D. Phil. Coursework Syllabus

(Bioinformatics Program)

CENTRE OF BIO-INFORMATICS

University of Allahabad

Allahabad

D. Phil Coursework Papers:

1. **Research Methodology and Communication Skills [Credit 4]**

2. **Mathematical & Statistical Methods in Bioinformatics [Credit 4]**

3. **Computational Biology & Structural Bioinformatics [Credit 8]**

4. **Elective Paper [Credit 8]**

   - **Elective paper should be decided by the D.Phil. supervisor.**

RESEARCH METHODOLOGY AND COMMUNICATION SKILLS

Course Code: C-1   **Credit Units: 4**

Course Objective:
This course enables students to develop their understanding of research methods, and confidence in designing a research project, choosing and executing appropriate methods, and assessing its intellectual/academic rigor.

Course Contents:

Unit I
Research – Definition – Importance and Meaning of research – Characteristics of research – Types of Research – Steps in research – Identification, Selection and formulation of research problem – Research questions – Research design – Formulation of Hypothesis – Review of Literature

Unit II
Sampling techniques: Sampling theory – types of sampling – Steps in sampling – Sampling and Non-sampling error – Sample size – Advantages and limitations of sampling. Collection of Data: Primary Data – Meaning – Data Collection methods – Secondary data – Meaning - Relevance's, Limitations and cautions, Statistics in Research.

Unit III
Type of research articles. Scientific paper format. Writing, evaluating, presenting, and publishing the results of scientific research. Choosing the appropriate journal, Peer review system, Impact Factor, h-index, Citation, and their importance, ISSN & ISBN numbers, Indexed & referred journals

Unit IV
Case studies of areas of current research. Formulating a research plan and writing a research proposal, seminar on current topics, Ethical issues & Copyright issues in research.

MATHEMATICAL & STATISTICAL METHODS IN BIOINFORMATICS

Course Code: C-2   **Credit Units: 4**

Course Objective:
The main objective of this course is to enable students to make meaningful interpretations of the biological data using statistical methods.

Course Contents:

Unit I: MATHEMATICAL METHODS IN BIOLOGICAL SEQUENCE ANALYSIS
Markov chains and hidden Markov models, Fourier analysis, Shannon information theory, EM algorithms, Genetic algorithms, Support Vector Machine.

Unit II: ELEMENTARY PROBABILITY AND DISTRIBUTION THEORY
Random experiments, sample space, events, definition of probability and elementary results, conditional probability and Bayes theorem, random variables, probability mass function and probability density function, Binomial distribution, Multinomial distribution, Poisson distribution, Normal/ Gaussian distribution.

Unit III: INFERENCE- TESTS OF HYPOTHESES
Basic elements of testing of hypothesis, t-test, Chi-square test, Fishes exact test, F-test, Significance of a test, P- value Testing, Z score, Non-parametric Wilcoxon test

Unit IV: REGRESSION AND CORRELATION METHODS
Pearson Correlation, Spearman's Rank Correlation, Linear regression model, Least squares methods, Estimating model parameters, Residual sum of squares, Analysis of variance for regression model.

COMPUTATIONAL BIOLOGY & STRUCTURAL BIOINFORMATICS

Course Code: C-3   **Credit Units: 8**

Objective: This course will enable the students to achieve skills in Bioinformatics that are essential for application in Computational Biology & Biotechnology.

Course Contents:

Unit I: Introduction to Bioinformatics
Overview of Bioinformatics and computational biology. Importance of databases – EMBL, NCBI databases - protein sequence databases (SwissProt) - structure databases (PDB), bibliographic databases, analysis of biological data - methods of alignment - methods for optimal alignments; using gap penalties and scoring matrices- multiple sequence alignment, tools for MSA (MUSCLE, T-coffee), Similarity Searching Tools: BLAST and FASTA, Theory and Algorithms, variants of BLAST and FASTA, PSI-BLAST and PHI BLAST, Statistical Significance.

Unit II: Molecular Evolution & Phylogeny
The concept of evolutionary tree, Models of sequence evolution (HKY, TN93, GTR), Phylogenetic tree reconstruction methods; Neighbor-Joining, Maximum Parsimony, Maximum Likelihood, and Bayesian Inference. Molecular Clock, Rate heterogeneity, Phylogeneomics: Super-tree & Super-matrix method, Tree evaluation methods: Bootstrapping and jack-knifing methods.

Unit III: Bimolecular Interactions & Computational Systems Biology
Introduction to biological networks, Importance of biological networks, Types of biological networks (PPI networks; Gene-regulatory networks; Transcriptional regulatory network; Metabolic networks; Signalling networks), computational prediction of PPIs, Network properties, Network motifs, Network layouts, Modelling and Simulation of Biological Systems. Database of biological networks (STRING, DIP, BioGRID, iHOP, and KEEG) and Network Analysis software (Cytoscape and Network Analyzer). GO enrichment and Pathway analysis.

Unit IV: Predictive methods using DNA and protein sequences
Concepts of motif, pattern and profile, Gene Identification methods, Gene predictions strategies – Prokarytotic and Eukaryotic, Identification and characterization of proteins, Protein structure prediction methods: Secondary and tertiary approaches.

Unit V: Introduction to Chemoinformatics, Molecular modelling & Drug designing
Chemical similarity search methods, Designing focused chemical libraries, Comparative protein modelling, Molecular docking and virtual high-throughput screening, Search algorithms, Scoring methods, Challenges in Molecular docking, Introduction of receptor flexibility through multiple receptor conformations (MRC docking), Hit-to-lead optimization, Introductory concept of Activity Cliffs.