DISCIPLINE SPECIFIC ELECTIVE COURSE (DSE -2): Biostatistics & Bioinformatics for Plant Sciences

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 (if any)
		Lecture	Tutorial	Practical / Practice		
Biostatistics & Bioinformatics for Plant Sciences **DSE-2**	**4**	**2**	**0**	**2**	Class XII pass	**Nil**

Learning Objective:

- To train students in using computational and mathematical tools to solve biological problems.

Learning Outcomes:

At the end of this course students will be able to:

- use the various online databases and resources for accessing biological data.
- use the different methods of alignment of DNA, RNA and protein sequences and interpret the significance of the same.
- understand the descriptive and inferential statistical tests for interpretation of experimental data.

Unit 1- Introduction to Bioinformatics **3 hours**

Historical background; Aims and scope; Bioinformatics in Genomics, Transcriptomics, Proteomics, Metabolomics; Applications of bioinformatics in crop improvement

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 (only BLAST)), introduction to UniProt, PDB, PlantPepDB.

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

Unit 5- Introduction to Biostatistics **2 hours**

Definition, Basics of descriptive and inferential statistics; Limitations and applications.

Unit 6- Data and sampling methods **3 hours**
Primary and secondary data; Sampling methods (in brief); tabulation and presentation of data.

Unit 7- Measures and deviations of central tendencies **4 hours**
Dispersion - range, standard deviation, mean deviation, standard error, skewness and kurtosis, quartile deviation –merits and demerits; Coefficient of variation.

Unit 8-Correlation and Regression **3 hours**
Correlation - types and methods of correlation (I. E. Karl Pearson and Spearman Rank method), Introduction to simple regression equation; similarities and dissimilarities between correlation and regression.

Unit 9- Statistical tests **3 hours**
Statistical inference - 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 **60 hours**

1. Biological databases (NCBI, UniProt, PlantPepDB)
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. Calculation of standard deviation and coefficient of variation through manual calculation and using Microsoft Excel, using only ungrouped data)
8. Calculation of correlation coefficient values by Karl Pearson's /Spearman Rank methods (through manual calculation and using Microsoft Excel)
9. Student's t-test (using Microsoft Excel), chi square test (Manual and using Microsoft Excel)

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. Zar, J.H. (2012). Biostatistical Analysis, 4th edition. London, London: Pearson Publication.
5. 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. Pandey, M. (2015). Biostatistics Basic and Advanced. New Delhi, Delhi: M V Learning.
5. Khan, I.A., Khanum, A., Khan S., (2020). Fundamentals of Biostatistics, 6th edition. Ukaaz Publications, Hyderabad, India.

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