

Course: MICROBIOLOGY PG

Semester	4
Paper Number	MMCB 4413
Paper Title	<b><u>BIostatISTICS &amp; BIOINFROMATICS</u></b>
No of credits	6
Non composite/composite	Composite
No. of periods assigned	6
Course description/objective	<ul style="list-style-type: none"><li>• To know the basic concepts of Biostatistics</li><li>• To have the ideas about tools and databases of Bioinformatics</li><li>• To integrate the biological experimental data with biostatistics and bioinformatics</li></ul>
Reference List	Biostatistics - Daniel Basic Bioinformatics by ManjuBansal Essential Bioinformatics by Jin Xiong
Evaluation	Theory: 70 (60 End sem + 10 CIA) Practical: 30 (10 End sem + 20 CIA) <b>Question Paper format: theory end semester</b> <b>Module 1: 30 marks</b> SHORT QUESTION: FROM 7 QTNS <b>ANSWER 5 (EACH 2 MARKS)</b> = 5X2=10 LONG QUESTION: FROM 6 QTNS <b>ANSWER 4 (EACH 5 MARKS)</b> = 4X5=20  <b>Module 2: 30 marks</b> SHORT QUESTION: FROM 7 QTNS <b>ANSWER 5 (EACH 2 MARKS)</b> = 5X2=10 LONG QUESTION: FROM 6 QTNS <b>ANSWER 4 (EACH 5 MARKS)</b> = 4X5=20  Viva: End sem 10 marks

**MMCB4413**

**BIostatISTICS & BIOINFROMATICS**

## **THEORY: 70**

### ❖ **MODULE 1: Biostatistics (25 MARKS)**

Probability and statistics- conditional probability, population, variables, collection, tabulation and graphical representation of data, frequency distribution, central tendency and skewness, Probability distribution- Binomial, Poisson & Gaussian distributions, additive and multiplicative laws of probability, concept of correlation, regression, method of least squares, random number generation, testing and use; probability density and cumulative distribution function, systematic and random sampling; accidental and systematic errors, error function, propagation of error. Testing of hypothesis- test for mean, chi-square test ANOVA . Prey predator model (numerical modelling)- An introduction. Basic idea of applied multivariate techniques.

### ❖ **MODULE 2: Bioinformatics (45 MARKS)**

Biological databases- Overview, modes of database search , mode of data storage, file formats- (Flat file format, db-tables), flatfile formats of GenBank, UniprotEMBL, DDBJ, PDB.

Different types of Biological databases - nucleic acid, genome, protein sequence and structure, gene expression databases, Database of metabolic pathways. Data submission & retrieval from NCBI, EMBL, DDBJ, Uniprot, PDB, Gene ontology (GO), promoter prediction and gene prediction analysis [SSC+AB]

Structure predictions for nucleic acids - Approaches for the prediction of RNA secondary and tertiary predictions, energy minimization and base covariance models, Determination of primary, secondary and tertiary structures of RNA using sequence editor, mfold, RNA composer, tRNAscanSE, multalin [AB]

Sequence Alignments Local and global Sequence alignment, pairwise and multiple sequence alignment, scoring an alignment, scoring matrices, dynamic programming algorithm (Smith-Waterman and Needleman-Wunch algorithm), analysis of BLAST results; sequence motif discovery. Molecular visualization using Rasmol/Jmol, [KS]

Phylogeny and Phylogenetic trees: Types of phylogenetic trees, Phylogenetic analysis- Basic concepts of phylogenetic analysis, rooted/uprooted trees, Different approaches of phylogenetic tree construction - UPGMA, Neighbour joining, Maximum Parsimony, Maximum likelihood (preliminary idea). [SSC]

Study of model organisms and its database, Genome, transcriptome, proteome, Genomics Proteomics, analysis of Microbiome, transcriptomics metabolomics integrated study, Lipidomics, Glycomics, Enzyme database, Pathway Analysis using KEGG and Reactome [SSC]

Structure prediction Of Proteins- Primary structure analysis, secondary structure prediction using psipred, Jpred, homology modeling using Swissmodel. Protein database, Protein structure and function databases, Protein classification databases (SCOP, CATH), Protein structure prediction tools (Primary, secondary, Tertiary structure) [SSC]

Comparative Modeling of Proteins - Homology, Ab Initio. Secondary Structure Prediction, Background Algorithms, Analysis and Evaluation of structures. Basic Concepts of Virtual Screening and Computer Aided Drug Discovery.

Computational Genomics : Concepts of Genome Annotation and Mining, Major features of completed genomes: E.coli, S.cerevisiae, Arabidopsis, Human , Prediction of Essential Signatures using pattern matching algorithms, Basics of Transcriptome and Metagenome Analysis

### **PRACTICAL : 30**

**Biostatistics**-Working with biological Material: Flower, blood cell, Air sample data analysis [AKM]

**Bioinformatics-**

Database retrieval system- SRS of EBI and DBGet, data retrieval from NCBI, UNIPROT, PDB, PUBCHEM, PubMed database. Analysis of gene sequence from nucleotide database (Primary Databases). Introduction of various bibliographic databases. Getting the gene sequences by exploring and querying the nucleic acid databases.

Introduction to molecular docking, Rasmol [KS]

Nucleic acid structures: RNA folding (coding and non-coding RNA), RNA loops, conformational study. DNA-binding proteins, RNA-binding proteins, Secondary and tertiary structural analysis of nucleic acids structure. [AB]

Protein structure and functional databases and Structure prediction tools, Genome browser, Protein Databases, Gene finding tools, Phylogenetic analysis, Multiple Sequence Alignment and Phylogenetic Tree Construction [SSC]

Homology Analysis using BLAST, ORF Prediction, Primer Designing, Comparative Modeling of proteins using different servers and evaluation, Generating 2D and 3D structures of RNA

**Reference:**

Biostatistics - Daniel

Basic Bioinformatics by Manju Bansal

Essential Bioinformatics by Jin Xiong