GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Master of Science (Industrial Biotechnology)

Level: PG

Course / Subject Code: IB01001041

Course / Subject Name : Bioinformatics

1. Learning Outcomes

Learning Outcome Component	Learning Outcome (Learner will be able to)
Domain Knowledge	• Understand the fundamentals of Bioinformatics, performs analysis, develops workflow/pipelines for biological data analysis.
Critical thinking, Logic Building, Problem Solving	• Deciphering real-world biological problems in computational biology. • Assess different problem-solving approaches. • Building logical step-by-step solutions to solve biological data handling/ analyses problems.
Exposure and Cross-discipline Understanding	• Illustrate the implementation of mathematics, statistics & programming in Biological System data handling and processing.
Effective Communication	• Communicate ideas clearly and effectively.
Professional & Ethical Behaviour	• Transparency, honesty and ethical reasoning in devising tools of data handling and processing.

2. Course Duration: The course duration is 45 sessions of 60 minutes each.

3. Course Contents:

Module No:	Module Content	No. of Sessions	70 Marks (External Evaluation)
I	**Biological databases** Introduction, Primary & Secondary database, Sequence file formats, Introduction to structures, Protein Data Bank (PDB), Molecular Modelling Database (MMDb), Structure file formats, Visualizing structural information, Database of structure viewers, Collection of sequences, sequence annotation, sequence description.	8	14
2	**Sequence alignment and database searching** Evolutionary basis of sequence alignment, Optimal alignment methods, Substitution scores & gap penalties, Statistical	8	14

	significance of alignments, Database similarity searching, FASTA, BLAST, Low complexity regions, Repetitive elements, Multiple Sequence Alignment: Progressive alignment methods, Motifs and patterns, Clustral, Muscle; Scoring matrices, Distance matrices.		
3	**Phylogenetic analysis** Alignment, tree building and tree evaluation, Comparison and application of Unweighted Pair Group Method with Arithmetic Mean (UPGMA), Neighbour Joining (NJ), Maximum Parsimony (MP), Maximum Likelihood (ML) methods, Bootstrapping, Jackknife; Software for Phylogenetic analysis. DNA barcoding: Methods tools and databases for barcoding across all species, Applications and limitations of barcoding, Consortium for Barcode of Life (CBOL) recommendations, Barcode of Life Database (BOLD).	6	11
4	**Structural biology** 3-D structure visualization and simulation, Basic concepts in molecular modeling: different types of computer representations of molecules; External coordinates and Internal Coordinates, Molecular Mechanics, Force fields etc. Secondary structure elucidation using Peptide bond, phi, psi and chi torsion angles, Ramachandran map, anatomy of proteins – Hierarchical organization of protein structure –like CATH (class, architecture, topology, homology), SCOP (Structural Classification of Proteins), FSSP (families of structurally similar proteins).	6	10
5	**Classification and comparison of 3D structures** DNA & RNA secondary and tertiary structures, t-RNA tertiary structure; Protein Secondary structure prediction: Algorithms viz. Chou Fasman, GOR methods, Tertiary Structure prediction: Fundamentals of the methods for 3D structure prediction (sequence similarity/identity of target proteins of known structure, fundamental principles of protein folding etc.) Homology/comparative modeling, fold recognition, threading approaches, and ab initio structure prediction methods; CASP (Critical Assessment of protein Structure Prediction); Computational design of promoters, proteins & enzymes.	6	5
6	**Application in drug design**	3	4

	Chemical databases like NCI/PUBCHEM; Fundamentals of Receptor-ligand interactions; Structure-based drug design: Identification and Analysis of Binding sites and virtual screening; Ligand based drug design: Structure Activity Relationship – QSARs & Pharmacophore; In silico predictions of drug activity and ADMET.		
7	**Analysis of microarray data** Designing of oligo probes; Image processing and normalization; Microarray data variability (measurement and quantification); Analysis of differentially expressed genes; Experimental designs.	2	5
8	**Biological algorithms** Comparison with computer algorithms, string structures, Introduction to programming in computational biology through Python.	3	2
9	**System biology** System-level understanding of biological systems, use and integration of data from transcriptomics, proteomics and metabolomics; Concepts in glycomics, interactomics and fluxomics.	3	5
10	**Practicals** 1. Introduction to major online resources and interpreting search results NCBI, GenBank/ DDBJ/ EMBL, EMBOSS, Ensembl, PubMed, NR, UniProt, InterPro, OMIM, KEGG, Expassy, SWISS-PROT, PROSITE, PFam, PIR, InerPro, PSSM, TGCA. PDBPARAM, SCOPE, CATH, BioGRID. 2. DeepViewer, MODELLER, Auto Dock, Pymol. 3. Genomics and proteomics analysis using online tool Galaxy. 4. Similarity searches using tools like BLAST and interpretation of results. 5. Multiple sequence alignment using ClustalW. 6. Phylogenetic analysis of protein and nucleotide sequences. 7. Using RNA structure prediction tools. 8. Use of various primer designing and restriction site prediction tools.	---	(30 marks CEC)

	9. Use of different protein structure prediction databases (PDB, SCOP, CATH). 10. Construction and study of protein structures using Deepview/PyMol. 11. Homology modelling of proteins. 12. MATLAB working environment: Installing and running simple MATLAB programs, MATLAB key features, MATLAB (Matrix Laboratory). 13. Use of Statistical packages like SPSS (Statistical Package for Social Sciences)/ SAS (Statistical Analysis System). 14. Performing various statistical analysis like T-test, ANOVA, Regression, Chi-square, PLS (Partial Least Squares) and PCA (Principle Component Analysis).		

4. Pedagogy:

- ICT enabled Classroom teaching
- Case study
- Practical/live assignment
- Interactive classroom discussions

5. Evaluation:

Students shall be evaluated on the following components:

	Internal Evaluation	**(Internal Assessment – 20 Marks)**
A	• Assignments	10 marks
	• Class Presence	5 marks
	• Record maintenance	5 marks
B	**Mid-Semester Examination**	**(Internal assessment-30 Marks)**
C	**End-Semester Examination**	**(External assessment-70 Marks)**

LO – PO Mapping: Correlation Levels:

1 = Slight (Low); 2 = Moderate (Medium); 3 = Substantial (High), "-"= no correlation

Sub Code: 1310104	**PO1**	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**
LO1:Programming environment and Domain Knowledge	3	2	3	3	2	2	2
LO2: Critical thinking, Logic Building, Problem Solving	3	3	3	3	3	3	3

LO3: Exposure and Cross-discipline Understanding	3	2	3	3	2	3	3
LO4:Effective Communication	3	3	3	3	3	3	2
LO5: Professional & Ethical Behaviour	3	3	3	3	3	3	3

6. Reference Books:

No	Author	Name of the Book	Publisher	Year of Publication / Edition
1	David W. Mount	Bioinformatics – Sequence and Genome analysis	Cold Spring Harbor Laboratory Press, New York	Latest edition
2	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 Latest Edition

Note: Wherever the standard books are not available for the topic appropriate print and online resources, journals and books published by different authors may be prescribed.

7. List of Journals/Periodicals/Magazines/Newspapers / Web resources, etc

- https://nptel.ac.in/courses/102106065- NOC- Bioinformatics: Algorithm and Applications
- (https://galaxy-tut.edu.au/)
- (https://wiki.galaxyproject.org/Teach/GTN)

Course Outcomes:

On completion of this course, students should be able to:

• Develop an understanding of basic theory of computational tools.

• Gain working knowledge of computational tools and methods.

• Appreciate relevance of computational tools for investigating specific contemporary biological questions.