

UG SEMESTER 7

Semester: 7	
Course : Major	
Paper Title: Bioinformatics and Biostatistics	
Paper code: C4MB230722T/P	Credits: 6
Hours/week : 3T/4P	
Category: Core/MDC/SEC/VAC : CORE	
Theory / Practical / Composite : Composite	
No of Modules : None	
Course Overview: • 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 ac	
Course Outcome:	
(CO1): Remember	
1. Remembering different types of databases. 2. Remembering the basic rules of algorithms used in local and global alignment. 3. Remembering the statistical methods.	
(CO2): Understand	
• Understanding the basic concepts of bioinformatics. • Understanding the logics behind using different tools. • Understanding the use of databases. • Understanding the prediction tools • Understanding the use of the statistical tools.	
(CO3): Apply	
• Applying the knowledge of bioinformatics in solving biological problems. • Applying the bioinformatics tools in designing primers etc. • Applying molecular docking models to find out new potential molecules. • Applying the concept of bioinformatics tools in drawing phylogenetic tree. • Applying the statistical tools to interpret biological, medical, and public health data	
(CO4): Evaluate	
• Evaluation of biological significance of the bioinformatics tools. • Evaluation of the biological significance of BLAST analysis in identification of new species of microorganism, etc. • Evaluation of the biological results with the statistical tools.	
Prerequisites: NA	
SYLLABUS	

UNIT/Module	CONTENT	HOURS or NUMBER OF CLASSES	CO Mapping	COGNITIVE LEVEL
I.	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	4	CO1,CO2, CO3,CO4	K1,K2, K3,K4
II.	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	6	CO1,CO2, CO3,CO4	K1,K2, K3,K4
III.	Genomics, Transcriptomics, Proteomics, Metabolomics, Glycomics • Different Databases and analysis tools for Genome, transcriptome, proteome, metabolome (KEGG pathway), Glycome, Reactome	4	CO1,CO2, CO3,CO4	K1,K2, K3,K4

IV.	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,	10	CO1,CO2, CO3, CO4	K1,K2, K3, K4
V.	Molecular docking • Drug likeness screening, ADMET analysis, Concept of pharmacophore mapping and pharmacophorebased screening. Molecular docking (rigid and flexible)	4	CO1,CO2, CO3, CO4	K1,K2, K3, K4
VI.	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.	4	CO1,CO2, CO3,CO4	K1,K2, K3,K4
VII.	• Mean and Variance of Discrete and Continuous Distributions namely Binomial, Geometric and Normal Distribution. Fitting of Distributions.	4	CO1,CO2, CO3,CO4	K1,K2, K3,K4
VIII.	• Principles of statistical analysis of biological data. Difference between sample and	4	CO1,CO2, CO3,CO4	K1,K2, K3,K4

	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	20	CO1,CO2, CO3,CO4	K1,K2, K3,K4
Text Books				
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: Basic Concepts and Methodology for the Health Sciences, 10ed, ISV. Wayne W. Daniel, Chad L. Cross				
Suggested readings				
1. Powerpoint slides and teaching videos given in the class and web sources				
Evaluation Theory CIA: 30 Semester Exam:70				
Paper Structure for Theory Semester Exam: Full marks 70 Short questions: 10 (each 2 marks) from 12 (10x2=20) Long questions: 5 (each 10 marks) from 7 (5x10=50)				

Course outcomes (COs) and Cognitive Level Mapping

COs	CO Description	Cognitive levels
CO1	Remember : 1. Remembering different types of databases. 2. Remembering the basic rules of algorithms used in local and global alignment. 3. Remembering the statistical methods.	K1
CO2	Understand • Understanding the basic concepts of bioinformatics. • Understanding the logics behind using different tools. • Understanding the use of databases. • Understanding the prediction tools • Understanding the use of the statistical tools.	K2
CO3	Apply • Applying the knowledge of bioinformatics in solving biological problems. • Applying the bioinformatics tools in designing primers etc. • Applying molecular docking models to find out new potential molecules. • Applying the concept of bioinformatics tools in drawing phylogenetic tree. • Applying the statistical tools to interpret biological, medical, and public health data	K3
CO4	Evaluate • Evaluation of biological significance of the bioinformatics tools. • Evaluation of the biological significance of BLAST analysis in identification of new species of microorganism, etc. • Evaluation of the biological results with the statistical tools.	K4