Course No.	Title of the Course	Course Structure	Pre-requisite
BTBTC14	Bioinformatics	3L - 0T - 2P	None

COURSE OUTCOMES (COs):

The student will be able to:

1. Describe the process of development of Bioinformatics as an amalgamation of Biology, Computer Science, Statistics and Internet Technologies.
2. Summarize the various types of basic biological data, and associated databases/tools.
3. Apply various Bioinformatics algorithms for storing, searching, determining, predicting and comparing molecular information.
4. Interpret and evaluate the significance and confidence of biological information obtained from use of various Bioinformatics tools.
5. Formulate novel Bioinformatics solutions at every step from gene to protein function.

COURSE CONTENT:

MODULE-I

Introduction to Bioinformatics: Human genome project; genome sequencing and assembly; Mapping – physical and genetic maps; The NCBI data model: Organization and salient characteristics

MODULE-II

Biological databases-primary sequence databases: Composite sequence databases- Secondary databases-composite protein pattern databases-structure classification databases. Sequence formats, submission and retrieval

MODULE-III

Sequence analysis: Exact string matching – naïve, preprocessing and semi-numerical methods, Dynamic programming for pairwise alignment, global and local sequence alignment, Fast sequence database search, scoring matrices, scoring and significance of alignments; Fasta and Blast

MODULE-IV

Multiple sequence alignment: Introduction, sequence representation and classification, Sequence alignment algorithms, common pitfalls and mistakes; Introduction and uses of Phylogenetic analysis; methods for phylogenetic analysis; statistical significance and tree interpretation.

MODULE-V

Data mining and analytical tools for genomic and proteomic studies: Gene and genome annotation, Genome Information Resources; Protein identity based on composition, Motifs and patterns, secondary structure prediction, tertiary structure prediction; Molecular dynamics and simulations -basic concepts, force fields, protein- ligand interaction.

PRACTICAL:

1. NCBI data model: Sequence databases
2. NCBI data model: Structure and Map viewer
3. NCBI data model: Literature databases
4. Local and Global similarity alignment
5. Database search and statistics
6. Multiple sequence alignment
7. Phylogenetic tree construction and analysis
8. Gene annotation
9. Homology modelling

SUGGESTED READINGS:

1. Andreas D. and B. F. Francis Ouellette, *Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins*, 3rd Edition.
2. David W. Mount, *Bioinformatics: Sequence and Genome Analysis*, 2nd Edition.
3. Siegfried, Donna Rae, *Biology For Dummies*.
4. Phillip Compeau and Pavel Pevzner, *Bioinformatics Algorithms: An Active Learning Approach (Vol. 2)*, 2nd Edition, August 2015.
5. Michael S. Waterman, Introduction *to Computational Biology: Maps, Sequences and Genomes*.
6. Daniel. C. Liebler, *Introduction to Proteomics*, Humana press, 2002,198 pages.
7. Andrew Leach, *Molecular Modelling: Principles and Applications Paperback*
8. A. Malcolm Campbell and Laurie J. Heyer, *Discovering genomics, proteomics and bioinformatics*, Cold Spring Harbor Press 2003 Paperback
9. Coursera Bioinformatics specialization: https://www.coursera.org/specializations/bioinformatics
10. Rosalind Bioinformatics Problem solving: http://rosalind.info/problems/locations/