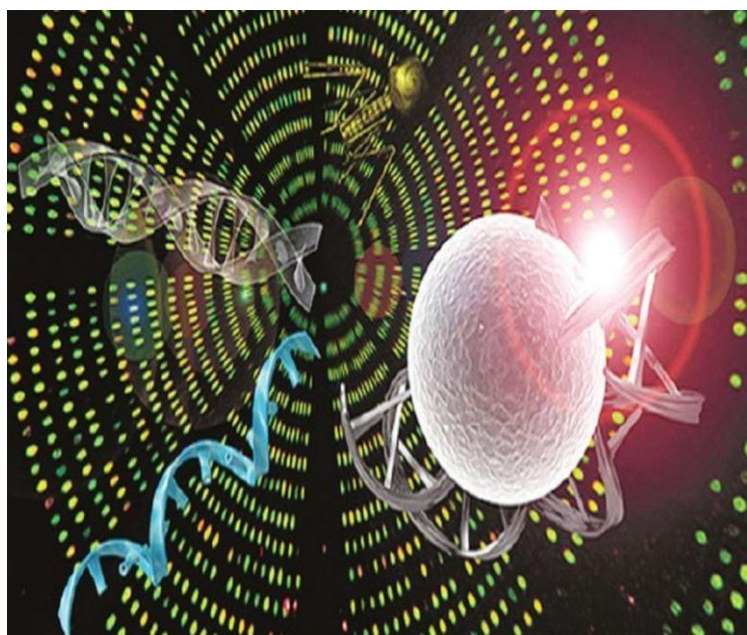




M.Sc. Bioinformatics Curriculum



Department of Computer Science
Jamia Millia Islamia (Central University)
Jamia Nagar, New Delhi-110025
<https://www.jmi.ac.in/computerscience>

1. INTRODUCTION

Master of Science in Bioinformatics, i.e., M.Sc. (Bioinformatics), is a 2-years (4-Semesters) length ‘academic, self-financed, and evening’ Programme of professional nature. It is primarily aimed to enable biological science graduates with bioinformatics skills for a career in the Post-Genomics era. However, its courses cater well specifically to biological science graduates but also suited to especially inspired aspirants from other backgrounds. It is now firmly accepted that computing and information technologies apply equally well to further research on proteins, genes, etc. Furthermore, the new age biological research highly driven by high-end computing and computational Modeling is inevitable. Therefore, the Programme includes training in the significant areas of Bioinformatics. In particular, the rigorous exposure to interdisciplinary areas aims to equip the students with strong analytical and technical skills to work competently in diverse areas, including teaching, research, and development.

2. PROGRAMME HIGHLIGHTS

| | | |
|----|--------------------------------|---|
| A. | Programme Name | M.Sc. (Bioinformatics) |
| B. | Nature | Regular, Self-Financed & Evening (Credit-Based SemesterSystem) |
| C. | Duration | 2 Years (4 Semesters) |
| D. | Programme Credits | 95 |
| E. | Theoretical Courses | 11 Courses (44 Credits) |
| F. | Lab-Oriented Theory Courses | 04 Courses (19 Credits) |
| G. | Laboratory Courses | 05 Courses (10 Credits) |
| H. | Elective Theory Courses | 02 (09 Credits) |
| I. | Major Project | 01 Course (20 Credits) |
| J. | Minor Project | 01 Course (02 Credits) |
| K. | OS Platforms | Windows, Unix, and Linux |
| L. | Programming Languages | C++, Python/Perl/Java, R, PHP |
| M. | Databases | Oracle, MySQL |
| N. | Bioinformatics Tools/Databases | BLAST, NCBI, EMBL, GenBank, DDBJ, Swiss Prot, RCSB, MEGA, KEGG, EMBOSS Transeq, Primer BLAST, Tandem Repeat Finder, Repeat Masker, ORF Finder, NebCutter, M-Fold, RNA-Fold, RADAR and P-Fold, SALI Lab, BHAGEERATH <i>ab-initio</i> , SPDBV, SmiLib, WhichCyp, ALOGPS, FINDSITE, AUTODOCK VINA, SPLINTER & SwissDock, Cytoscape, STRING, etc. |

3. PROGRAMME STRUCTURE

In order to achieve the aims and objectives set forth at the outset, a program structure under the semester-based credit system is prescribed as follows. The program structure prescriptions include the titles of the theory/lab courses, L-T-P, and assigned credits. Further, the respective detailed syllabi of theory courses are mainly based on the latest available edition of the book prescribed first in the list, with minor portions or special-topics-of-interest are based on the respective reference books. Moreover, prescriptions of ten generic practical assignments are included as suggestive lab exercises, on which the concerned teachers may work on specific assignments for student(s).

M.Sc. Bioinformatics Programme Structure

| SEM | CODE | PAPER-TITLE | L-T-P | CREDIT | SUMMARY |
|----------------------|--------|---|--------|--------|--------------------------------------|
| I | MBI-11 | Introduction to Bioinformatics | 3-1-0 | 4 | L: T:P = 15:5:10 Credit=25 |
| | MBI-12 | Computing Fundamentals | 3-1-0 | 4 | |
| | MBI-13 | Programming in C++ | 3-1-0 | 4 | |
| | MBI-14 | Comparative and Functional Genomics | 3-1-0 | 4 | |
| | MBI-15 | Programming in R | 3-1-2 | 5 | |
| | MBI-16 | Lab-I (Bioinformatics Tools) | 0-0-4 | 2 | |
| | MBI-17 | Lab-II (C++& Linux) | 0-0-4 | 2 | |
| II | MBI-21 | Systems Biology and Applications | 3-1-0 | 4 | L: T:P = 15:5:10 Credit=25 |
| | MBI-22 | Data Structure | 3-1-0 | 4 | |
| | MBI-23 | Essential Mathematics & Biostatistics | 3-1-0 | 4 | |
| | MBI-24 | Database Management System | 3-1-0 | 4 | |
| | MBI-25 | Elective-I | 3-1-2 | 5 | |
| | MBI-26 | Lab-III (MySQL/Oracle) | 0-0-4 | 2 | |
| | MBI-27 | Lab-IV (Systems Biology and Advance Bioinformatics Tools) | 0-0-4 | 2 | |
| III | MBI-31 | NGS Data Analysis | 3-1-0 | 4 | L: T: P=15:5:10 Credit=25 |
| | MBI-32 | Structural Bioinformatics and Drug Design | 3-1-0 | 4 | |
| | MBI-33 | R Advanced Algorithm for Computational Biology | 3-0-2 | 4 | |
| | MBI-34 | Web-based Programming | 3-1-2 | 5 | |
| | MBI-35 | Elective-II | 3-1-0 | 4 | |
| | MBI-36 | Lab-V (Drug Design + NGS) | 0-0-4 | 2 | |
| | MBI-37 | Minor Project | 0-0-4 | 2 | |
| IV | MBI-41 | Major Project | 0-2-20 | 20 | Credit=20 |
| TOTAL CREDITS | | | | | 95 |

Marks distribution for project evaluation – Project Report: 100 Marks; Implementation: 200 Marks; Presentation + Viva-voce: 100 Marks.

~~ ELECTIVES ~~

| Elective I | | Elective II | |
|------------|-------------------------------------|-------------|---|
| MBI-25.01 | Programming in Python & Biopython | MBI-35.01 | Artificial Intelligence in Bioinformatics |
| MBI-25.02 | Programming in Perl and Bioperl | MBI-35.02 | Machine Learning Techniques in Bioinformatics |
| MBI-25.03 | Object-Oriented Programming in Java | MBI-35.03 | Soft Computing Techniques in Bioinformatics |
| | | MBI-35.06 | Health Informatics |
| | | MBI-35.07 | Evolutionary Bioinformatics |
| | | MBI-35.08 | Advance Bioinformatics Algorithms |
| | | MBI-35.09 | Advance Bioinformatics Applications |
| | | MBI-35.10 | Advanced Systems Biology |
| | | MBI-35.11 | Immunoinformatics |
| | | MBI-35.12 | Parasite Biology |
| | | MBI-35.13 | Proteomics and Metabolomics |
| | | MBI-35.14 | Pharmacogenomics & Genome Informatics |

| | |
|-----------|--|
| MBI-35.15 | Biomedical Data Science |
| MBI-35.16 | Precision Medicine Informatics. |
| MBI-35.17 | Computational Software engineering. |
| MBI-35.18 | Algorithms and programming in health care. |

Detail Syllabi - 2023

M.Sc. Bioinformatics Programme

MBI-11: Introduction to Bioinformatics (3-1-0)

- 1. Introductory Concepts:** Introduction, Branches of Bioinformatics, aim and scope of Bioinformatics, Applications of Bioinformatics in Various Areas, Overview of Available Bioinformatics Resources on the Web, Protein and Genome; Information Resources and Analysis Tools; Established Techniques and Methods; Sequence and molecular File Formats FASTA, GenBank, FASTQ and Structured File Formats.
- 2. Biological Databases:** Databases in Bioinformatics- Classification Schema of Biological Databases, Biological Annotation and Data Curation, Biological Sequence Databases: Data Retrieval-NCBI, Protein database,PIR, Swiss Prot, Pfam, GEO , EMBL, DDBJ, KEGG, Protein 3D Structure and Classification database- RCSB, CATH,SCOP, DALI and TM-Align.
- 3. Sequence Alignment:** Why sequence alignment, Concept of alignment, Scoring matrices, PAM, BLOSUM, Alignment of pairs of sequences, Local and Global Alignment, Dynamic Programming Algorithm(DPA)- Needleman-Wunsch Algorithm, Smith-Waterman Algorithm., Statistics of Sequence Alignment Score; E-Value, P-Value, ClustalW, Heuristic Methods-FASTA, BLAST and its variant, Multiple Sequence Alignment ,Hidden Markov Model
- 4. Phylogenetic Analysis:** Distance and Character Based Methods and Software, Computing Tools for Phylogenetic Analysis, Distances, GROWTREE, PAUP, PHYLIP and MEGA; Construction and Visualization Phylogenetic Tree; and Application of Phylogenetic Analysis.
- 5. Protein Modeling and Drug Design:** Protein structure and Modeling, The Drug discovery process. .

REFERENCES

Zhumur Ghosh and Bibwkanand Mallick: Bioinformatics Principles and Applications, Oxford University Press
S.C.Rastogi,N.mendiratta and P.Rastogi: Bioinformatics Methods and Applications, PHI

Jin Xiong: Essential Bioinformatics, Cambridge University Press

Tramontano: Introduction to Bioinformatics, Chapman and Hall Series.

MBI-12: Computing Fundamentals (3-1-0)

- 1. Computing Concepts:** Basic Computing Systems, Layers of a Computing System, History of Computing, History of Computing Software, Stored-Program Concept and von Neumann Architecture. Fetch-Execute Cycle, RAM and ROM, Types of RAM and ROM, Secondary and Tertiary Storage Devices, Cache Memory, Memory Hierarchy, Input-Output Devices, Touch Screens.
- 2. Data Representation and Number Systems:** Binary Values and Computers, Data and Computers, Analog and Digital Data; Binary Representation. Number Systems: Binary, Octal, Decimal, and Hexadecimal. Conversions of Data from one Number System to another Number System. Representation of Numeric Data – Negatives and Real Data Representation. Representing Texts - ASCII and Unicode Character Sets. Binary Arithmetic – Addition and Subtraction of Numbers in Different Number Systems.

Gates and Circuits: Computers and Electricity; Logic Gates – AND, OR, NOT, XOR, NAND and NOR Gates. Gate Processing; Gates with More Inputs; Constructing Gates; Transistors; Circuits – Combinatorial Circuits: Adders and Multiplexers. Circuit as Memory; Integrated Circuits; CPU Chips.

3. Problem solving and Algorithm Design : Problem Solving, Problem-Solving strategies, Algorithms, The computer problem solving process, Pseudocode, Pseudocode Functionality, Top-Down design methodology, Object –Oriented Methodology

High Level Programming Languages: Translation process, Compilers, Interpreters, Programming language paradigms, Functionality of Imperative languages, Boolean expressions, Strong Typing, Input/Output structures, Control structures, Composite data type.

4. Operating Systems: Introduction and Examples of Operating Systems, Roles of Operating Systems: Memory, Process, and CPU Management. Batch Processing, Categories of Operating Systems: Timesharing, Multitasking, Multithreading, and Multiprogramming Operating Systems. Process States and Process Control Block.

5. Computer Networks: Introduction to Computer Networks. Network Topologies, Types of Networks, Internet vs. Intranet, Switch, Bridge, and Routers. Firewalls, Network Protocols, TCP/IP, Network Addresses, Domain Name System, Internet services: E-mail, FTP, ISP, etc.

REFERENCES

Dale & Lewis: Computer Science Illuminated, Narosa Publishing House
Kedall & Kendall: Systems Analysis and Design, Prentice Hall India

MBI-13: Programming in C++ (3-1-0)

1. Programming Basics: Introduction to Programs and Algorithms; Problem Solving Aspect (Algorithm Devising); Algorithm Design Aspect (Top-down Design); Algorithm Implementation; Program Verification. Flowchart. C++ Character Set; Keywords; Identifier, Constants and Variables; Data Types; Operators & Expression, Precedence & Associability of Operators; Managing Console I/O – Reading and Writing Characters.

2. Control Structures & Functions: Decision Making Structures – If Statement, If-Else Statement, Nested If-Else Statement, Else-If Ladder, Switch Statement, Looping Structures – While Statement, Do-While Statement, For Statement, Continue and Break Statements. Functions: Library Functions; User-Defined Functions; Call by Value and Call by Reference; Recursive Functions, Storage Class & Scope of Variables.

3. Classes and Objects: Unstructured Programming; Structured Programming; Object Oriented Programming; Abstract Data Type (ADT); Class; Object; Encapsulation; Polymorphism; Inheritance; Pros and Cons of Object Oriented Methodology, Friend Functions, Friends Classes; Inline Functions; Constructors, Parameterized Constructor, Destructor, Static Data Member, Static Member Functions, Creating Object; Passing Objects to a Function.

4. Arrays, Overloading & Inheritance: Concept of Array, Operations on Arrays, Array of Objects; Pointers, Pointer to Objects, this Pointer, Pointer to Derived Type, Reference Parameters, Passing References to Objects, new and delete Operators, Function Overloading: Rules to Overload a Function, Overloading Constructor Functions, Copy Constructors; Operator Overloading, List of Operators that cannot be Overloaded, Overloading Operator as a Member Function, Inheritance and its types, Base-Class Access Control, Multiple Inheritance, Virtual Base Class, Virtual functions.

5. Generic Function, Exception and File Handling: Generic Functions and its Benefits, Explicitly Overloading a Generic Function, Overloading a Function Template, Generic Function Restriction, Generic Class. Exception Handling, try and catch Blocks, throwing User Defined Exception. C++ Streams, C++ File Handling, Opening and Closing a File, Reading and Writing a Text File, Random Access, Reading and Writing Object to a File.

REFERENCES

Balaguruswamy Object Oriented Programming with C++, Tata McGraw Hill
Herbert Schildt: Complete Reference C++, Tata McGraw Hill
R1 H.M. Deitel& P.J. Deitel: C++ How to Program, Pearson Education.

MBI-14: Comparative and Functional Genomics (3-1-0)

1. Comparative genomics: Databases for genomics resources, Technologic principles of currently utilised methods within the field of functional genomics. Sequence alignment: global versus local, Inferring gene function from relatedness to other genes, Finding Open Reading Frames, overall structure and organisation of the genome in microorganisms, plants and animals. Extra-chromosomal elements: Mitochondrial genome, mitochondrial plasmids. Organization of nuclear and organellar genomes; Repetitive DNA-satellite DNAs and interspersed repeated DNAs, Transposable elements, LINES, SINES, Alu family and their application in genome Mapping.

2. Molecular Phylogenetics: Fine structure of gene, split genes, pseudogenes, non-coding genes, overlapping genes and multi-gene families. Genome sequencing methods, Genome annotation at different levels, Comparative genome sequencing. Molecular Phylogenetics Origins, Applications of Molecular Phylogenetics. Phylogenomic comparisons, introduction to programs, Phylogenetic analyses: tree terminology and parsimony, Phylogenetic analyses to determine relationships and interpret character evolution. Single gene disorders- conventional and contemporary methods. Candidate gene identification; Genetic polymorphism and disease susceptibility.

3. Databases and Software for Genome Biology: Gene Expression Analysis using Microarrays and RNA-Seq, Application of DNA microarrays for the analysis of gene expression, protein-DNA binding, chromatin structure, chromatin modifying complexes, and RNA polymerase occupancy. Error models and data normalization techniques for high-resolution array technologies. Clustering genes into sets and discovering gene set features that can be used for diagnostic purposes. Importance of chromatin structure in contemporary modelling, relationship between chromatin structure and transcriptional regulation.

4. Functional Genomics-I: Determining the Functions of Individual Genes, Pathway and GO annotation systems, Non-coding DNA, gene prediction methods and tools. Genome Evolution, Genome and Transcriptome Assembly Algorithms, Profile HMMs for Protein Family Modelling, Human Genome Project, Genomics and Personalized medicine, Sex-linked inheritance: Conceptual basis, sex influenced traits, mechanism of sex determination. Mutation variability across the genome. Mutation detection, diagnosis and therapy. Genes and Environment - heritability, penetrance and expressivity.

5. Functional Genomics-II: Integration of genomic, epigenetic and external effects in the context of understanding gene and genome function, Genome regulation: Introduction, levels of regulation, evidences and experimental designs/methodologies, role of genetic analysis in understanding gene function and regulation. Transcriptional control - DNA sequence elements and transcription factors. Mutation – Classification, mechanism, repair, role in genetic analysis and evolution. Post-transcriptional regulation – Alternative RNA splicing, RNA editing, RNA transport and localization, RNA stability, Regulation of translation – RNA structure, control at initiation, codon usage, Post-translational modifications. RNA-mediated control of gene regulation.

REFERENCES

Hartl & Ruvolo. Genetics: Analysis of Genes and Genomes. Jones & Bartlett Learning
Mount. Bioinformatics sequence and Genome analysis. Cold Spring Harbor Laboratory Press
Deonier, Tavaré, & Waterman. Computational Genome Analysis: An Introduction. Springer



MBI-15: Programming in R (3-1-2)

1. Basic Data Types: Numeric, Integer, Complex, Logical, Character, **Vector:** Combining Vectors, Vector Arithmetic, Vector Index, Numeric Index Vector, Logical Index Vector, Named Vector Members.

2. Matrix & Probability Distributions: Matrix Construction, List: Named List Members, Data Frame, Data Frame Column Vector, Data Frame Column Slice, Data Frame Row Slice and Data Import. Probability Distributions: Binomial Distribution, Poisson distribution, Continuous Uniform Distribution, Exponential Distribution, Normal Distribution, Chi-squared Distribution, Student t Distribution and F Distribution

3. Qualitative Data: Frequency Distribution of Qualitative Data, Relative Frequency, Distribution of Qualitative Data, Bar Graph, Pie Chart, Category Statistics. Quantitative Data: Frequency Distribution of Quantitative Data, Histogram, Relative Frequency Distribution of Quantitative Data, Cumulative Frequency Distribution Cumulative Frequency Graph, Cumulative Relative Frequency Distribution Cumulative Relative Frequency Graph, Stem-and-Leaf Plot and Scatter Plot

4. Numerical Measures & Interval Estimation: Mean, Median, Quartile, Percentile, Range, Interquartile Range, Box Plot, Variance, Standard Deviation, Covariance, Correlation Coefficient, Central Moment, Skewness and Kurtosis. Interval Estimation: Point Estimate & Interval Estimate of Population Mean with Known/Unknown Variance, Sampling Size of Population Mean, Point Estimate of Population Proportion, Interval Estimate of Population Proportion and Sampling Size of Population Proportion.

5. Packages & Bioinformatics Applications: Designing Packages, Popular Packages for Bioinformatics applications, Introduction to Bioconductor, R packages for Microarray/NGS Data analysis, Protein Structure visualization, etc.

REFERENCES

Sandip Rakshit: R Programming for Beginners. McGraw Hill Education
Lander: R for Everyone: Advanced Analytics and Graphics. Pearson Education India

MBI-16: Lab-I (Bioinformatics Tools) (0-0-4)

Implementation of at least ONE specific assignment concerning each of the following:

1. Retrieving, viewing and printing of the specific protein sequence (by accession no. or name) using a public database site.
2. Exploring the NCBI, ExPASy, www.ebi.ac.uk/Tools etc. websites for information and tools available there.
3. Pairwise alignment of Protein and DNA sequences & data interpretation.
4. Local and global alignment of sequence data and comparing both results.
5. Retrieving DNA and/or protein sequences of a given item (by name or accession number) from

GENBANK. Performing a sequence similarity search using the BLAST.

6. Retrieving this protein sequence of a given organism and downloading the structure of this protein from existing database. Short-listing protein sequences of highest similarity from the list of BLAST search result and doing a multiple sequence alignment (Using CLUSTALW). Finding out the regions of exact/good match in the protein sequences of these sequences.
7. Aligning nucleotide sequences; designing a degenerate primer of 20 bases from nucleotide alignment data, and calculate the level of degeneracy of this primer.
8. Learning about the Phylip/MEGA program and its uses for the construction of phylogenetic trees.
9. Searching and downloading protein structure data using Entrez. Viewing the structure using public domain software.
10. Protein structures: Visualizing and analysis of inter atomic distances, H-bond calculations, secondary structure analysis and salt bridge analysis of protein structures using different software. Prediction of 3D structure of protein.

MBI-17: Lab-II (C++& Linux) (0-0-4)

Implementation of at least ONE specific assignment concerning each of the following:

1. Basic Linux commands
2. Basic C++ I/O functions/operators
3. Conditional statements: if-else, and switch
4. Looping and unconditional branching statements
5. Arrays and matrix operations
6. Defining classes and creating objects
7. Defining constructors, destructions, Array of objects
8. Inheritance& Virtual Functions
9. Template – Implementing genetic searching and sorting algorithms
10. File and file operations

MBI- 21: Systems Biology and Applications (3-1-0)

- 1. Systems Biology:** System-level-Understanding of Biological Systems, Introduction, Measurement Technologies and experimental methods, Comprehensive Measurements, Measurement for Systems Biology, Next-generation Experimental Systems. System structure identification, Bottom-up-approach, Top-down-approach. Application areas of Systems Biology.
- 2. System Behavior Analysis:** Simulation, Analysis Methods, Robustness of Biological Systems, Lessons from Complex Engineering Systems. System Control; Redundancy, Modular Design, Control, Structural Stability, Impacts of systems Biology.
- 3. Modeling Genetic Networks:** Why Modeling is necessary, What type of Modeling is appropriate, Modeling the activity of a single gene, Gene Regulatory Network(GRN) Understanding gene regulation, Understanding the Biology, Biochemical Processes; transcription, exons & introns, splicing, translation, post translation modification. Overview of Models; Boolean, Differential equation, stochastic Models, Kinetic Logic Model.
- 4. The Analysis of Cancer Associated Gene Expression Matrices:** Separators, Identifications of separators in noisy data, Genetic algorithms, Statistical validation of separators extracted from gene expression matrices, Generative models, Randomization based generative models.
- 5. Reverse Engineering from Gene Expression Data:** The DBRF Method for Inferring a Gene Network from Large-Scale Steady-State Gene Expression Data; The difference based regulation finding methods, Inference of a Redundant Gene Regulatory Network, Computational Experiments; Network Models. Automated Reverse Engineering of Metabolic Pathways by Means of Genetic Programming.

REFERENCES

Hiroaki Kitano: **Foundations of Systems Biology**, MIT Press

Bower & Bolouri: **Computational Modeling of Genetic & Biochemical Networks**, MIT Press

Alon: **Introduction to Systems Biology of Biological Circuits**, Chapman & Hall/CRC.

MBI- 22: Data Structure (3-1-0)

- 1. Arrays and Matrices:** Array as an Abstract Data Type (ADT); One Dimensional Array; Multi Dimensional Array; Matrix Representation using 2D Arrays - Row-major Order, Column-major Order; Special Matrices: Diagonal, Tridiagonal, Lower Triangular and Upper Triangular Matrices; Sparse Matrices: Representation and Transpose; Addition of Sparse Matrices.
- 2. Linked Lists:** Single Linked List; Static Representation of Linked List; Dynamic Representation of Linked List; Operations on Single Linked List — Creating, Traversing, Insertion, Deletion, Copy; Merging, Searching; Circular Linked List; Double Linked List; Circular Double Linked List.
- 3. Stacks and Queues:** Introduction to Stacks; Array Representation of Stack; Linked Representation of Stack; Operations on Stacks; Applications of Stack - Infix Expression to Postfix Conversion, Evaluation of Postfix Expression, Implementation of Recursive Functions, Towers of Hanoi. Introduction to Queues; Array Representation of Queue; Linked Representation of Queue; Applications of Queue — CPU Scheduling.
- 4. Trees:** Basic Concepts of Tree; Binary Trees; Types of Binary Tree; Properties of Binary Trees; Representation of Binary Trees. Operations on Binary Trees: Creation, Insertion, and Deletion of Nodes; Binary Tree Traversals — In-order, Pre-order, and Post-order Traversals; Expression Tree; Binary Search Trees (BST); Operations on BST: Search, Insertion and Deletion; B Tree and B+ Tree.
- 5. Searching and Sorting Techniques:** Searching vs. Browsing. Searching Techniques: Linear search and Binary Search. Sorting Techniques: Bubble Sort, Selection Sort, Insertion Sort, Merge Sort, Quick Sort, Radix sort, etc.
- 6. Graphs:** Basic Concepts Related to Graph; Difference between Tree and Graph; Properties of Graph; Graph Representations: Adjacency Matrix; Weighted Graph Representations; Graph Traversing Methods: Breadth-First Traversal and Depth-First Traversal.

Samanta: **Classic Data Structure**, Prentice Hall of India

Sahni: **Data Structures Algorithm and Applications**, University Press Kruse, Tondo & Leung **Data Structures and Program Design in C**, Pearson Education

MBI-23: Essential Mathematics & Biostatistics (3-1-0)

- 1. Basic Mathematics:** Vectors: Vector algebra, Dot and Cross products. Matrices: Algebra of Matrices, Transpose and inverse, Diagonalization of Matrices and Characteristic roots. Basic Differentiation and Integration Techniques.
- 2. Central Tendency and Dispersion:** Terminologies: Population, Sample, Variable, Parameter, Primary and Secondary Data; Screening, Representation and Classification of Data; Frequency Distribution, Tabulation and Graphical Representations; Mean, Geometric Mean, Harmonic Mean, Median and Mode; Quartiles and Percentiles; Measures of Dispersion: Range, Variance, Standard Deviation, and Coefficient of Variation.
- 3. Probability and Distributions:** Sample space, events, Equally likely events. Probability and types; Different Approaches, Independent Events, Addition and multiplication rules, Conditional probability, Rules for Calculating Probabilities.
- 4. Correlation & Regression:** Methods of studying simple correlation: Scatter Diagram, Karl Pearson's Co-efficient of Correlation, Spearman's Rank Correlation. Multiple and Partial Correlation, Correlation ratio and intra-class correlation; Regression Analysis: Introduction: Meaning and Purpose of Regression. Simple and Multiple Linear Regressions using Least Square Principle, Coefficient of Determination.
- 5. Analysis of Variance and Testing Hypothesis:** Assumptions for Use of the ANOVA, One-Way ANOVA, F-test, Two-way ANOVA for the Randomized Block Design, Linear Model Representation of the ANOVA. Sampling, Test of Significance, Testing Hypotheses about the Variance, Testing Hypotheses about the difference of two means, Testing Hypotheses about the difference of two Proportions, Tests of Hypotheses and Sample Size. Parametric and nonparametric test. The Goodness-of-Fit Test; Chi Square Test.

REFERENCES

- Daniel & Cross: Biostatistics-Basic Concepts and Methodology for the Health Sciences, Wiley**
Majajan: Methods in Biostatistics, Jaypee
Forthofer, Lee & Hernandez: Biostatistics: A Guide to Design, Analysis, and Discovery, Academic Press.

MBI-24: Database Management System (3-1-0)

- 1. Basic Concepts & ER Diagram:** Data, Database and Database Management System; Database vs. Traditional File Systems; Three-Tier Architecture of DBMS, Data Independence; Categories of DBMS: Hierarchical, Network and Relational Database, Data and Database Models, Categories of Database Models, Entity relationship (ER) Model: Basic Concepts & their Representations, Entity, Entity Type, Entity Set, Attributes and Keys, Relationships, Relationship Types, Structural Constraints, Weak Entity, Naming Conventions, Design Issues in ER Model. ER Diagrams.
- 2. Relational Database Model:** Structure of Relational Model; Domains, Attributes, Tuples, and Relations; Characteristics of Relations; Relational Constraints – Domain Constraints, Key Constraints, Entity Integrity, and Referential Integrity Constraints; Relational Database Schema.
- 3. Functional Dependencies and Normalization:** Informal Design Guidelines for Relation Schemas; Functional Dependencies; Inference Rules for Functional Dependencies; Normalization using Functional Dependencies – First Normal Form (1NF), Second Normal Form (2NF), Third Normal Form (3NF), and Boyce-Codd Normal Form (BCNF).
- 4. Relational Algebra and SQL:** Relational Algebra Operations – Select, Project, Rename, Union, Intersection, Set Difference, Join, and Division Operations; Aggregate Functions and Groupings. SQL Concepts, Schema and Table Deletion; Table Modification; Insert, Delete, and Update Statements; SELECT-FROM-WHERE Structure; Renaming Attributes; Nested Queries and Set Comparisons; Aggregate Functions.
- 5. Advanced Database Concepts:** Emerging database technologies and applications: Spatial databases, Mobile databases, Multimedia databases, Genome data management, Overview of Data Warehouse, OLAP, Big data, biological big data, and big data analytics.

REFERENCES

- Elmasri and Navathe: Fundamentals of Database Systems, Pearson Ed.**
Han & Kamber: Data Mining Concepts and Techniques, Morgan Kaufman Publisher.
Loney & Koch: ORACLE 9i – The Complete Reference, TataMcGraw-Hill.

MBI-25: Elective-I (3-1-0)

MBI-25.1: Programming in Python & Biopython

- 1. An Introduction to Python Programming:** Working with Python, An interpreter for python, Relational operators, Logical operators, Bitwise operators, Variables and assignment Statements, Keywords, Script mode.
- 2. Basic Concepts:** Control structures, if-else conditional statement, Looping statements, Nested loops, break, continue and pass, Debugging, Scope of variables, Strings, String manipulations, Regular Expressions, Built-in Functions, I/O Functions, Function Definition and Call, Importing user-defined modules, Command-line arguments, Mutable and Immutable objects, Recursion.
- 3. Advanced Concepts:** Lists, Accessing lists, Working with lists, Operations, related Functions and Methods, Tuples, Accessing tuples, Working with tuples, Operations, related Functions and Methods, Dictionary, Working with dictionary, Accessing values in dictionaries, Working with dictionaries, Operations, related Functions and Methods. Files and Exceptions: File Handling, Writing structures to a file., Errors and Exceptions, Handling exceptions using try-except, File processing examples.
- 4. OOP concepts:** OOPs concepts, Classes and objects, Constructor, Destructor, Attributes, Encapsulation, Data Hiding and Data Abstraction, , Inheritance, Polymorphism, Overloading, overriding., Inbuilt Object-Oriented functions and modules. Managing Databases using SQL.
- 5. Biopython:** Introduction to Biopython, Installation, Inbuilt modules related to sequence objects, sequence annotation objects, sequence analysis, sequence input/output, sequence alignment objects and tools, Applications of Biopython. Overview of Scikit module.

References

- Taneja & Kumar: Python Programming: A Modular Approach, Pearson**
Kenneth & Lambert: Fundamental of Python. Course Technology
Chang, Chapman, et al. Biopython Tutorial and Cookbook (ebook).

MBI-25.2: Programming in Perl & Bioperl

- 1. Introduction:** History of Perl, Availability, Support, Versions, Installation. Basic Concepts, Significance of Perl in Bioinformatics.
- 2. Basics Constructs:** Scalar Data, Numbers, Strings, Scalar Operators, Scalar Variables, Scalar Operators and Functions. Arrays and List Data: Literal Representation, Variables, Array Operators and Functions, Scalar and List Context. Hashes: Hash Variables, Literal Representation of a Hash, Hash Functions, Hash Slices. Control Statement blocks, Loops and Conditions. Basic Input/Output.
- 3. Advance Constructs and Features:** Regular Expressions: Concepts of Regular Expressions, Simple Usage of Regular Expressions, Patterns, Matching Operators, Substitutions, Split and Join functions, Subroutines: System and User Functions, Local Operator, Length, Parameter Lists, Lexical Variables, File Handles and File Tests: Opening and Closing a File handle, Using Pathnames and Filenames, Die, Using File handles.
- 4. Object-Oriented Perl:** Introduction to Modules, Creating Objects and References. CGI Programming: The CGI.pm Module, CGI Program in Context, Simple CGI Programs, Passing Parameters via CGI, Perl and the Web.
- 5. Application of Perl in Bioinformatics:** Representing String and Sequenced Data in Perl, Program to Store a DNA Sequence, Concatenating DNA Fragments, DNA to RNA Transcription, Reading Proteins from File, Finding Motifs, Counting Nucleotides, Exploding Strings into Arrays, Operating on Strings, Generating random DNA, Analyzing DNA, Genbank, Sequence and Annotation, Parsing PDB Files, Parsing BLAST Output, Reading FASTA files. BioPerl Overview, Installation Procedures; Fundamental Constructs and Special Features; BioPerl Modules, Creating BioPerl Objects.

MBI-25.3: Object-Oriented Programming in Java

- 1. Introduction:** Environment and Programming Structure: Java White Paper Buzzwords, History of Java, Choosing a Development Environment: Command-Line Tools, Running a Graphical Application, A Simple Java Program, Comments, Data Types, Variables, Operators, Input and Output, Control Flow, Big Numbers, Arrays.

2. Class, Objects and Inheritance: Introduction to OOP, Predefined Classes, User Defined Classes, Static Fields and Methods, Method Parameters, Object Construction, Packages, Class Path, Class Design; Inheritance: Super-classes and Subclasses, Types of Inheritance, Polymorphism, Abstract class, Object: The Cosmic Super class, Generic Array Lists, Inheritance Guidelines.

3. Interfaces, String Handling and Exceptions: Interfaces, Object Cloning, Interfaces and Callbacks, Inner Classes, Proxies, String Handling APIs: String, Immutable String, String Comparison, String Concatenation, Substring, Methods of String Class, StringBuffer Class, StringBuilder Class, Creating Immutable Class, StringTokenizer Class. Exception Handling.

4. Generic Programming & Collection: Definition of Generic Class, Generic Methods, Bounds for Type Variables, Generic Code and the Virtual Machine, Restrictions and Limitations, Inheritance Rules for Generic Types, Wildcard Types, Reflection and Generics. Collection Interfaces, Concrete Collections, The Collections Framework, Algorithms, Legacy Collections

5. Java GUI Programming & JDBC: Introduction to Swing, Creating a Frame, Positioning a Frame, Displaying Information in a Component, Displaying Images, Event Handling, Basics of Event Handling, Actions, Mouse Events, The AWT Event Hierarchy; JDBC: Basic JDBC Programming Concepts, Installing JDBC, Statements, Executing Queries, Result Sets.

REFERENCES

Horstmann & Cornell: Core Java Volume I: Fundamentals, Pearson Education
Horstmann & Cornell: Core Java Volume II: Advanced Features, Pearson Education.
Dietel & Dietel: Java How To Program, Pearson Education

MBI-26: Lab-III (MySQL/Oracle) (0-0-4)

Implementation of at least ONE specific assignment concerning each of the following:

1. SQL statements to create, update, and delete databases and tables
2. SQL statements to insert, update, and delete records from tables
3. SQL statements to create, update, and delete views
4. Simple SQL queries to retrieve information from a database
5. Nested SQL queries to handle complex information retrieval requirements.
6. SQL queries using aggregate functions like count, average, sum, etc.
7. PL/SQL blocks using basic data types and operators
8. PL/SQL blocks using branching and looping constructs
9. Database triggers using PL/SQL
10. Database functions/procedures using PL/SQL.

MBI-27: Lab-IV (Systems Biology and Advance Bioinformatics Tools) (0-0-4)

Implementation of at least ONE specific assignment concerning each of the following:

1. Visualization of biological networks (graph) using Cytoscape/other tools.
2. Topological analysis of biological networks using Cytoscape/other tools.
3. Statistical analysis of high-throughput gene expression data.
4. Implementation of gene regulatory network reconstruction algorithm.
5. Implementation of modelling methods for metabolic network
6. Working with CellDesigner or other biological network editing tools.
7. GO enrichment and DAVID analysis
8. Searching of adjacent and patterned repeats of nucleotide sequences using Tandem Repeat Finder or Repeat Masker.
9. Open Reading Frames using ORF Finder. Finding probable restriction sites for different enzymes using NebCutter tool.
10. RNA structure prediction using M-Fold, RNA-Fold, RADAR and P-Fold server.

MBI-31: NGS Data Analysis (3-1-0)

1. Introduction to NGS: Generation to DNA sequencing technologies, A Typical NGS Experimental Workflow, Different NGS Platforms – Illumina, Ion Torrent Semiconductor Sequencing, Pacific Biosciences SMRT, ONT Nanopore; Major Applications of NGS

2. Base Calling, Quality Control & Read Mapping: Base Calling, FASTQ File Format, Base Quality Score, NGS Data Quality Control and Preprocessing; Reads Mapping – Mapping Approaches and Algorithms, Selection of Mapping Algorithms and Reference Genome Sequences, SAM/BAM as the Standard Mapping File Format, Mapping File Examination and Operation, Tertiary Analysis, NGS Data Storage, Transfer, and Sharing, Computing Power Required for NGS Data Analysis, Bioinformatics Skills & Software Required for NGS Data Analysis.

3. Transcriptomics by RNA-Seq: Principle of RNA-Seq; Experimental Design: Factorial Design, Replication and Randomization, Sample Preparation, Sequencing Strategy; RNA-Seq Data Analysis: Data Quality Control and Reads Mapping, RNA-Seq Data Normalization, Identification of Differentially Expressed Genes, Differential Splicing Analysis, Visualization of RNA-Seq Data, Functional Analysis of Identified Genes; RNA-Seq as a Discovery Tool. Small RNA Sequencing: Data Generation, Preprocessing, Mapping, Identification of Known and Putative Small RNA Species, Normalization, Identification of Differentially Expressed Small RNAs, Functional Analysis of Identified Small RNAs.

4. Genotyping and Genomic Variation Discovery: Data Preprocessing, Mapping, Realignment, and Recalibration; Single Nucleotide Variant (SNV) and Indel Calling: SNV Calling, Identification of de novo Mutations, Indel Calling, Variant Calling from RNA-Seq Data, Variant Call Format (VCF) File, Evaluating VCF Results. Structural Variant (SV) Calling: Read-Pair-Based SV Calling, Breakpoint Determination, De novo Assembly-Based SV Detection, CNV Detection, Integrated SV Analysis; Annotation of Called Variants, Testing of Variant Association with Diseases or Traits.

5. De novo Genome Assembly & ChIP-Seq Analysis: Genomic Factors and Sequencing Strategies for de novo Assembly, Genomic Factors That Affect de novo Assembly, Sequencing Strategies for de novo Assembly; Assembly of Contigs, Sequence Data Preprocessing, Error Correction, and Assessment of Genome Characteristics, Contig Assembly Algorithms; Scaffolding, Assembly Quality Evaluation, Gap Closure, Limitations and Future Development. Principle of ChIP-Seq, Experimental Design: Experimental Control, Sequencing Depth, Replication; Read Mapping, Peak Calling, and Peak Visualization, Differential Binding Analysis, Functional Analysis, Motif Analysis, Integrated ChIP-Seq Data Analysis.

REFERENCES

Wang: Next-Generation Sequencing Data Analysis, CRC Press.

Data & Nettleton (Eds). Statistical Analysis of Next Generation Sequencing Data. Springer.

Sung. Algorithms for Next-Generation Sequencing. CRC Press.

MBI-32: Structural Bioinformatics and Drug Design (3-1-0)

1. Fundamentals of macromolecular structure: Functional significance of structure, principles of protein structure: secondary structures, motifs, domains, tertiary and quaternary structures. Relationship between sequence and 3D structure of a protein. Structural implications of the peptide bond; rigid planar peptide unit; cis and trans configuration; conformations of a pair of linked peptide units; Ramachandran plot and Protein Data Bank. Conformations of biomacromolecules, Thermodynamics of protein folding.

2. Protein Structure Visualization Tools (PyMOL, Chimera, VMD), Protein Structure Comparison Tools and Methods, Protein Structural Alignment (Dali, CE, TM-Align), Protein Structure Classification and Databases (PDB, CATH, SCOPE, CDD), Drug design software, Enzyme kinetics and inhibition. Rational approaches to lead discovery based on traditional medicine, Random screening, Non-random screening, serendipitous drug discovery. Pharmacophore, Lock-Key principle and induced fit theory.

3. Structure Prediction: Protein Secondary Structure Prediction, Chou-Fasman, GOR Method, Artificial Neural Networks, Protein Secondary Structure Prediction Tools and Servers (DSSP, STRIDE, Jpred, Psipredetc), Protein Tertiary Structure Prediction, Homology Modelling, Refinement of the homology model, Structure Validation Strategies, Comparison of various strategies in homology modelling, Prediction of protein structures by threading, Energy Minimization methods and Conformational Analysis, global conformational minima determination.

4. Drug Design: Introduction to drug discovery, Current approaches and philosophies in drug design, Molecular Modelling and virtual screening techniques, lead identification, Concept of pharmacophore mapping and pharmacophore-based Screening, Analysis of the receptor, Docking: Rigid docking, flexible docking, manual docking, docking based screening, Scoring Functions, Binding energy calculations *De novo* drug design., Drug likeness screening, ADMET properties.

5. Quantitative Structure Activity Relationship (QSAR): SAR versus QSAR, QSAR methodology, Quantum-chemical descriptors, COMFA and COMSIA. Molecular mechanisms of diseases and drug action. Pharmacokinetics: Models and Applications, Drug-receptor interactions. Pharmacodynamics, Biomolecular interactions and binding thermodynamics, Biochemical and cell-based assays. Molecular recognition and specificity. Future developments.

REFERENCES

Gu & Bourne (Editors): Structural Bioinformatics, Wiley-Blackwell

Jhoti & Leach (Editors): Structure-based Drug Discovery, Springer

MBI-33: Advanced Algorithm in Computational Biology (3-1-2)

- 1. Algorithms in Computing:** Biological and Computer algorithm, Algorithm design techniques- Exhaustive search, branch-and bound, Dynamic Programming, Randomized Algorithm and machine learning. Time and space complexity of algorithms., Search Algorithms: Random walk, Hill climbing, simulated annealing.
- 2. Combinatorial Pattern Matching:** Hash Tables, Repeat Finding, Exact Pattern Matching; Genetic Algorithm: Basic Concepts, Reproduction, Cross over, Mutation, Fitness Value, Optimization using GAs; Applications of GA in bioinformatics.
- 3. Markov Chains and Hidden Markov Model:** Introduction to Markov Chains and Hidden Markov Model (HMM), Viterbi Algorithm -the Forward Algorithm, The Backward Algorithm and Posterior State Probabilities. Pairwise Alignment using HMM.
- 4. Support Vector Machines:** Introduction, hyperplane separation (maximum and soft margin hyperplanes), linear classifier, Kernel functions, Large Margin Classification, Optimization problem with SVM, Applications of SVM in bioinformatics. Bayesian network: Bayes Theorem, Inference and learning of Bayesian network, BN and Other Probabilistic Models.
- 5. RNA Structure Analysis:** Introduction to RNA. Terminologies Related to RNA secondary Structure. Inferring Structure by Comparative Sequence Analysis, RNA secondary Structure Prediction, Base Pair Maximization and the Nussinov Folding Algorithm. Covariance Models: SCFG-based RNA profiles.

REFERENCES

1. An introduction to bioinformatics algorithms by Neil C. Jones, Pavel Pevzner. MIT Press.
2. Durbin, Sean, Krogh & Mitchison. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, Cambridge University Press
3. Rastogi S. Bioinformatics, Concept, Skills and Applications, CBS
4. Baxevanis & Francis: Bioinformatics-A Practical Guide to the Analysis of Genes and Proteins. Wiley Mount D. Bioinformatics: Sequence and Genome Analysis. CBS.
7. Quality checking and trimming using freely available software (e.g. FastQC).

MBI-34: Web-based Programming (3-1-2)

1. Introduction to Internet and Web Programming: Brief History of the Internet, WWW, Web System Architecture, Internet vs. Intranet, URL, Protocol used in Internet: TCP/IP, SMTP, PPP, HTTP(s), Services on the Internet, Web Servers, Overview of Web Authoring Tools, Design Frameworks.

2. Hypertext Markup Language (HTML) and HTML5: Introduction to HTML and HTML5, Basic Structures of a HTML/HTML5 Documents, HTML5 Semantic Elements: Header, Footer, Article, Section. Ordered & Unordered Lists, Hyperlinks, Working with Table, Working with Forms: Form and Input Tags, Text Box, Radio Button, Checkbox, Select Tag and Pull Down Lists, Hidden, Submit and Reset, Attributes of HTML5 Form Elements: Number, Date, Time, Calendar, Range, HTML5 Drag and Drop, Graphics: Canvas, SVG, Multimedia Elements: Audio, Video.

3. Cascading Style Sheets (CSS) and JavaScript: Benefit of CSS, CSS Properties, CSS Styling: Background, Text Format, Controlling Fonts, Working with Lists and Tables, CSS ID and Class, Web Page Layout and Editing with CSS, Writing JavaScript into HTML, Basic Programming using JavaScript, JavaScript Client Validations, Dialog Boxes, Event Handling.

4. Introduction to Hypertext Preprocessor (PHP): Evaluation of PHP, Basic Syntax, Defining variable and constant, Data Types, Operator and Expression, Handling HTML Form With PHP: Capturing Form Data, Dealing with Multi-valued Field, Conditional Statement, Iterations, Arrays, Working with Functions: Call-by-value and Call-by-reference, Recursive Function, Working with String. String Matching with Regular Expression: Pattern Matching in PHP, Replacing Text, Splitting a String with a Regular Expression.

5. Web Server & Database Connectivity: Introduction to WAMP/XAMPP Server – Configuration and Web Application Deployment, PHP Server Variables, State Management - Query String (URL Rewriting), Hidden field, Cookies, Session. Database Connectivity with MySQL: Connection with MySQL Database, Basic Database Operations – Insert, Delete, Update and Select, Setting Query Parameter, Executing Query, Join.

REFERENCES

Robert W. Sebesta, Programming the World Wide Web, Addison Wesley

Bayross, Web Enabled Commercial Application Development Using HTML, JavaScript, DHTML and PHP, BPB
Dick Oliver, Michael Morrison, Sams Teach Yourself HTML and CSS in 24 Hours, Pearson Education.

Ivan Bayross, HTML 5 and CSS 3 Made Simple, BPB

MBI-35: Elective II (3-1-0)

Syllabus is based on the selection of Elective Course by the Students.

MBI-36: Lab-V (Drug Design + NGS) (0-0-4)

Implementation of at least ONE specific assignment concerning each of the following:

1. Generating drug library using blaster tool and analysing drug properties using Drug Classification tool
2. Predicting cytochromes P450 inhibition using WhichCyp tool & analysing ADMET properties using ALOGPS2.
3. Predicting binding site in the protein molecule using FINDSITE tool & estimating binding free energy using BAPPLserver.
4. Designing ligand molecule using eDesign tool, Ligand-Protein Docking using AUTODOCK VINA&SwissDock tool.
5. Searching docking resources for structure-based drug design using SPLINTER.
6. Working with NGS databases (NCBI-SRA, etc.), NGS file formats, File format conversion.
7. Quality checking and trimming using freely available software (e.g. FastQC).

8. Read mapping using BWA or Bowtie or any other freely available software, Interpretation and Visualization of Sequence Alignment Map (SAM) file.
9. Identification of SNPs using GATK Pipeline.
10. RNA-Seq and Chip-seq analysis using cloud-based server (e.g. Galaxy).

MBI-37: Minor Project (0-0-4)

Consult Course Co-ordinator for complete guidelines.

MBI-41: Major Project (0-2-20)

Consult Course Co-ordinator for complete guidelines.



Department of Computer Science

Jamia Millia Islamia (Central University)

Jamia Nagar, New Delhi-110025

<https://www.jmi.ac.in/computerscience>

