

Course name : Biological Sequence Analysis and Algorithms

Course Code : IT453

Credit : 4 credits

Course Offered to : M.Sc III Semester

Course Description:

This course will provide the necessary introduction about the sequence data analysis carried out by biologists in general and also will cover few important computational algorithms in that context. Emphasis will be given to computational methods employed to identify/characterize/analyse various biological features.

Pre-requisite : Nil

Pre-requisite (Desirable): Idea of very basic concepts in molecular biology especially with gene and genome sequence organisation and its features.

Course outcomes (CO):

1. Students will understand and grasp the rationale behind the computational methods employed for various biological problems
2. Students will be motivated to undertake small research projects dealing with sequence data
3. This course may expose them to the latest and challenging computational methods/algorithms currently undertaken by researchers in the domain of sequence data analysis

Tentative Course Plan:

Introduction and Overview of Bioinformatics , Survey of both primary and secondary Databases relating to sequences, structures, gene expression, and regulatory and signal pathways (3 hours)

Overview of Sequence data analysis in general – Software tools and web servers (3 hours)

Sequence alignments algorithms : Scoring matrices and scoring functions – pairwise alignment local and Global – Multiple sequence alignment – Database search – BLAST versions (6 hours)

Statistics and search method for Sequence patterns, profiles, motifs etc – MEME, Weight matrix, Profile, sequence Logo etc (3 hours)

Gene Identification problem : Codon usage, fourier and markov models (3 hours)

Overview of advanced concepts : Information theory, machine learning and probabilistic modelling (3 hours)

Phylogenetic analysis: Concept of distance – Distance, parsimony and likelihood methods (4 hours)

RNA secondary structure – RNA folding algorithms and RNA databases (2 hours)

Computational Genomics – current topics (3 hours)

Total Theory	30 hours
Tutorials	10 hours

References:

Text Book:

1. Durbin et al (2003). Biological sequence analysis : probabilistic models of Proteins and Nucleic acids. Cambridge University Press (Indian Edition)

Reference Book:

1. Jonathan Pevesner . Bioinformatics and Functional Genomics (3rd Edition) Wiley Blackwell
2. Handbook of Computational Molecular Biology Ed by Srinivas Aluru. (2005). Indian Edition Chapman & Hall/CRC