69 BIOINFORMATICS

Unit 1: Molecular Biology and Biotechnology

Structure of DNA and RNA, Basics of replication, transcription and translation. Post-transcriptional and translational modifications. Transcriptional and translation control of prokaryotes and eukaryotes. Features of genetic code in prokaryotes and eukaryotes. Gene expression. General principles of recombinant DNA technology, restriction enzymes. Methods of gene transfer-plasmid and viruses as vectors, genomic and cDNA library construction, chromosome walking. Basics of genome organization and mapping, Non-coding RNA, Genome Editing, Gene silencing. Bio-chips.

Unit 2: Preliminaries of Bioinformatics

Overview of available genomic resources on the web; NCBI/ EBI/ EXPASY etc; Nucleic acid sequence databases; GenBank/ EMBL/ DDBJ; Database search engines: Entrez, SRS. Overview/concepts in sequence analysis; Pairwise sequence alignment algorithms: Needleman & Wunsch, Smith & Waterman; BLAST, FASTA; Scoring matrices for Nucleic acids and proteins: PAM, BLOSUM, Multiple sequence alignment: PRAS, CLUSTALW.

Unit 3: Genome assembly

Types and methods of genome sequence data generation; Shot gun sequencing method; Problems of genome assembly, Approaches of genome assembly: Comparative Assembly, DE novo Assembly; Read coverages; Sequencing errors, Sequence Quality Matrix, Assembly Evaluation; Challenges in Genome Assembly. Various tools and related methods of genome assembly: MIRA, Velvet, ABySS, ALLPATHS-LG, Bambus2, Celera Assembler, SGA, SOAPdenovo etc.

Unit 4: Evolutionary Biology

Phylogenetic trees and their comparison: Definition and description, various types of trees; Consensus (strict, semi-strict, Adams, majority rule, Nelson); Data partitioning and combination Tree to tree distances, similarity; Phylogenetic analysis algorithms: Maximum Parsimony, Distance based: UPGMA, Transformed Distance, Neighbors-Relation, Neighbor-Joining. Probabilistic models of evolution, Maximum likelihood algorithm; Approaches for tree reconstruction: Character optimization; delayed and accelerated transformation, Reliability of trees, Bootstrap, jackknife, decay, randomization tests.

Unit 5: Statistical Genomics

Frequency distributions, Graphical representations and Descriptive statistics. Elements of probability theory, Conditional probability, Bayes’ theorem. Random variable- discrete and continuous. Mathematical expectation. Moment generating and characteristic functions. Probability distributions-Binomial, Poisson and Normal distribution and its application. Sampling distributions and its properties- Chi-square, t, and F. Testing of hypotheses - types of errors, level of significance and power of a test, Tests of significance based on Z, t, chi-square and F distributions. Concept of sampling, Sampling vs. Complete Enumeration, Simple Random Sampling. Correlation and regression analysis.

Fundamentals of Population genetics: Gene and genotypic frequencies. Random mating and equilibrium in large populations, Hardy –Weinberg law, Effect of systematic forces on changes in gene frequency- Selection, mutation and migration. Equilibrium between forces in large population. Polymorphism. Fisher's fundamental theorem of natural selection. Polygenic systems for quantitative characters, Principles of Quantitative genetics: Values, Means and Variances, Concepts of breeding value, dominance, average effect of gene and epistatic interactions. Detection and Estimation of Linkage, Genetic variance and its partitioning. Correlation between relatives. Genotype and environment interaction. Estimation of genetic parameters. Genome wide association study.

Unit 6: Protein Structure Prediction

Nature of proteomic data; Overview of protein data bases; SWISSPROT, UniProtKB; PIR-PSD, PDB, Prosite, BLOCKS, Pfam/Prodom etc.; Structure analysis: Exploring the Database searches on PDB and CSD, WHATIF Molecular visualization tools; Visualization of tertiary structures, quaternary structures, architectures and topologies of proteins using molecular visualization softwares such as RasMol, Cn3D, SPDBV, Chime, Mol4D etc.

Structure prediction tools and homology modeling: Prediction of secondary structures of proteins using different methods with analysis and interpretation of the results; Comparison of the performance of the different methods for various classes of proteins. (Fasman method, Garnier Osguthorpe Robson (GOR), Neural Network based; methods); NLP approach for secondary structure prediction of RNA; Introduction to mfold and Vienna packages; Prediction of tertiary structures of proteins using Homology Modeling approach: SWISSMODEL, SWISS-PDB Viewer; along with analysis and interpretation of results. Molecular dynamics simulation and docking.

Unit 7: Biological Database Management System

Database Management System (DBMS): definition and purpose of DBMS, Advantages of DBMS, DBMS Architecture- Three level Architecture for DBMS –internal, conceptual and external levels and their Schemas and Mapping, Role of Schemas, Data Abstraction, Data independence-Physical and Logical data independence. Data Models-Relational, Network, Hierarchical. Relational data models (binary, ternary, quaternary & n-ary relations) Components: Relation, Tuple, cardinality, degree. Network model- Entity Relationship (E-R) model- Components of E-R model: Entities, Attributes, Relationships. Relational Databases-, Important terms in relational database system, Primary and Foreign keys. Relational Data Integrity and constraints: Domain Constraints, Entity Integrity, Referential constraints. Normal forms. Structured Query Language- Commands, Queries, Data Definition Language (DDL), Data Manipulation Language (DML). Primary, secondary and derived biological databases, submitting sequence to the Database and retrieval.

Unit 8: Bio-programming and Computational Biology

Object oriented programming, classes, objects, Abstract data types, Data types, Operators (Arithmetic, Logical and Comparison) and expressions. Data encapsulation- modules and interfaces; Polymorphism - Static and dynamic binding, Inheritance: class and object inheritance. Perl: Introduction, Scalar, Arrays and List Data, Control Structures, Hashes, String Handling, Regular Expressions; Subroutines, File handling, BioPERL modules. Machine learning techniques, Supervised and unsupervised learning, Discsion tree, hidden markov model, Artificial Neural Network, Support Vector Machine and genetic algorithms. Pre-processing of gene expression data; Data Normalization techniques, Data quality control: Modelling of errors, Imputation etc; High-throughput screening.