

Chapter 4

Bioinformatics Database Resources

Icxa Khandelwal

Jaypee University of Information Technology, India

Aditi Sharma

Jaypee University of Information Technology, India

Pavan Kumar Agrawal

G. B. Pant Engineering College, India

Rahul Shrivastava

Jaypee University of Information Technology, India

ABSTRACT

Various biological databases are available online, which are classified based on various criteria for ease of access and use. All such bioinformatics database resources have been discussed in brief in this book chapter. The major focus is on most commonly used biological/bioinformatics databases. The authors provide an overview of the information provided and analysis done by each database, information retrieval system and formats available, along with utility of the database to its users. Most widely used databases have been covered in detail so as to enhance readers' understanding. This chapter will serve as a guide to those who are new to the field of bioinformatics database resources, or wish to have consolidated information on various bioinformatics databases available.

INTRODUCTION

The National Center for Biotechnology Information (NCBI) defines bioinformatics as: “the field of science in which biology, computer science, and information technology merge into a single discipline”. Bioinformatics can be considered an amalgam of three sub-disciplines:

1. Development of new algorithms as well as statistics so that the relationship between the elements of huge datasets can be determined.

DOI: 10.4018/978-1-5225-8903-7.ch004

2. Analysis as well as interpretation of biological data i.e. various types of sequences and structures.
3. Development of tools and software to ensure efficient access as well as management of biological data (Toomula, 2011).

The bioinformatics database resources focus primarily on the third sub-discipline of bioinformatics. A database can be defined as a computerized and organized storehouse of related information that provides a standardized way for searching, inserting and updating data. The data stored in these databases is persistent and organized. Database Management System (DBMS) is a software application that deals with the user, other applications, and the database itself in order to perform analysis and capture data in a systematic manner.

Bioinformatics databases or biological databases are storehouses of biological information. They can be defined as libraries containing data collected from scientific experiments, published literature and computational analysis. It provides users an interface to facilitate easy and efficient recording, storing, analyzing and retrieval of biological data through application of computer software. Biological data comes in several different formats like text, sequence data, structure, links, etc. and these needs to be taken into account while creating the databases.

There are various criteria on the basis of which the databases can be classified. On the basis of structure, databases can be classified as a text file, flat file, object-oriented and relational databases. On the basis of information, they can be classified as general and specialized databases. Most commonly, they are classified on the basis of the type of data stored in primary, secondary and composite databases (Kumar, 2005).

CLASSIFICATION OF DATABASES

Type 1

Databases can be classified on the basis of structure as Abstract Syntax Notation (ASN.1), Flat files, Object oriented databases, Relational databases, and XML. Table 1 provides a comparison of various types of databases on the basis of structure

- **ASN.1:** This format comprises of a syntax and description of how a particular data type can be represented physically in a data stream or sequential file (Buneman, Davidson, Hart, Overton, & Wong, 1995). This format has been adopted by NCBI for the representation of sequential data. It is one of the major file formats in GenBank (Cooray, 2012).
- **Flat Files:** This implementation is based on only one table, which incorporates the complete data i.e. all the attributes for each variable. Each row of the table specifies a different record. Specified delimiters are used to differentiate among records. Maintenance of data stored is a major drawback of this type of databases. Integration of two or more databases is difficult due to redundancy in data and variation in the format used.
- **Object Oriented Databases:** Object oriented databases can handle complex data types and can be easily integrated with Object Oriented Programming Languages (OOPL) (Codd, 1970). They can be defined as a collection of objects. Objects represent an instance of an entity and comprise attributes as well as methods (Hasegawa, 2008).

34 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/bioinformatics-database-resources/228619

Related Content

Removal of Toxic Pollutants From Soil Using Microbial Biotechnology

Anupam Pandey, Priyanka Harishchandra Tripathi, Satish Chandra Pandey, Vinay Mohan Pathak and Tapan Kumar Nailwal (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 1456-1476).

www.irma-international.org/chapter/removal-of-toxic-pollutants-from-soil-using-microbial-biotechnology/228678

Multi-Agent Systems in Three-Dimensional Protein Structure Prediction

Leonardo de Lima Corrêa and Márcio Dorn (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 1031-1068).

www.irma-international.org/chapter/multi-agent-systems-in-three-dimensional-protein-structure-prediction/228657

Subspace Clustering of DNA Microarray Data: Theory, Evaluation, and Applications

Alain B. Tchagang, Fazel Famili and Youlian Pan (2019). *Biotechnology: Concepts, Methodologies, Tools, and Applications* (pp. 210-264).

www.irma-international.org/chapter/subspace-clustering-of-dna-microarray-data/228625

Brain-Computer Interfaces for Control of Upper Extremity Neuroprostheses in Individuals with High Spinal Cord Injury

Rüdiger Rupp, Martin Rohmand Matthias Schneiders (2014). *Emerging Theory and Practice in Neuroprosthetics* (pp. 237-264).

www.irma-international.org/chapter/brain-computer-interfaces-for-control-of-upper-extremity-neuroprostheses-in-individuals-with-high-spinal-cord-injury/109892

Biosynthesis of Silver Nanoparticles for Study of Their Antimicrobial Effect on Plasma-Treated Textiles: Silver Coating of Plasma-Treated Cotton Fabric

Shazia Shukrullah, Muhammad Anwar, Muhammad Yasin Naz and Inzamam Ul Haq (2022). *Emerging Developments and Applications of Low Temperature Plasma* (pp. 149-166).

www.irma-international.org/chapter/biosynthesis-of-silver-nanoparticles-for-study-of-their-antimicrobial-effect-on-plasma-treated-textiles/294715