

Course code	Course Name	L-T-P Credits	Year of Introduction
CS465	BIOINFORMATICS	3-0-0-3	2016

Course Objectives:

- To introduce concepts and data representations in bioinformatics
- To introduce fundamentals of Sequence alignment and Gene Recognition
- To discuss predictive methods using DNA and Protein Sequences

Syllabus:

Introduction to bioinformatics and molecular biology: Databases tools and their uses, Data searches and Pairwise Alignments, Multiple Sequence Alignments, Molecular Phylogenetic, Genomics and Gene Recognition, Protein and RNA structure Prediction

Expected Outcome:

The Students will be able to :

- interpret the concepts of bioinformatics
- identify different types of biological sequence
- analyse multiple sequences and find conserved regions
- predict RNA and Protein secondary structures
- analyse genomic sequences and identify encoded gene regions

References:

1. S C Rastogi, N Mendiratta and P Rastogi, " Bioinformatics: Methods and Applications" , ISBN : 978-81-203-4785-4, published by PHI Learning Private Limited, New Delhi, 2015.
2. D E Krane and M L Raymer, Fundamental Concepts of Bioinformatics, ISBN 978-81-7758-757-9, Pearson Education, 2006.
3. Andreas D.Baxevanis, B F Francis Ouellette, "Bioinformatics - A Practical Guide to the Analysis of Genes and Proteins", Third Edition, 2005-2006, ISBN: 978-81-265-2192-0, published by John Wiley & Sons INC. , U.K.
4. Neil C Jones and Pavel A Pevzner, An Introduction to Bioinformatics Algorithms, MIT press, 2004.

Course Plan			
Module	Contents	Hours	End Sem. Exam Marks
I	Bioinformatics and Computational Biology, Nature & Scope of Bioinformatics. The central dogma of molecular biology and bio-sequences associated with it, RNA classification –coding and non coding RNA- mRNA, tRNA, miRNA and sRNA, RNAi. DNA and RNA structure – Nucleic Acid structure and function, Genetic Code, Genes and Evolution	6	15%
II	Importance of databases - Biological databases-primary sequence databases, Composite sequence databases- Secondary databases- nucleic acid sequence databases - Protein sequence data bases - structure databases, Types of databases, Data retrieval tools - Entrez	8	15%

FIRST INTERNAL EXAM			
III	Sequence alignment – local/global, pairwise sequence alignment, scoring methods. Needleman and Wunsch algorithm, global and local alignments. Multiple sequence alignment. Scoring matrices: basic concept of a scoring matrix, Matrices for nucleic acid and proteins sequences, PAM and BLOSUM series, principles based on which these matrices are derived. Differences between distance & similarity matrix.	8	20%
IV	Introduction, Advantages, Phylogenetic Trees, Tree topologies, Methods for phylogenetic analysis- Distance Matrix methods, Character based methods. HMM (Hidden Markov Model): Introduction to HMM, Forward algorithm, Viterbi algorithm, applications in Bioinformatics	6	15%
SECOND INTERNAL EXAM			
V	General introduction to Gene expression in prokaryotes and eukaryotes- Prokaryotic Genomes – Gene structure, GC content, Gene Density, Eukaryotic Genomes- Gene structure, GC content, Gene Density, Gene Expression, Transposition, Gene prediction approaches.	8	20%
VI	Protein and RNA structure Prediction: Predicting RNA secondary structure - Nussinov Algorithm, Energy minimisation methods - Zuker Algorithm. Amino Acids, Polypeptide Composition, Protein Structures, Algorithm for protein folding, Structure prediction	6	15%
END SEMESTER EXAM			

Question Paper Pattern (End semester exam)

- There will be **FOUR** parts in the question paper – A, B, C, D
- Part A**
 - Total marks : 40**
 - TEN** questions, each have **4 marks**, covering **all the SIX modules** (**THREE** questions from **modules I & II**; **THREE** questions from **modules III & IV**; **FOUR** questions from **modules V & VI**). **All the TEN** questions have to be answered.
- Part B**
 - Total marks : 18**
 - THREE** questions, each having **9 marks**. One question is from **module I**; one question is from **module II**; one question **uniformly** covers **modules I & II**.
 - Any TWO** questions have to be answered.
 - Each question can have **maximum THREE** subparts.
- Part C**
 - Total marks : 18**

- b. **THREE** questions, each having **9 marks**. One question is from **module III**; one question is from **module IV**; one question *uniformly* covers **modules III & IV**.
- c. **Any TWO** questions have to be answered.
- d. Each question can have *maximum THREE* subparts.

5. Part D

- a. **Total marks : 24**
 - b. **THREE** questions, each having **12 marks**. One question is from **module V**; one question is from **module VI**; one question *uniformly* covers **modules V & VI**.
 - c. **Any TWO** questions have to be answered.
 - d. Each question can have *maximum THREE* subparts.
6. There will be **AT LEAST 60%** analytical/numerical questions in all possible combinations of question choices.

