GUJARAT TECHNOLOGICAL UNIVERSITY

School of Applied Sciences and Technology
Integrated M.Sc. Biotechnology

BASICS OF BIOINFORMATICS & BIOSTATISTICS

Prerequisite:
Students should have knowledge of basic computer software, biochemistry and molecular biology

Rationale:
The objectives of this course are to provide students with theory and practical experience of use of common computational tools and databases which facilitate investigation of molecular biology and evolution-related concepts. The objective of this course is to introduce students to statistical methods and to understand underlying principles, as well as practical guidelines of "how to do it" and "how to interpret it" statistical data.

Course Scheme:

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
				Theory		Tutorial/ Practical		
L	T	P	C	University exams (ESE)	Progressive Assessment (PA)	External Practical /viva Exam(ESE)	Internal evaluation Practical /viva Exam(PA)	
3	0	2	4	70	30	30	20	150

Course Content:

Unit No.	Content	No. of Hours	Weightage (%)
1	**Introduction to bioinformatics and databases** Definition, History & milestones, objectives and applications of Bioinformatics in Biotechnology, Agriculture, Health care. Biological data, Types of Biological Data, Management of Biological data, Introduction to Databases: Introduction to Biological Databases, Types of Databases, Literature Databases: PUBMED, PMC, Europe PMC Nucleic acid and protein	15	25

	databases: GenBank, EMBL, DDBJ, Uniprot-KB, SWISSPROT Database Retrieval and Deposition Systems: SRS, Entrez, Bankit, Seqin. Biotechnological Databases: EST, SNP, And GEO, TCGA Pattern Databases: InterPro, Prosite, Pfam.		
2	**Sequence alignment** scoring matrices, PAM, BLOSUM, Local and global alignment concepts; Dot matrix sequence comparison; Dynamic programming; Needleman-Wunch algorithm, Smith- Waterman algorithm; Statistics of alignment score. Multiple sequence alignment	10	16
3	**Prediction of protein structure** Prediction of protein secondary and tertiary structure from the amino acid sequence-Use of sequence patterns for protein structure prediction, Identifying Structural features of proteins from sequence elements (motifs, domains etc.). Prediction and Identification of BCell Epitopes using protein sequences	15	25
4	**Introduction to biostatics** Types of biological data (ordinal scale, nominal scale, continuous and discrete data), frequency distribution and graphical representations (bar graph, histogram, box plot and frequency polygon), cumulative frequency distribution, populations, samples, simple random, stratified and systematic sampling. Measures of Location, Properties of Arithmetic Mean, median, mode, range, Properties of Variance and Standard Deviation, Coefficient of Variation.	10	17
5	**Statistical significance and interpretation of results** Steps in testing statistical significance, selection and computation of test of significance and interpretation of results; Sampling distribution of mean and standard error, Large sample tests (test for an assumed mean and equality of two population means with known S.D.), z-test; Small sample tests (t-test for an assumed mean and equality of means of two populations when sample observations are independent); Parametric and Non parametric tests (Mann-Whitney test); paired and unpaired t-test, chi square test.	10	16
	Total Hours:	60	

Textbook:

1. David W. Mount, Bioinformatics – Sequence and Genome analysis, Cold Spring Harbor Laboratory Press, New York Latest edition
2. Norman T.J. Bailey, Statistical Methods in Biology, Cambridge University Press, 3rd Edition

Reference Books:

1. A.D. Baxevanis and B.F.F. Ouellette (Eds) Bioinformatics: a Practical Guide to the Analysis of Genes and Proteins John Wiley and Sons Latest Edition,
2. P. N. Arora and P. K. Malhan, Biostatistics, Himalaya Publishing House, 2nd Edition

Course Outcomes:

No.	Course Outcomes	RBT Level*
1	Theoretical understanding of Statistics	UN,RM,AP,AN
2	Critical thinking and Problem Solving	UN,RM,AP,AN
3	Deciphering real-world biological problems in computational biology.	UN,RM,AP,AN
4	Understand the fundamentals of Bioinformatics, performs analysis, develops workflow/pipelines for biological data analysis.	UN,RM,AP,AN

*RM: Remember, UN: Understand, AP: Apply, AN: Analyze, EL: Evaluate, CR: Create

Suggested Course Practical List:

Not Applicable

List of Laboratory/Learning Resources Required

1. https://nptel.ac.in/courses/102106065
2. https://onlinecourses.nptel.ac.in/noc19_bt19/preview