**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 | | |
| Bioinformatics and Biostatistics (BS-DSC-701) | 4 | 2 | | 2 | Class XII with Biology | NA |

**Learning Objectives**
The Learning Objectives of this course are as follows:

- The objective of this course is to impart basic understanding of bioinformatics and biostatistics. The course will introduce the broad scope of bioinformatics by discussions on the theory and practices of computational methods in biology.
- This course also aims to provide students with a practical hands-on experience with common bioinformatics tools and databases.
- This course also provides foundational skills and the knowledge of analysis of scientific data and students will gain a deeper understanding of its relevance and applications in various fields of biological sciences.

   Learning outcomes
- Students will understand the basics of bioinformatics and computational biology and develop awareness of the interdisciplinary nature of this field.
- Students will learn about Biological Databases and the types of databases.
- Students will understand protein structure using visualization softwares.
- Students will be able to gain understanding of sequence alignments and analyze phylogeny using alignment tools.
- Students will understand different applications of genomics in gene prediction and obtain knowledge on applications of bioinformatics from genomes to personalized medicine.
- Students will understand the basic concepts of sampling methods and data classification and presentation; variables and statistical methods.
- Students will acquire the understanding of interpreting the scientific data that is generated during scientific experiments.

# SYLLABUS FOR DSC-19

**2.2 Course Contents**

**Theory**

**Credits: 2**                                               **Total Hours: 30**

**Unit 1: Introduction to Bioinformatics & Biological Databases**          **7 hours**
Historical background; Aims and scope; Bioinformatics in Genomics, Transcriptomics, Proteomics, Metabolomics, Systems biology; Applications of bioinformatics in Plant Sciences and Computational Drug Discovery; Introduction to biological databases - Primary, secondary and composite databases; Study of following databases: NCBI (GenBank, PubChem, PubMed and BLAST); Introduction to UniProt, PDB, PlantPepDB.

**Unit 2: Sequence Alignment & Molecular Phylogeny**          **8 hours**

Similarity, identity and homology; Concepts of alignment (gaps and penalty); Alignment – pairwise and multiple sequence alignments, Scoring Matrices: PAM and BLOSUM. Introduction to Molecular Phylogeny, methods of construction of phylogenetic trees: maximum parsimony (MP), maximum likelihood (ML) and distance (Neighbour-joining) methods (to be discussed briefly).

**Unit 3: Introduction to Biostatistics & sampling methods**          **5 hours**
Biostatistics - definition and basic principles. Variables - measurements, functions, limitations and uses of statistics; Primary and secondary data; Sampling methods (in brief); tabulation and presentation of data.

**Unit 4: Measures of central tendency, Correlation, Regression & Statistical inference**
                                                              **10 hours**
Measures of central tendency - mean, median, mode, merits & demerits. Measures of dispersion - range, standard deviation, mean deviation, standard error, quartile deviation –merits and demerits; Coefficient of variance. Correlation - types and methods of correlation (I. E. Karl Pearson and Spearman Rank method), Introduction to simple regression equation, Hypothesis – (simple hypothesis), student's  t-test, chi-square test.

### 2.3    Practical

**Credits: 2**                                               **Total Hours: 60**

1. Sequence Retrieval from NCBI  & GenBank
2. Protein Structure download from PDB and Structure visualization ( Jmol/Pymol)
3. Pairwise sequence alignment and its interpretations using BLAST.
4. Multiple Sequence alignment using CLUSTALw and Phylogenetic tree construction.
5. Gene Prediction using Gene Prediction Tools
6. Protein structure prediction using homology modeling and structure validation using Ramachandran Plot .
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 deviations, quartile deviations, standard error and coefficient of variance.

**9.** Student's t-test (using Microsoft Excel only), chi square test (Manual and using Microsoft Excel).

**2.4 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 ofGenes 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 theanalysis 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

2.5 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.