GUJARAT TECHNOLOGICAL UNIVERSITY

Program Name: Bachelor of Engineering
Level: UG
Branch: Biotechnology
Subject Code: BE04004031
Subject Name: Bioinformatics

w. e. f. Academic Year:	**2024-2025**
Semester:	**4**
Category of the Course:	**Professional Core Course**

Prerequisite:	Fundamentals of Biochemistry, Introduction to Computer Science and Molecular Biology Basics
Rationale:	Bioinformatics combines biology and computing to manage, analyze, and interpret large biological datasets, such as genomes and protein structures. It enables discovery of genes, understanding diseases, and drug development through computational tools. By organizing complex data, it turns raw sequences into meaningful biological insights. Its core purpose is accelerating research in genetics, evolution, and personalized medicine.

Course Outcomes:

Sr. No.	CO statement	Marks% weightage
CO-1	Recall significant advances in historical timeline of bioinformatic developments	15
CO-2	Identify, locate and retrieve specific records from various biological databases	15
CO-3	Predict concealed signatures in RNA and Protein sequences	25
CO-4	Construct virtual 3D models of protein structures	25
CO-5	Analyze healthcare data with bioinformatic tools	20

Teaching and Examination Scheme:

Teaching / Learning Scheme (in Hours per semester)					Total Credits	Assessment Pattern and Marks					Total Marks
						Theory		Tutorial / Practical			
L	T	P	PBL*	Total no of hours per semester		ESE (E)	PA / CA (M)	PA/ CA (I)	PBL (I)	ESE (V)	
45	0	30	45	120	4	70	30	20	30	50	200

- **Problem Based Learning (PBL) aims to accommodate learning beyond syllabus as per clause 9.4 of NBA manual.**

Where L = Lecture, T= Tutorial, P= Practical, TW/SL = Term-Work / Self-Learning, TH = Total Hours, PA = Progressive Assessment, ESE = End-Semester Examination

Content:

Sr. No.	Content	Total Hrs
1	**Introduction to Bioinformatics** A brief history and evolution of bioinformatics as a discipline. Overview of key bioinformatics resources: NCBI (GenBank, BLAST, Entrez) and UniProt (Swiss-Prot, TrEMBL). Introduction to biological databases: primary, secondary, and specialized databases. Sequence analysis fundamentals: sequence file formats (FASTA, GenBank), scoring matrices (PAM, BLOSUM), pairwise sequence alignment (global and local alignments), dynamic programming algorithms (Needleman Wunsch, Smith-Waterman). Multiple sequence alignment techniques (Clustal Omega, MUSCLE), molecular evolution principles, phylogenetic tree construction methods (Neighbor-Joining, Maximum Parsimony, Maximum Likelihood). Gene prediction: ab initio methods and evidence-based methods. Functional annotation strategies using tools like InterProScan, Pfam, and GO annotation.	**15**
2	**Structure Databases and Structural Bioinformatics** Introduction to structure databases: Protein Data Bank (PDB), SCOP, CATH. Structural analysis tools: DSSP for secondary structure assignment, prediction of secondary structural elements (SSEs) using methods like PSIPRED. 3D structure prediction: template-based modeling (homology modeling using SWISS-MODEL, MODELLER), and template-free (ab initio) methods. Introduction to deep learning approaches for structure prediction (e.g., AlphaFold2 and RoseTTAFold). Visualization tools: PyMOL, Chimera, and validation of structural models (Ramachandran plots, Verify3D, ProSA-web).	**10**
3	**Genomics and Applications:** Basic concepts: genes, genomes, chromosomes. Human genome project and implications for biology and medicine. Advances in DNA sequencing technologies: Sanger sequencing, Next-Generation Sequencing (NGS), Third-Generation Sequencing (PacBio, Oxford Nanopore). Introduction to evolutionary genomics: gene duplication, horizontal gene transfer, genome rearrangements. Introduction to metagenomics: analyzing microbial communities without culturing, applications in environment, industry, and health. Future trends in genomics: personalized medicine, synthetic genomics, and genome editing technologies (CRISPR-Cas9).	**10**

4	**Modern Bioinformatics:** AI and Machine Learning Integration: Applications of machine learning in sequence annotation, structure prediction, functional genomics, and biomarker discovery. Introduction to neural networks, support vector machines (SVM), random forests, and deep learning models in bioinformatics. Systems Biology and Network Bioinformatics, Structural Bioinformatics with AI, Cloud Computing and Big Data in Bioinformatics, Ethics, Data Privacy, and Open Science:	**10**
	TOTAL	**45**

Suggested Specification table with Marks (Theory): (For B.E. only)

Distribution of Theory Marks					
R Level	U Level	A Level	N Level	E Level	C Level
20	**20**	**20**	**20**	**10**	**10**

R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Reference Books:

1. Andreas D. Baxevanis, Gary D. Bader, David S. Wishart, Bioinformatics, Wiley, 2019, 4th Edition
2. Arthur M. Lesk, Introduction to Bioinformatics, Oxford University Press, 2019, 5th Edition
3. John M. Archibald, Genomics: A Very Short Introduction, OUP Oxford, 2018, Illustrated Edition
4. Bioinformatics Lab Manual

List of Experiments / Practical Demonstrations

1. Biological Databases: NCBI- GENBANK, PUBCHEM, UNIPROT, PDB, CATH, SCOP & KEGG.
2. Sequence Alignment Tools: Pairwise- BLAST, MSA- CLUSTAL OMEGA.
3. Phylogenetic Analysis: Tree Construction and Analysis.
4. Sketching and Visualization of small molecules - Biovia Discovery Studio.
5. Protein 3D structure visualization – Biovia Discovery Studio.
6. Protein Modelling – Biovia Discovery Studio.
7. Next Generation RNA Sequencing – Data visualization and analysis.
8. Metagenomics - Data visualization and analysis.

Major Equipment

1. Computers with internet access
2. High-performance workstation (optional for molecular modeling)

Open-Source Software / Learning Websites

1. NCBI, UniProt, PDB
2. BLAST, Clustal Omega, SWISS-MODEL
3. Chimera, PyMOL
4. NPTEL Bioinformatics

List of suggested activities for Problem Based Learning:

1. Case study on protein modelling of a disease target.
2. Comparative genomics project (two organisms).
3. Mini project using NGS dataset (public data).

****