GENERIC ELECTIVE (GE-6)

Credit distribution, Eligibility and Pre-requisites of the Course

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course
		Lecture	Tutorial	Practical/ Practice		
Informatics and Statistics for Biology and Allied Sciences ALS BOT GE 06	**4**	**2**	**0**	**2**	Class XII pass with Biology and Chemistry	NA

Learning Objectives

The Learning Outcomes of this course are as follows:

- to build an understanding *insilico*/computational approaches in various aspects of understanding biology and biological research.
- to build analytical skills and integrate the principles of statistical analyses for robust interpretation of biological observations.

Learning outcomes

By studying this course, students will be able to:

- learn the basics of bioinformatics and develop awareness of the interdisciplinary nature of this field.
- learn about biological databases, sequence retrieval, alignment, and phylogenetic analysis using various tools.
- understand the basic concept of sampling methods, data classification, presentation and statistical analysis.

Unit 1: Introduction to Bioinformatics (4 Hours)

Historical background, Aims and scope, bioinformatics in Genomics, Transcriptomics, Proteomics, Metabolomics, Systems biology and drug discovery, Applications and Limitations in bioinformatics.

Unit 2: Biological databases (4 Hours)

Introduction to biological databases - Primary, secondary and composite databases. Study of following databases: NCBI (GenBank, PubChem, PubMed and its tools (BLAST)), introduction to EMBL, DDBJ, UniProt, PDB and KEGG.

Unit 3: Basic concepts of Sequence alignment (4 Hours)

Similarity, identity and homology. Concepts of alignment (gaps and penalty); Alignment – pairwise and multiple sequence alignments.

Unit 4: Molecular Phylogeny (4 Hours)

Introduction to Molecular Phylogeny, methods of construction of phylogenetic trees: maximum parsimony (MP), maximum likelihood (ML) and distance (Neighbour-joining) methods.

Unit 5: Biostatistics (2 Hours)

Biostatistics – definition, Basics of descriptive and inferential statistics; Limitations and applications of biostatistics.

Unit 6: Data types and presentation (3 Hours)

Primary and secondary data; Sampling methods (in brief); tabulation and presentation of data.

Unit 7: Descriptive Statistics (4 Hours)

Measures of central tendency - mean, median, and mode; Measures of dispersion - range, standard deviation, and standard error.

Unit 8: Correlation and Regression (3 Hours)

Types and methods of correlation, Introduction to simple regression equation; similarities and dissimilarities between correlation and regression.

Unit 9: Statistical inference (3 Hours)

Hypothesis – (simple hypothesis), student's t test, chi-square test.

(Note: Numerical based questions of Unit 7, 8 and 9 should be covered only in practical)

PRACTICAL (60 hours)

1. Biological databases (NCBI, EMBL, UniProt, PDB)
2. Literature retrieval from PubMed
3. Sequence retrieval (protein and gene) from NCBI (formats - FASTA, GenBank and GenPept formats)
4. Protein Structure retrieval from PDB (in pdb format) and visualization by viewing tools (Ras Mol/ J mol/Mol*/Swiss 3D Viewer/Pymol)
5. Multiple sequence alignment (MEGA/Clustal omega)
6. Construction of phylogenetic tree (PHYLIP/ MEGA/ Clustal omega).
7. Making of Bar diagrams, Pie chart, Histogram, Frequency polygon, Cumulative frequency curve (any four) in the given data set using Microsoft Excel
8. Calculation of mean, mode, median, standard deviation and standard error (through manual calculation and using Microsoft Excel) (use only ungrouped data)
9. Calculation of correlation coefficient values by Karl Pearson's /Spearman Rank methods (through manual calculation and using Microsoft Excel)
10. Student's t-test (using Microsoft Excel only), chi square test (Manual and using Microsoft Excel).

Essential/recommended readings

1. Ghosh, Z., & Mallick, B. (2008). *Bioinformatics – Principles and Applications* (1st ed.). Oxford University Press.
2. Baxevanis, A.D., Ouellette, B.F., John (2005). *Bioinformatics: A Practical Guide to theAnalysis of Genes and Proteins* (3rd ed.). Wiley & Sons, Inc.
3. Roy, D. (2009). *B* (1st ed.). Narosa Publishing House.
4. Andreas, D. Baxevanis& B.F. Francis Ouellette. (2004). *Bioinformatics: A practicalguide to the analysis of genes and proteins* (3rd ed.). John Wiley and Sons.
5. Khan, I.A., & Khanum, A. (2004). Fundamentals of Biostatistics (5th ed.). Ukaazpublications.
6. Campbell, R.C. (1998). *Statistics for Biologists*. Cambridge University Press.

Suggestive readings

1. Pevsner, J. (2009). *Bioinformatics and Functional Genomics* (2nd ed.). Wiley Blackwell.
2. Xiong, J. (2006). *Essential Bioinformatics* (1st ed.). Cambridge University Press.
3. Mount, D.W. (2004). *Bioinformatics: Sequence and Genome analysis* (2nd ed.). ColdSpring Harbor Laboratory Press, USA.
4. Zar, J.H. (2012). *Biostatistical Analysis* (4th ed.). Pearson Publication.
5. Pandey, M. (2015). *Biostatistics Basic and Advanced*. M V Learning.

Note: Examination scheme and mode shall be as prescribed by the Examination Branch, University of Delhi, from time to time.