

Bioinformatics introduction to biological databases

Gulshan Kumar
School of Fisheries

Introduction

- Bioinformatics is an interdisciplinary field mainly involving molecular biology and genetics, computer science, mathematics, and statistics to address data intensive biological problems. **Data intensive, large-scale biological problems are addressed from a computational point of view.** The most common problems are modeling biological processes at the molecular level and making inferences from collected data.

- A bioinformatics solution usually involves the following steps:
 - • Collect statistics from biological data.
 - • Build a computational model.
 - • Solve a computational modeling problem.
 - • Test and evaluate a computational algorithm.

- **Biological databases**
- *A database* is a computerized archive **used to store and organize data in such a way that information can be retrieved easily via a variety of search criteria.**
- Types of databases
- In bioinformatics, and indeed in other data intensive research fields, databases are often categorised as primary or secondary.

- **Biological Databases**
- The data repositories more relevant to the biological sciences include:
 - nucleotide and protein sequences
 - protein structures
 - genomes
 - genetic expression
 - bibliography

- Main sequence databases:
- [NCBI](#) (The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.)
- [EMBL](#) (European Molecular Biology Laboratory)
- Main protein databases:
- [Uniprot](#) ([UniProt](#) is a protein database that includes information divided in two sections: Swiss-Prot and TrEMBL. UniProt aims to store sequence and functional information for the proteins.)
- [PDB](#) (Protein Data Bank) stores 3D structures for proteins and nucleic acids.)
- [MMDB](#)

- Bibliography:
- [Pubmed](#)
- [Web of Science](#)