GENERIC ELECTIVES (BOT-GE-10)

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 **BOT-GE-10**	**4**	**2**	**0**	**2**	**Nil**	**Nil**

Learning Objectives:

- To build an understanding *in silico*/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

The student will understand:

- 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.

SYLLABUS OF BOT-GE-10

Unit 1: Introduction to Bioinformatics **Weeks: 1.5**

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 **Weeks: 02**

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 **Weeks: 02**

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

Unit 4: Molecular Phylogeny **Weeks: 02**

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

Unit 5: Biostatistics **Week: 01**

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

Unit 6: Data types and presentation **Weeks: 1.5**

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

Unit 7: Descriptive Statistics **Weeks: 02**

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

Unit 8: Correlation and Regression **Weeks: 1.5**

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

Unit 9: Statistical inference **Weeks: 1.5**

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)

Practicals:

1. Biological databases (NCBI, EMBL, UniProt, PDB) **(Week: 02)**
2. Literature retrieval from PubMed. **(Week: 01)**
3. Sequence retrieval (protein and gene) from NCBI (formats - FASTA, GenBank and GenPept formats). **(Week: 02)**
4. Protein Structure retrieval from PDB (in pdb format) and visualisation by viewing tools (Ras Mol/ J mol/Mol*/Swiss 3D Viewer/Pymol). **(Week: 02)**
5. Multiple sequence alignment (MEGA/ Clustal omega). **(Week: 02)**
6. Construction of phylogenetic tree (PHYLIP/ MEGA/ Clustal omega). **(Week: 02)**
7. Making of Bar diagrams, Pie chart, Histogram, Frequency polygon, Cumulative frequency curve (any four) in the given data set using Microsoft Excel. **(Week: 01)**
8. Calculation of mean, mode, median, standard deviation and standard error (through

manual calculation and using Microsoft Excel) (use only ungrouped data). **(Week: 01)**

9. Calculation of correlation coefficient values by Karl Pearson's /Spearman Rank methods (through manual calculation and using Microsoft Excel). **(Week: 01)**
10. Student's t-test (using Microsoft Excel only), chi square test (Manual and using Microsoft Excel). **(Week: 01)**

Suggested readings:

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

Additional Resources:

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

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