

Course: Discipline Specific Elective

Semester	5
Paper Number	HMBDS5012T/P
Paper Title	Bioinformatics and Biostatistics
No. of Credits	6 (Th:4, Pr:2)
Theory/Composite	Composite
No. of periods assigned	Th: 4 Pr: 3
Course description/objective	<ol style="list-style-type: none">1. analysis of gene and protein sequences to reveal protein evolution2. to develop computational approaches to study and predict protein, DNA, RNA structures, motifs and understanding of their functions
Reading/Reference Lists	<ol style="list-style-type: none">1. Saxena Sanjay (2003) A First Course in Computers, Vikas Publishing House2. Pradeep and Sinha Preeti (2007) Foundations of Computing, 4th ed., BPB Publications3. Lesk M.A.(2008) Introduction to Bioinformatics . Oxford Publication, 3rd International Student Edition4. Rastogi S.C., Mendiratta N. and Rastogi P. (2007) Bioinformatics: methods and applications, genomics, proteomics and drug discovery, 2nd ed. Prentice Hall India Publication5. Primrose and Twyman (2003) Principles of Genome Analysis & Genomics. Blackwell
Evaluation	CIA: 20 End-Sem:80 (Th:50 and Pr:30) Question paper format of Th paper (Mod 1 & 2: 25 Marks each) For each module: Objective questions 5 marks (5 questions out of 7) 2 questions of 10 marks each (2 questions out of 3)

BIOINFORMATICS AND BIOSTATISTICS (THEORY)
SEMESTER –V

HMBDS5012T

TOTAL HOURS: 52

CREDITS: 4

Module 1

Marks: 25

Unit 1 Introduction to Bioinformatics and Biological Databases **No. of Hours: 6**

Biological databases - nucleic acid, genome, protein sequence and structure, gene expression databases, Database of metabolic pathways, Mode of data storage - File formats - FASTA, Genbank and Uniprot, Data submission & retrieval from NCBI, EMBL, DDBJ, Uniprot, PDB

Unit 2 Sequence Alignments, Phylogeny and Phylogenetic trees **No. of Hours: 8**

Local and Global Sequence alignment, pairwise and multiple sequence alignment. Scoring an alignment, scoring matrices, Analysis of BLAST results

Types of phylogenetic trees, Different approaches of phylogenetic tree construction - UPGMA, Neighbour joining, Maximum Parsimony, Maximum likelihood (preliminary idea)

Unit 3 Genome organization and analysis **No. of Hours: 6**

Diversity of Genomes: Viral, prokaryotic & eukaryotic genomes
Genome, transcriptome, proteome, metabolome (KEGG pathway), Glycomics, Major features of completed genomes: *E.coli*, *S.cerevisiae*, *Arabidopsis*, Human

Unit 4: Structure prediction **No. of Hours: 6**

Primary structure analysis, secondary structure prediction using psipred, Jpred, homology modeling using Swissmodel. Molecular visualization using Rasmol/Jmol,

Determination of primary, secondary and tertiary structures of RNA using sequence editor, mfold, RNA composer, tRNA scanSE, multalin

Module 2

Marks: 25

Biostatistics

Unit 1:

No. of Hours: 6

Measures of central tendency , Measures of dispersion ; skewness, kurtosis; Curve Fitting; Correlation and Regression.

Elementary Probability and basic laws; Discrete and Continuous Random variable, Mathematical Expectation;

Unit 2:

No. of Hours: 6

Emphasis on examples from Biological Sciences; Mean and Variance of Discrete and Continuous Distributions namely Binomial, Poisson, Geometric, Weibull, Logistic and Normal distribution. Fitting of Distributions.

Unit3:

No. of Hours: 14

Principles of statistical analysis of biological data. Sampling parameters. Difference between sample and Population, Sampling Errors, difference between parametric and non-parametric statistics; Sampling Distributions, Standard Error, Testing of Hypothesis, Level of Significance and Degree of Freedom; Large Sample Test based on Normal Distribution, Small sample test based on t-test, Z- test and F test; Confidence Interval; Distribution-free test - Chi-square test.

DSE-1: BIOINFORMATICS AND BIOSTATISTICS (PRACTICAL)

HMBDS5012P

TOTAL HOURS:39

CREDITS: 2

1. Introduction to bioinformatics databases (any three): NCBI/PDB/DDBJ, Uniprot, PDB, Prediction of different features of a functional gene.
2. BLAST analysis
3. Sequence retrieval, Sequence alignment & phylogenetic analysis using clustalW & phylip 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, Genscan/Genetool
4. Protein structure prediction: primary structure analysis, secondary structure prediction using psipred, Jpred, homology modeling using Swissmodel. Molecular visualization using Rasmol/Jmol,
5. Determination of primary, secondary and tertiary structures of RNA using sequence editor, mfold, RNA composer, tRNA scanSE, multalin
6. Problems based on measures of central tendency, Problems based on measures of dispersion.
7. Problems based on measures of skewness and kurtosis.

8. Problems based on correlation coefficient.
9. Problems based on regression.
10. Application based problems using probability distributions.
11. Fitting of distributions. (binomial, poisson, normal)
12. Problems based on small sample tests and confidence interval .(z-test, t-test,F-test)
13. Problems based on large sample tests.