Biological databases are libraries of life sciences information, collected from scientific experiments, published literature, high-throughput experiment technology, and computational analysis. They contain information from research areas including genomics, proteomics, metabolomics, microarray gene expression, and phylogenetics.[2] Information contained in biological databases includes gene function, structure, localization (both cellular and chromosomal), clinical effects of mutations as well as similarities of biological sequences and structures.

Biological databases can be broadly classified into sequence, structure and functional databases. Nucleic acid and protein sequences are stored in sequence databases and structure databases store solved structures of RNA and proteins. Functional databases provide information on the physiological role of gene products, for example enzyme activities, mutant phenotypes, or biological pathways. Model Organism Databases are functional databases that provide species-specific data. Databases are important tools in assisting scientists to analyze and explain a host of biological phenomena from the structure of biomolecules and their interaction, to the whole metabolism of organisms and to understanding the evolution of species. This knowledge helps facilitate the fight against diseases, assists in the development of medications, predicting certain genetic diseases and in discovering basic relationships among species in the history of life.

Biological knowledge is distributed among many different general and specialized databases. This sometimes makes it difficult to ensure the consistency of

information. Integrative bioinformatics is one field attempting to tackle this problem by providing unified access. One solution is how biological databases cross-reference to other databases with accession numbers to link their related knowledge together.

Relational database concepts of computer science and Information retrieval concepts of digital libraries are important for understanding biological databases. Biological database design, development, and long-term management is a core area of the discipline of bioinformatics. Data contents include gene sequences, textual descriptions, attributes and ontology classifications, citations, and tabular data. These are often described as semi-structured data, and can be represented as tables, key delimited records, and XML structures.

dd

Nucleic Acids Research Database Issue:

An important resource for finding biological databases is a special yearly issue of the journal *Nucleic Acids Research* (NAR). The Database Issue of NAR is freely available, and categorizes many of the publicly available online databases related to biology and bioinformatics. A companion database to the issue called the Online Molecular Biology Database Collection lists 1,380 online databases. Other collections of databases exist such as MetaBase and the Bioinformatics Links Collection.

Access:

Most biological databases are available through web sites that organise data such that users can browse through the data online. In addition the underlying data is usually available for download in a variety of formats. Biological data comes in many formats. These formats include text, sequence data, protein structure and links. Each of these can be found from certain sources, for example:

- Text formats are provided by Pub Med and OMIM.
- Sequence data is provided by GenBank, in terms of DNA, and UniProt, in terms of protein.

- Protein structures are provided by PDB, SCOP, and CATH.

Species-specific databases

Species-specific databases are available for some species, mainly those that are often used in research (Model Organisms). For example, EcoCyc is an *E. coli* database. Other popular model organism databases include Mouse Genome Informatics for the laboratory mouse, *Mus musculus*, the Rat Genome Database for *Rattus*, ZFIN for *Danio Rerio* (zebrafish), PomBase for the fission yeast *Schizosaccharomyces pombe*, FlyBase for *Drosophila*, WormBase for the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*, and Xenbase for *Xenopus tropicalis* and *Xenopus laevis* frogs.

See the following also:

- Biobank
- Biological data
- Chemical database
- Death Domain database
- European Bioinformatics Institute
- Gene Disease Database
- Integrative bioinformatics
- List of biological databases
- Model organism databases
- NCBI
- PubMed (a database of biomedical literature)