GUJARAT TECHNOLOGICAL UNIVERSITY
Integrated Master of Science (Biotechnology)
Semester: 7
Subject Name: Bioinformatics
Subject Code: 1370404

Prerequisite: Candidates enrolling in the integrated MSc in Industrial Biotechnology with an elective in Bioinformatics are expected to possess a solid foundation in core biological sciences such as molecular biology, genetics, biochemistry, and microbiology. Additionally, a working knowledge of programming languages (e.g., Python or R), as well as fundamental skills in mathematics and statistics, is essential to support the computational analysis of biological data. This background will enable students to effectively engage with topics like sequence analysis, transcriptomics, and systems biology, and to integrate bioinformatics tools into industrial biotechnology applications.

Rationale: Integrating an elective in Bioinformatics into the MSc Industrial Biotechnology curriculum is essential for equipping students with the computational skills necessary to harness and interpret the vast amounts of biological data generated in modern industrial settings. This subject bridges the gap between traditional wet-lab biotechnology and digital data analysis, enabling future biotechnologists to optimize processes such as metabolic engineering, strain development, and product yield enhancement through data-driven insights.

Course Scheme:

Teaching Scheme			Total Credits	Assessment Pattern and Marks				Total Marks
L	T	PR	C	Theory		Practical		
				ESE (E)	PA(M)	ESE (V)	PA (I)	
4	0	0	4	70	30	0	0	100

Course Content:

Sr. No.	Course Content	No. of Hours	% of Weightage
1	**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 significance of alignments, Database similarity searching, FASTA, BLAST, Low complexity regions, Repetitive elements, Multiple Sequence Alignment: Progressive alignment methods, Motifs and patterns,	**8**	**14**

	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** 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.	**3**	**4**

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. 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). Performing various statistical analysis like T-test, ANOVA, Regression,Chi-square, PLS (Partial Least Squares) and PCA (Principle Component Analysis).	---	(30 marks CEC)

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

Course Outcome:

After Completion of the Course, Student will able to:

No	Course Outcomes	RBT Level
1	Understand the fundamentals of Bioinformatics, performs analysis, develops workflow/pipelines for biological data analysis.	UN
2	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.	AN
3	Illustrate the implementation of mathematics, statistics & programming in Biological System data handling and processing.	EL
4	Communicate ideas clearly and effectively.	EL
5	Transparency, honesty and ethical reasoning in devising tools of data handling and processing.	CR

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

List of Experiments: (Minimum 6 experiments need to be performed)

****