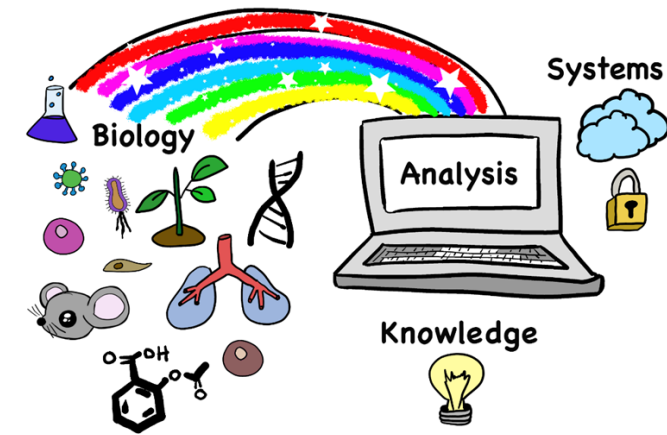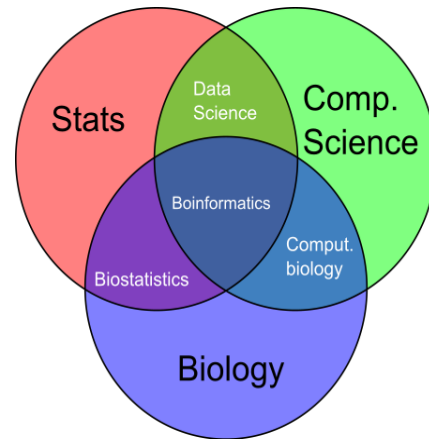# Introduction to Bioinformatics

Shatakshi Kulkarni

Definition

❑ Bioinformatics is a field of science in which biology, computer science and information technology merge into a single discipline

❑ It is an interdisciplinary field of extracting, storing, organizing, managing, and analyzing biological data to generate useful knowledge.

Need & Scope

❏ As the application of information technology to biology, bioinformatics pervades the whole of biology, including genetics, biochemistry, ecology and medicine.

❏ However, much of the publicity and emphasis that bioinformatics has received in the last few years has been on DNA and protein sequence analysis.

❏ Given the large amount of sequence data available and the rate at which it is growing, this is where the need for computer analysis has been felt the most.

❏ Huge advances in the fields of molecular biology and genomics especially after Human Genome Project. To handle massive amount of data.
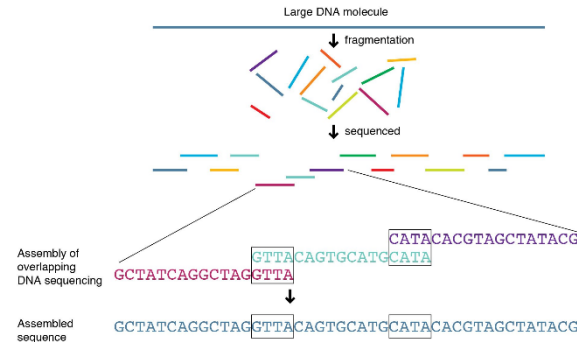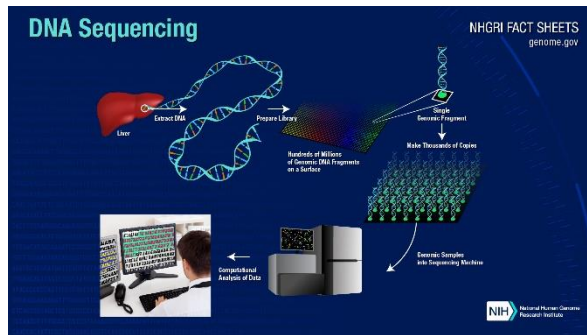
Applications
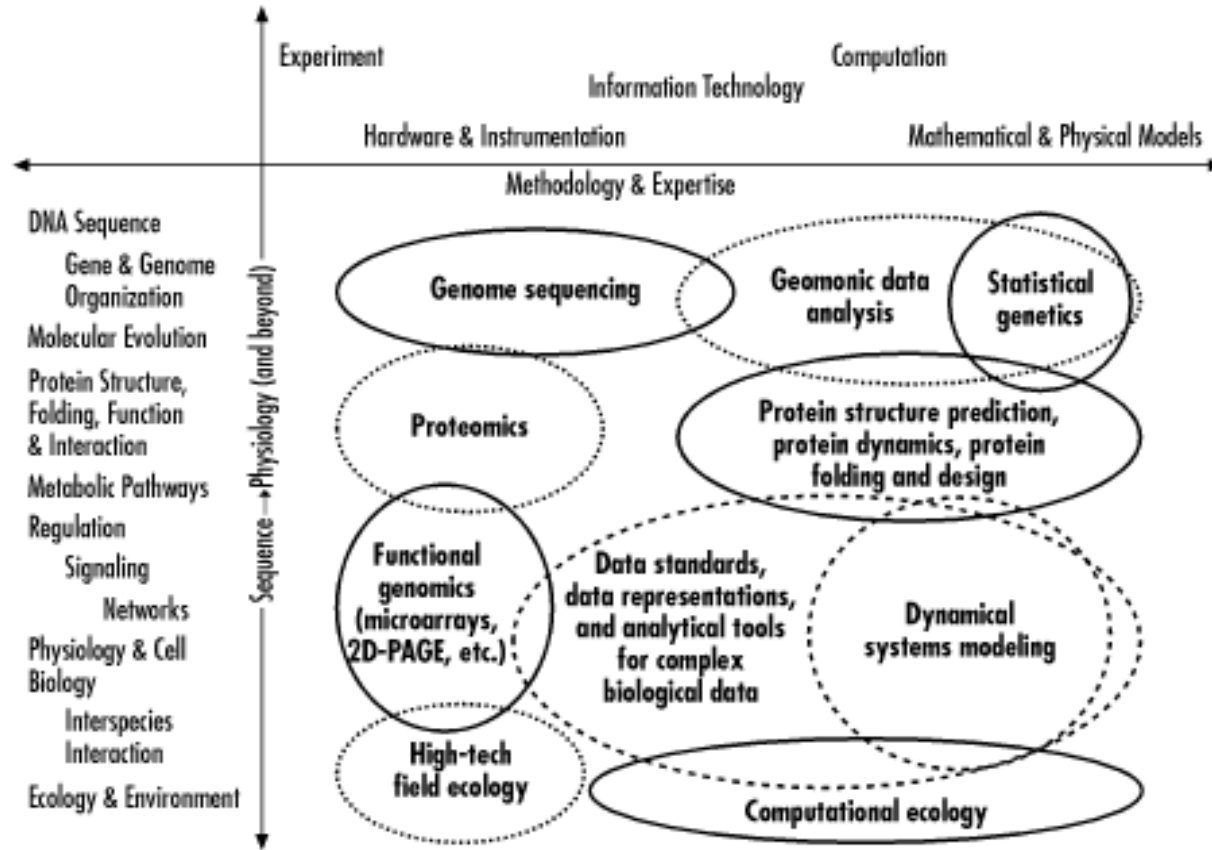
❑ Data Management:

1. The biological data like DNA sequences, protein sequences is stored, managed and retrieved
2. Biological Databases are present from where data can be retrieved
3. DBMS/Database management systems like SQL can be used

❑ Sequence & Genomic Data Analysis:

1. Analysis of DNA, RNA and Protein sequences
2. Genome sequencing (NGS), assembly (overlapping sequences) and annotation (identify coding regions)
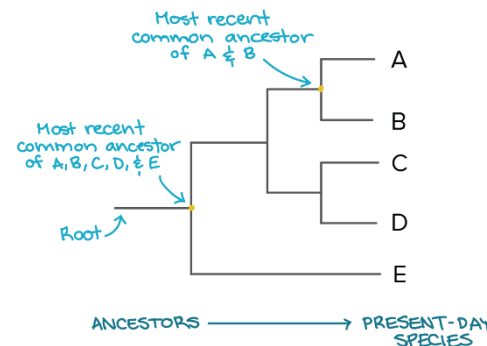

❑ Evolutionary Biology and Phylogenetic

1. Sequencing and assembly is followed by sequence mapping with existing data to find out similarities and differences
2. Comparative genomics and evolutionary analysis provide insights into genome evolution, gene duplication, and genetic variation across species

❑ **Structure Prediction:**
1. Bioinformatics plays a crucial role in predicting and analyzing the three-dimensional structures of proteins, nucleic acids, and other biomolecules.
2. Involves protein structure prediction, molecular modeling, protein-ligand docking, and analysis of protein-protein interactions.
3. Structural bioinformatics tools aid in understanding the function, dynamics, and interactions of biomolecules at the atomic level.

Sequence → Structure
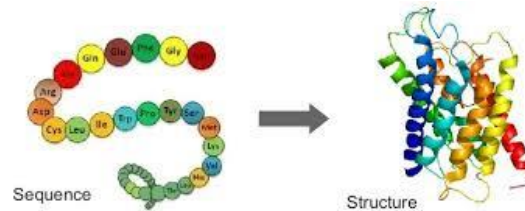
❑ **Drug Development:**
1. Target identification and validation by genome and pathway analysis
2. Biomarkers discovery
3. Modeling the relationship between chemical structure and biological activity to optimize drug candidates
4. Screening large libraries of compounds in silico to identify potential drug candidates

❑ **Nutrigenomics & Personalized medicine:**
1. Nutrigenomics is studying individual's genetic makeup influences their response to nutrients and how nutrients can affect gene expression
2. Personalized medicine (also known as precision medicine) is an approach to healthcare that tailors medical treatment to the individual characteristics of each patient, including genetic, environmental, and lifestyle factors

❑ **Health Informatics:**
1. Predictive analysis and risk prediction using Electronic Health records/EHR is studied in health informatics where ML and statistics play a crucial role
2. Disease prediction, biomarker discovery, clinical trials and research come under health informatics
3. Using bioinformatics to stratify patients based on genetic and molecular profiles, ensuring that clinical trials include relevant and homogeneous groups, which can lead to more accurate and reliable results

Applications

1. Medicine (Molecular, Personalized, Preventative) & Drug Development
2. Biotechnology
3. Crop and livestock improvement (Insect resistance)
4. Ecology and population studies
5. Microbial genome applications
6. Gene therapy
7. Antibiotic resistance
8. Evolutionary studies
9. Forensic analysis
10. Bio-weapon creation
11. Improve nutritional quality
12. Veterinary Science

Limitations

1. Biological redundancy

2. Lack of data

3. Errors in sequence alignment leading to error in structural analysis and phylogenetic analysis