

MST304: BIOINFORMATICS**UNIT I**

Introduction to computers and bioinformatics- Types of operating systems, concepts of networking and remote login, basic fundamentals of working with unix/Linux. Biological databases- Introduction to NCBI, NCBI data bases, BLAST, BLASTn, BLASTp, PSI-BLAST, modes of database search, mode of data storage (Flat file format, db-tables), flatfile formats of GenBank, EMBL, DDBJ, PDB. Sequence alignment –Concept of local and global sequence alignment, Pairwise sequence alignment, Structure alignment, STAMP: structural alignment of multiple proteinsscoring an alignment, substitution matrices, multiple sequence alignment.. Principle of Protein structure and conformational space, pfam (Protein family prediction).

UNIT II

Phylogenetic analysis- Basic concepts of phylogenetic analysis, rooted/uprooted trees, approaches for phylogenetic tree construction (UPGMA, Neighbor joining, Maximum parsimony, Maximum likelihood). Cluster analysis; Phylogenetic clustering by simple matching coefficients; Sequence Comparison; Sequence pattern; Regular expression based pattern; Theory of profiles and their use in sequence analysis; Hidden Markov models; Concept of HMMS; Baum-Welch algorithm; Use of profile HMM for protein family classification; Pattern recognition methods.

UNIT III

Methods for modeling: Homology modeling; Loop modeling, Comparative modeling, Threading, Refinement of model,Protein structure prediction; Structure comparison of macromolecules with reference to proteins; Force fields; Molecular energy minimization; Monte Carlo and molecular dynamics simulation, Protein Modeling, Molecular Simulations_basic information.

UNIT IV

Generation and analysis of high throughput sequence data- Assembly pipeline for clustering of HTGS data, format of “.ace” file, quality assessment of genomic assemblies, International norms for sequence data quality, Clustering of EST sequences, concept of Unigene.

Annotation procedures for high through-put sequence data- Identification of various genomic elements (protein coding genes, repeat elements, strategies for annotation of whole genome, functional annotation of EST clusters, gene ontology (GO) consortium.

UNIT V

Structure predictions for nucleic acids and proteins- Approaches for the prediction of RNA secondary and tertiary predictions, energy minimization and base covariance models, Basic approaches for protein structure predictions, comparative modeling, fold recognition/threading and ab-initio prediction. Drug Designing- Molecular Docking, Virtual Screening, ADMET analysis, click chemistry.

Text/References

1. Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins by Baxevanis A.D. and Ouellette, Third Edition. John Wiley and Son Inc., 2005.
2. Bioinformatics Sequence and Genome Analysis by Mount D.W., CSHL Press, 2004.
3. Introduction to Bioinformatics by Tramontano A., Chapman & Hall/CRC, 2007.
4. Understanding Bioinformatics by Zvelebil, M. and Baum, Chapman & Hall/CRC, 2008.