

Semester	7
Course	Major
Paper Code	C4MB230722T/P
Paper Title	Bioinformatics and Biostatistics
No. of Credits	4
Theory / Practical / Composite	Composite
Minimum No. of preparatory hours per week a student has to devote	4 hours/week
Number of Modules	No modules
Syllabus	BIOINFORMATICS Unit 1: Introduction to Bioinformatics and Biological Databases • Biological databases – Types, nucleic acid, genome, protein sequence and structure (PDB, Alpha-fold, PIR, Uniprot) classification (SCOP, CATH) and family database (Pfam,PROSITE) gene expression databases. • File formats - FASTA, Genbank, • Data submission & retrieval from NCBI, EMBL, DDBJ, Uniprot, PDB Unit 2: Sequence Alignments, Phylogeny and Phylogenetic trees • Local and Global Sequence alignment (EMBOSS water and needle), The Smith Waterman algorithm; pairwise and multiple sequence alignment and associated tools, substitution matrices-PAM, BLOSUM, scoring an alignment, scoring matrices, Analysis of BLAST results • Different types of Phylogenetic tree, Different approaches of phylogenetic tree construction and associated tools – UPGMA, Neighbour joining, Maximum Parsimony & Maximum likelihood (preliminary idea). Tools-DendroUPGMA, MEGA Unit 3: Genomics, Transcriptomics, Proteomics, Metabolomics, Glycomics • Different Databases and analysis tools for Genome, transcriptome, proteome, metabolome (KEGG pathway), Glycome, Reactome Unit 4: Structure prediction and analysis tools • Gene identification tool-Picking out a given gene from genomes using Genscan or other softwares (promoter region identification, repeat in genome, ORF

prediction). Gene finding tools (Glimmer, GENSCAN), Primer designing-Genetool

- Primary structure analysis (pIMw], secondary structure prediction using psipred, Jpred, threading analysis by MUSTER, homology modelling using Swiss-model, Molecular visualization using Rasmol/Jmol,
- Determination of primary, secondary and tertiary structures of RNA using sequence editor, mfold, RNA composer, tRNA scanSE,

Unit 5: Molecular docking

- Drug likeness screening, ADMET analysis, Concept of pharmacophore mapping and pharmacophore-based screening. Molecular docking (rigid and flexible)

BIOSTATISTICS

Unit 6:

- Emphasis on examples from Biological Sciences; Measures of central tendency, measures of dispersion; ideas of skewness; Applications of Correlation Coefficient and Regression.
- Elementary Probability, Discrete and Continuous Random Variable.

Unit 7:

- Mean and Variance of Discrete and Continuous Distributions namely Binomial, Geometric and Normal Distribution. Fitting of Distributions.

Unit 8:

- Principles of statistical analysis of biological data. Difference between sample and population, Sampling errors, Standard Error.
- Testing of hypothesis, Levels of Significance, Large Sample Test based on Normal Distribution.

PRACTICAL

1. Retrieving protein and DNA sequences using Entrez at NCBI
2. To retrieve metabolic pathways using KEGG, BioCyc
3. Pairwise alignment tools-FASTA, BLAST
4. Multiple sequence Alignment- Maltalin, CLUSTAL W

	5. Phylogenetic analysis-DendroUPGMA, MEGA 6. Retrieve metabolic pathway using Reactome 7. Web browsing of SwissProt, PIR, UniProtKB 8. Understanding of PDB and NDB files 9. Determination of primary, secondary and tertiary structures of RNA (different types of RNA) using RNA composer, tRNA scanSE, etc.
--	---

Learning Outcomes	• Understanding the basic concepts of bioinformatics • Study of databases and their use • Study of genomes of prokaryotes and eukaryotes and evolutionary relationship among species. • To develop computational approaches to study and predict protein, DNA, RNA structures, motifs and understanding their functions • To apply statistical methods/tools to biological, medical, and public health data • It helps the students to design robust experiments, minimize errors, analyse trial results accurately.	
Reading/Reference Lists	1. Lesk M.A.(2008) Introduction to Bioinformatics . Oxford Publication, 3rd International Student. Edition 2. Rastogi S.C., Mendiratta N. and Rastogi P. (2007) Bioinformatics: methods and applications, genomics, proteomics and drug discovery, 2nd ed. Prentice Hall India Publication 3. Biostatistics -Daniel	
Evaluation	Theory 60 (45+15) CIA-10+3+2	Practical 40 CA (38+2)
Paper Structure for Theory Semester Exam	Full Marks: 45 Short questions: 5 (each 1 mark) from 7 (5x1=5) Long questions: 4 (each 10 marks) from 6 (4x10=40)	