TWO-YEAR FULL-TIME M. Sc. PROGRAMME

In

BIOINFORMATICS

(Choice Based Credit System)

Department of Biotechnology

FACULTY OF SCIENCE

DEEN DAYAL UPADHYAYA GORAKHPUR UNIVERSITY, GORAKHPUR – 273009

2020

MASTER OF SCIENCE (Bioinformatics)
TWO-YEAR FULL-TIME PROGRAMME

Paper Code		Paper Name	Credit
Semester – 1 Odd Semester	**Core Course**		
	BI-501	Molecular Biology	2+1
	BI-502	IT tools and applications	2+1
	BI-503	Basic Bioinformatics	2+1
	Optional Course (Only One)		
	BI-510	Problem Solving Through C	2+1
	BI-511	Structural Biology	3
	Minor Course (Only one)		
	BI-520	Cell Biology	2+1
	BI-521	Biochemistry	2+1
	Ability Enhancement Compulsory Courses (AECC)AECC1		
	BI-530	Environmental Sustainability/Swachh Bharat Abhiyan activities/Biodiversity and its conservation	3
	Total		**18**
Semester – 2 Even Semester	**Core Course**		
	BI-504	Introduction JAVA Programming	2+1
	BI-505	Introduction To Genomics and Proteomics	2+1
	Optional Course (only one)		
	BI-512	Probability and Information theory	3
	BI-513	Data structure and algorithms	3
	Minor Course (only one)		
	BI-522	Biomathematics and Biostatistics	2+1
	BI-523	Introduction R Programming	3
	Ability Enhancement Courses (AEC): Skill enhancement course		
	BI-540	Mushroom Culture/Bio-fertilizer production/ Environmental Law/Tourism and Hospitality management/Life Skills and skill development/Yoga studies	**3**
	Total		**15**
Semester-3 Odd Semester	**Core Course**		
	BI-506	Database Management	2+1
	BI-507	Perl programming for bioinformatics	2+1
	BI-508	Optimization, machine learning and artificial intelligence	2+1
	Optional Course (only one)		
	BI-514	Complex Algorithms in Bioinformatics	2+1
	BI-515	PYTHON programming for Bioinformatics	3
	Minor Course (only one)		
	BI-524	Statistical Methods In Bioinformatics	3
	BI-525	Bio-safety and Scientific Communications	3
	Ability Enhancement Compulsory Courses (AECC)AECC2		
	BI-531	Human Values and professional Ethics/Gender Sensitization	3
	Total		**18**
Se me st	**Core Course**		
	BI-509	Molecular modeling and drug discovery	2+1

	Minor Course (Only one)		
	BI-526	Human Genetics and Genome Project	3
	BI-527	NGS data Analysis	2+1
	Compulsory Course		
	BI-550	Course Seminar	0+1
	BI-560	Thesis/Dissertation	15
		Total	**22**

Not included in CGPA

BI-550	Course Seminar	0+1
BI-560	Thesis/Dessertation	15

Semester-1

BI-501: MOLECULAR BIOLOGY

THEORY (Core)

Unit 1:

Prokaryotic and eukaryotic genome organization, structural elements of chromosome and construction of artificial chromosome. DNA replication: Enzymes, accessory proteins and mechanisms of prokaryotic and eukaryotic DNA replication.

Unit 2:

Fine structure of gene, molecular basis of spontaneous and induced mutations and their role in evolution; DNA damage and repair, DNA amplification and rearrangement. Anti-sense and Ribozyme Technology: Inhibition of splicing, polyadenylation and translation, disruption of RNA structure and capping, Biochemistry of ribozyme, hammerhead, hairpin and other ribozymes, applications of anti-sense and ribozyme technologies.

Unit 3:

Transcription: Organization of transcriptional units, mechanisms of transcription and its regulation in prokaryotes and eukaryotes, Operon concept, attenuation and antitermination controls, RNA processing (capping, polyadenylation, splicing), DNA methylation, heterochromatization, General and specific transcription factors, regulatory elements and mechanism of transcription regulation, transcriptional and post-transcriptional gene silencing, environmental regulation of gene expression.

Unit-4:

Translation: Genetic code, Prokaryotic and Eukaryotic translation, mechanisms for initiation, elongation and termination, regulation of translation, co-and post- translational modifications of proteins. Homologous Recombination and Site-specific recombination.

Books Recommended:

- Molecular Biology of the Gene (4th Edition) *J.D. Watson, N.H. Hopkins, J.W. Roberts, J.A. Steitz and A.M Weiner*, The Benjamin/ Cummings Publ. Co. Inc, California
- Molecular Biology of the cell (3rd Edition) by *Bruce Alberts, Dennis Bray, Julian Lewis, martin Raff, Keith Roberts and James D. Watson,* Garland Publishing, Inc, New York & London
- Gene Cloning and DNA Analysis (4th Edition) by *T.A Brown*, Blackwell Science
- Essential of Molecular Biology (3rd edition) by *G.M. Malacinski& D. Freifelder,* Jones & Bartlett Publisher

BI-502: IT TOOLS AND APPLICATIONS

Unit 1:

Computer Appreciation: Characteristics of Computers, Input, Output, Storage units, CPU, Computer System, Binary Number System Binary to Decimal Conversion, Decimal to Binary Conversion, Binary Coded Decimal (BCD) Code, ASCII Code.**Computer Organisation:** Central Processing Unit, Memory: Main Memory, Secondary Storage Devices. Input Devices, Output Devices, Multimedia, Computer Software**:** Relationship between Hardware and Software, system software, application software, compiler, names of some high level languages, free domain software.

Unit 2:

Operating Systems: Disk Operating System: Simple DOS Commands, Simple File Operations, Directory related Commands. **Microsoft Windows:** An overview of different versions of Windows, Basic Windows elements, Files management through Windows. Using essential accessories: Systems tools – Disk cleanup, Disk defragmenter, Notepad, Paint, wordpad. 1.3. **Linux:** An overview of Linux , Basic Linux elements: System Features, Software Features. File Structure, File handling in Linux, Installation of Linux: H/W, S/W requirements, Preliminary steps before installation, specifics on Hard drive repartitioning and booting a Linux System.

Unit 3:

Word Processing (Microsoft Word): Word Processing concepts: Saving, Closing, Opening an existing document, Selecting text, Editing text, Finding and replacing text, printing documents.

Spreedsheet Package (Microsoft Excel): Spreadsheet Concepts. Creating, Saving and Editing a workbook, Inserting, Deleting work Sheets, entering data in a cell/formula.

Unit 4:

Presentation Package (Microsoft Powerpoint): Creating, Opening and Saving Presentations, Creating the Look of Your Presentation, Working in Different Views, Working with Slides, Adding and Formatting Text.

Information Technology and Society: Application of Information Technology in Railways, Airlines, Banking, Insurance, Inventory Control, Financial Systems, Hotel Management, Education, Video Games, Telephone exchanges, Mobile Phones, Information Kiosks, Special effects in Movies.

Books Recommended:

- Computer Fundamentals, First Edition (6th Edition) *P. K. Sinha and P. Sinha*, BPB Publication.
- IT Tools and Applications, *Ramesh Bangaria*, Laxmi Publications.
- Microsoft Office 2010 for Windows: Visual QuickStart Guide *Steven A. Schwartz* Pearson Education India

BI-503: BASIC BIOINFORMATICS

Unit 1:

Introduction to Genes and Proteins: Genome Sequences ORFs, Genes, Introns, Exons, Splice Variants DNA/RNA Secondary Structure Triplet Coding Protein Sequences Protein Structure: Secondary, Tertiary, Quaternary The notion of Homology.

Unit 2:

Introduction to Internet Use, Search Engines, and Sequence Information Sources**:** WWW, HTML, URLs Browsers: Netscape / Opera / Explorer Search Engines Google, PUBMED. EMBL GENBANK Entrez Unigene Understanding the structure of each source and using it on the web. PDB SwissProt TrEMBL Understanding the structure of each source and using it on the web.

Unit 3:

Introduction to Data Generating Techniques: Restriction Enzymes, Gel Electrophoresis, Chromatograms Blots, PCR, Microarrays, Mass Spectrometry What data each generates, and what bioinformatic problems they pose.

Unit 4:

Sequence and Phylogeny Analysis: Detecting Open Reading Frames, Outline of Sequence Assembly, Mutation Matrices, Pairwise Alignments, Introduction to BLAST, using it on the web, interpreting results, Multiple Sequence Alignment, Phylogenetic Analysis.

Books Recommended:

- Bioinformatics: A practical guide to the analysis of genes and proteins (3rd Edition) D. Baxevanis and F. Oulette, Wiley Indian Edition.

- Bioinformatics Methods And Applications: Genomics Proteomics And Drug Discovery (3rd Edition) *Rastogi SC, Rastogi P, Mendiratta N.*. PHI Learning Pvt. Ltd.; 2008.
- Mount, D.W., Bioinformatics: 2001, Sequence and Genome Analysis. CSHL Press.
- Claverie, J.M. and Notredame C. 2003 Bioinformatics for Dummies. Wiley Editor.

BI-510: PROBLEM SOLVING THROUGH 'C' LANGUAGE

Unit 1:

Introduction to Programming

The Basic Model of Computation, Algorithms, Flow-Charts, Programming Languages, Compilation, Linking and Loading, Testing and Debugging, Documentation.

Introduction to 'C' language: Character set, Variable and Identifiers, Built-in-Data Types, Variable Definition, Arithmetic operators and Expressions, Constants and Literals, Simple assignment statement, Basic input/output statement, Simple 'C' programs.

Unit 2:

Conditional Statements and Loops: Decision making within a program, Conditions, Relational Operators, Logical Connectives, if statement, if-else statement, Loops: while loop, do while, for loop, nested loops, infinite loops, switch statement, structured programming.

Arrays: One dimensional arrays: Array manipulation; Searching, Insertion, Deletion of an element from an array; Finding the largest /smallest element in an array; Two dimensional arrays, Addition/ Multiplication of two matrices, Transpose of a square matrix; Null terminated strings as array of characters, Representation sparse matrices

Unit 3:

Functions: Top down approach of problem solving, Modular programming and functions, Standard Library of C functions, Prototype of a function: Formal parameter list, Return Type, Function call, Block structure, Passing arguments to a function: call by reference, call by value, Recursive Functions, arrays as function arguments.

Structures: Structure variables, initialization, structure assignment, nested structure, structures and functions, structures and arrays: arrays of structures, structures containing arrays, unions.

Pointers: Address operators, pointer type declaration, pointer assignment, pointer initialization, pointer arithmetic, functions and pointers, Arrays and Pointers, pointer arrays.

Unit 4:

Self Referential Structures and Linked Lists: Creation of a singly connected linked list, traversing a link list, Insertion into a linked list, Deletion from a linked list.

File Processing: Concepts of Files, File opening in various modes and closing of a file, Reading from a file, Writing onto a file

Books recommended:

- "The C Programming Language" 2nd by Brian W Kernighan / Dennis M Ritchie, AT & T Bell Laboratories, Murrey Hill, Prentice Hall, Eaglewood Cliffs, New Jersey. 07632.
- " ANSI C Programming" by Yashavant Kanetkar.
- " Programming in C" by Reema Thareja.
- " Computer Basics and C Programming" by Rajaraman V.

BI-511: STRUCTURAL BIOLOGY

Unit 1

System 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.

Unit 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.

Unit 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.

Unit 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

Books recommended:

- C, Branden & C. Tooze, “Introduction to Protein Structure”, Garland Publishing Inc., New York, 1991.
- T.E. Creighton, “Proteins – Structures and Molecular Properties”, W.H. Freeman & Co., New York, 1993.
- EE Conn & PK Stumpf, “Outlines of Biochemistry”, Wiley Eastern.
- R.R. Sinden, “DNA Structure and Function”, Academic Press, San Diego, 1994.
- Bioinformatics Sequence, Structure and Databanks Edited by D. Hidggins & W. Taylor, Oxford University Press, 2000.

BI-520: CELL BIOLOGY

Unit 1:
Structure of prokaryotic and eukaryotic cells,Cellular organelles: Plasma membrane, cell wall, cytoskeleton- their structural organization; Mitochondria; Chloroplast; Nucleus and other organelles and their organization and function, genetic constitution of mitochondria and chloroplast, artificial membrane Liposomes.

Unit 2:
Microscopic techniques: Principles and application of light, phase contrast, fluorescence, confocal, scanning and transmission electron microscopy, cytophotometry and flow cytometry, fixation and staining, Fluorescence in-situ hybridization (FISH), GISH (Genomic in-situ hybridization).

Unit 3:
Transport of nutrients, ions and macromolecules across membranes, Cell cycle: Mitosis, meiosis, role of cyclins and cyclin dependent kinases, regulation of Cdk-cyclin activity, Cdk inhibitors, induction of cancer with respect to cell cycle, molecular events and regulation in model systems, cell surface receptors, second messenger system, MAP kinase pathways, mechanism of signal transduction pathway.

Unit 4:
Molecular biology and biochemistry of cancer, oncogenes, tumor suppressor genes, chemical carcinogenesis, Cellular basis of differentiation and development- cell division, gametogenesis and fertilization, differential gene activity and cell differentiation, Morphogenetic determinants in egg cytoplasm, genetic regulation of early embryonic development in Drosophila, homeotic genes

- Lewin's CELLS, 3rd Edi. (2013) by George Plopper, David Sharp, Eric Sikorski
- Molecular Biology of the Cell, 6th Edi., (2014) Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, and Peter Walter.

- The Cell: Amolecular approach, 7th Edi., by Geoffrey M. Cooper, Robert E. Hausman
- Essential Cell Biology, 4th Edi., (2019) by Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander D.Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.

BI-521: BIOCHEMISTRY

Unit1

Introduction: Some important properties of water, dissociation of water and its ion product, Kw, buffer and buffering capacity pH, Bronsted acids, ionization of weak acids and bases, Henderson Hasselbalch equation, Titration curves, buffering action, Goods buffer, water & intermolecular forces, thermodynamics in biochemistry, law of thermodynamics, Gibbs free energy & biomolecules in water.

Unit 2

Amino Acids: Structure, properties, classification, function, ionization and characterization, naturally occurring modifications of amino acids in proteins, non-protein amino acids, structure of proteins: primary, Secondary (α-helix, β-Plated and random coils), peptide bond, Ramachandran plots and collagen structure, protein sequencing, protease mapping, characterization of peptides, tertiary and quaternary structures of proteins/enzymes.

Unit 3

Carbohydrates: introduction, classification, types and nomenclature of sugars. Optical isomerism, mutarotation, basic structure and functions of monosaccharides, oligosaccharides, polysaccharides, energy storage molecules-starch, glycogen, building blocks-cellulose, hemicellulose, and chitins. Carbohydrate metabolism, brief overview of glycolysis and Krebs cycle.

Unit 4

Nucleic acids: nucleosides and nucleotides, primary structure of nucleic acids, structure, properties and functions of DNA and RNA, secondary and tertiary level organization, different DNA forms, conformation, super coiling, stereochemistry: nucleoside, torsion angles, sugar conformation, DNA structure: different types of DNA and their structure, DNA motifs, DNA repeats and their significance, function and stability, spectroscopic study of DNA: dye binding, interaction, denaturation, and renaturation of DNA, thermal denaturation and Tm values.

Unit 5

Lipids: classification, structure, properties and function of fatty acids, triglycerides, phospholipids, glycolipids, sphingolipids, sterols, cerebrosides, steroids, prostaglandins, glycolipids and proteoglycans. Vitamins and their biological significance, Biologically important small molecules. Structure of nucleosomes and ribosomes.

Recommended books:

- Lehninger Principles of Biochemistry (2017), Seventh Edition by L. Nelson; Michael M. Cox, C.B.S. Publication
- Biochemistry by Lubert Stryer (2019), 9th edition W. H. Freeman publisher.

- Fundamentals of Biochemistry: Life at the Molecular Level, (2016) 5th Edition Donald Voet, Judith G. Voet, Charlotte W. Pratt
- Biochemistry, Vol I, II, III by GeofferyZubey, WCB press

Semester 2

BI-504: INTRODUCTION JAVA PROGRAMMING

Unit 1:
Overview of JAVA : The genesis of java, An overview of java, java virtual machine (JVM),Java development kit (JDK), Java Vs C++, Data types, Literals, Variables, and Arrays, Operators, Control statements, Introducing Class, closer look at Methods and class, Nested and inner class ,Exploring Java.lang, String handling ,Constructor, Garbage collection and finalize() method.

Unit 2:
Inheritance, Packages and interface- Types of inheritance, Access specifiers class inheritance, using super, method overriding, Abstract class, constructor in multilevel inheritance, using final with inheritance, Dynamic method dispatch, Defining package, CLASSPATH, Access protection, Importing package, Defining and implementing interface, Extending interface, Nested interface.

Unit 3:
Exception handling and Multithreading: Using try and catch, multiple catch classes, Nested try statements, throw, throws and finally, Built in exception, Uncaught exception, Creating own exception class, Java Thread Model – Main thread, Creating own Thread, Life cycle of thread, Thread priorities, Synchronization and messaging, Intertribal communication, Suspending, Resuming and stopping thread.

Unit 4:
Input Output and Networking : Byte stream and character stream ,Predefined stream, reading console input, writing consol output, PrintWriter class, Reading and writing files Networking – classes and interfaces, Socket and overview, TCP/IP client socket and server, Inet address, URL Connection, Datagram.

Applet, AWT and Event handling – Applet life cycle, Creating an applet, Using image and sound in applet, passing parameter. AWT- Overview of java.awt package, Component and Containers, control component and layout manager. Event handling –The delegation-event model, Event classes, Source of event, Event listener interfaces, handling mouse and keyboard event, Adapter class.

Books recommended

- Balaguruswamy E, 2009 Programming in JAVA 2, Tata Mcgraw-Hill, India.

- Russel J.P, 2001 JAVA Programming, Prentice Hall of India ,New Delhi.
- Patrick Naughton and Herbertz Schildt, The Complete Reference JAVA 2 Tata Mcgraw-Hill.
- Dietel nad Dietel, "Java: How to program Java 2", Second Edition, Pearson Education

BI-505: INTRODUCTION TO GENOMICS AND PROTEOMICS

Unit 1:

Advanced Biochemistry: An introduction to physical biochemistry, intermediary metabolism and molecular biology. Topics include a survey of structure, chemistry and function of proteins and nucleic acids; regulation of gene expression at the level of DNA, RNA, and protein synthesis.

Metabolism and Pathways: Pathways of carbohydrate, lipid and nitrogen metabolism and their metabolic control.

Unit 2:

Basic concepts in Genomics: Whole genome analysis, Genome sequencing technology. Comparative genomics – Paralogs and orthologs, Phylogeny, Human genetic disorders, Candidate gene identification, Concepts of Pharmacogenomics.

Micro array for gene expression: Target selection, customized microarray design, image processing and quantification, normalization and filtering, statistical analysis, public microarray data sources.

Unit 3:

Proteomics: Basics of Protein structure, Introduction to basic Proteomics technology, Bio-informatics in Proteomics, Basics of Proteome Analysis, Concepts in Enzyme Catalysis.

Unit 4:

Genome Project – The unfolding story:. Introduction to the concepts cloning and mapping, Construction of Physical maps, Basics of radiation hybrid maps, Sequencing : Related discoveries and technology development, Implications of the Human Genome Project, Basic Human Inheritance Patterns, Basics of Single Nucleotide Polymorphism detection and its implication, Practical Application of medical Genetics Technology.

Books recommended:

- Genomes. Author : T.A. Brown, John Wiley and Sons, Bios Scientific Publishers
- Discovering Genomics, Proteomics and Bioinformatics Campbell AM and Heyer LJ Perason Education (Low priced Editions).
- Jerome, P.E. 2002 Mathematics for Genome Analysis. Cambridge.

- Primrose and Twyman 2003 Principles of Genome Analysis & Genomics. Blackwell.

BI-512: PROBABILITY AND INFORMATION THEORY

Unit 1

Probability Theory: Introduction, Random Experiment, Sample Space, Events, Complementary Events, Union and Intersection of Two Events, Difference Events, Exhaustive Events, Mutually Exclusive Events, Equally Likely Events, Independent Events, Mathematical & Statistical definition of Probability, Axiomatic definition of probability, Addition Theorem, Multiplication Theorem, Theorems of Probability, Conditional Probability, Inverse Probability.

Unit 2

Probability Distributions& Random Variables: Introduction, Probability mass/density function (PMF/PDF), discrete distributions, continuous distributions, mean and variance of common distributions, Properties, Cumulative Distribution function (CDF), properties, relation with PDF, functions of random variables, Characteristic function, special random variables, Chebychev and Markov Inequalities, Central limit theorem.

Unit 3: Random Processes

Introduction, Random process, wide sense random process, strict sense and ergodic random process, Linear systems with random inputs.

Unit 4:

Information Theory: notion of information, concept of entropy, conditional and joint entropies, principle of maximum entropy, discrete memory less channel (DMC), Shannon's theorems and their applications, Variable length code, Huffman code, Shannon-Fano code.

BI-513: DATA STRUCTURES AND ALGORITHMS

Unit 1:

Introduction & computational complexity: Abstract data type and data structures, Classes and objects, Complexity of algorithms: worst case, average case and amortized complexity, notation for algorithm complexity, Algorithm analysis, Recurrence relations, Introduction to NP-completeness.

Unit 2:

Essential preliminaries: Overview of recursion, iteration, arrays, pointers, lists, stacks, queues. Dictionaries: Hash tables, Binary search trees, splay trees, Balanced Trees, AVL trees, 2-3 trees, B-Trees. Priority Queues : Heaps, binomial queues, Applications.

Unit 3:
Graphs: Shortest path algorithms (Dijkstra, Bellman-Ford, Floyd-Warshall), minimal spanning tree algorithms (Prim, Kruskall), depth-first, breadth-first search, and their applications.
Sorting: sorting methods and their analysis (shell sort, quicksort, merge sort, heap sort, radix sort), lower bound on complexity

Unit 4:
String matching: String matching algorithms (Rabin-Karp, Knuth-Morris-Pratt).

Algorithm design paradigms: Algorithm Design Paradigms: Greedy methods, divide and conquer, dynamic programming, backtracking, local search methods, branch and bound technique. Travelling salesman problem.

Books recommended:

- Lipshutz and Pai, 2010 Data Structure (Outline Schaum Series), McGraw Hill, Indian edition.
- Standish T. A., 1980 Data Structure Techniques, Addison-Wesley.
- Salzberg, B., 1988, File Structures, Prentice-Hall.
- Tharp, A.L., 1988, File Organization and Processing, John Wiley and Sons.
- Aho, A., Hopcroft, J. and Ullman, J., The Design and Analysis of Computer Algorithms, Addison Wesley.
- Goodman S.E. and Hedetniemi S.T., Introduction to the Design and Analysis of Algorithms. McGraw Hill.

BI-522: BIOMATHEMATICS AND BIOSTATISITCS

Unit I
Sets, Types of Sets, Subsets, Complement of Sets, union and Intersection of Sets, Difference of Sets, Demorgan's Law, Cartesian product of Sets. Basics of Probability, Permutation and Combination.

Unit II
Measure of central tendency and dispersion: Mean, median, mode, range, standard deviation, variance

Unit III
Correlation and Regression: Types, Karl-Pearson's correlation, Spearman's Rank correlation, Regression equation and fitting

Unit IV

Probability Distribution: Basics of Binomial, Poisson and Normal distributions and their application in biology. Random Variable; Discrete and Continuous Probability Distribution, Probability mass function, probability Density function, Mathematical Expectation.

Unit V

Matrices, Types of Matrices, Addition of matrices, Subtraction of matrices and Product of matrices. Properties of Matrix Multiplication. Transpose of Matrix, Symmetric and Skew-symmetric Matrices, Inverse of Matrix.

Books recommended:

- Shanti Narayan, A text book of Vector Calculus, S Chand & company, New Delhi
- H. Nell and D. Quadling, 'Pure Mathematics (Advanced Level Mathematics)'. Vol. 1,2,3, Cambridge University Press 2002.
- Edward Batschelet. 'Introduction to Mathematics for Life Scientists', 3rd Edition, Springer – Verlag, 1992
- J. Crawshaw and J. Chambers', 'Advanced level Statistics', 4th Edition, Nelson Thornes, 2002.

BI-523: INTRODUCTION R PROGRAMMING

Unit 1

Introduction: Introducing to R, R Data Structures, Help functions in R, Vectors, Scalars, Declarations, recycling, Common Vector operations, Using all and any, Vectorized operations, NA and NULL values, Filtering, Vectorised if-then else, Vector Equality, Vector Element names.

Matrices, Arrays And Lists: Creating matrices, Matrix operations, Applying Functions to Matrix Rows and Columns, Adding and deleting rows and columns, Vector/Matrix Distinction, Avoiding Dimension Reduction, Higher Dimensional arrays, lists, Creating lists, General list operations, Accessing list components and values, applying functions to lists, recursive lists.

Unit 2

Data Frames: Creating Data Frames, Matrix-like operations in frames, Merging Data Frames, Applying functions to Data frames, Factors and Tables, factors and levels, Common functions used with factors, Working with tables, Other factors and table related functions - Control statements, Arithmetic and Boolean operators and values, Default values for arguments - Returning Boolean values, functions are objects, Environment and Scope issues, Writing Upstairs - Recursion, Replacement functions, Tools for composing function code, Math and Simulations in R.

Unit 3

OOP: S3 Classes, S4 Classes, Managing your objects, Input/Output, accessing keyboard and monitor, reading and writing files, accessing the internet, String Manipulation, Graphics, Creating Graphs, Customizing Graphs, Saving graphs to files, Creating three-dimensional plots

Unit 4

Interfacing: Interfacing R to other languages, Parallel R, Basic Statistics, Linear Model, Generalized Linear models, Non-linear models, Time Series and Auto-correlation, Clustering

Aplication of R programing in Bioinformatics: Case studies.

Recommended books:

- A First Course in Statistical Programming with **R** by Braun & Murdoch.
- A Beginner´s Guide to **R**" by Zuur.
- Introduction to Scientific Programming and Simulation Using **R** by Chapman & Hall/CRC.
- **R** in a Nutshell by Adler.
- An Introduction to **R** by Venables & Smith

Semester-3

BI-506: DATABASE MANAGEMENT
THEORY
(Core)

Unit 1:

Introduction to Databases: Why Database systems, Data abstraction and data models, Instances and schemas, Database Administrator, Data Definition and manipulation languages, brief introduction to network and hierarchical models.

Entitles Relationship model: Entity and entity sets, relationships and relationship sets,, E-R diagram, reducing E-R diagrams to Tables and trees.

Relational Algebra and Calculus: Relational algebraic operations such as select, project, union, set difference, Cartesian product,, intersections, natural join, division, generalized projection, outer join etc., tuple relational calculus, domain relational calculus.

Unit 2:

Issues in designing relational databases: Pitfalls in relational database design, decomposition, importance of normalization, functional dependencies, Boyce-Code Normal form, third normal form and fourth normal form.

Query language and query optimization: Domain types in SQL, Schema definition in SQL, Types of SQL commands, SQL operators, tables, views, indexes, aggregate functions, insert, delete and update operations, join, union, intersection, minus etc. in SQL, queries, sub-queries, equivalence of queries

Unit 3:
Database system architecture: Introduction to centralized system, client server system, parallel system and distributed system.

Introduction to ASN.1 and NCBI data model: Why specialized data model is required for biological sequences, different data types supported by ASN.1and how they are used for storage of different types of information reading of NCBI data using freely available NCBI toolbox.

Books recommended:

- Date, C.J. Introduction to Database Systems (Vol I & II), 2004, 8th Edition. AddisonWesley.
- Ullman, J.D. 1989, Principles of Database and knowledge base Systems (Vol I & II), Computer Science Press New York.
- Gio Wiederhold, 1997 Database Design, McGraw Hill.
- Elmasri R. and Navathe S.B., 2007 Fundamentals of Database Systems. Fifth Edition.Pearson.
- Singh S.K., 2011 Database Systems- Concepts, Designs and Application. 2nd Edition. Pearson
- Silberschatz A. Korth H. F. Sudarshan S., 2010 Database System Concepts. Sixth Edition. McGraw-Hill. Date K., Swamynathan S. 2012 An Introduction to Database Systems. Eight Edition. Pearson.

BI-507: PERL PROGRAMMING FOR BIOINFORMATICS

Unit 1
Introduction: Introduction to Perl, Downloading and installation from Website, Writing and Running a Perl Program, Editing, Advantages.

Data Types: Scalar data and scalar variables: Number, String, Conversion between Numbers and Strings, Variable Interpolation, Arithmetic and Decimal Precision, Arrays: Initialization, Manipulation of Array elements; Associative Array (Hashes): Initialization, Manipulation of Elements of Array.

Arithmetic and Logical Operators: Arithmetic Operators, Assignment Operators, Increment and Decrement Operators, String Concatenation and Repetition, Operators precedence and

Associativity, Conditional Operators, Logical Operators, Operators for manipulating arrays, Operators for Manipulating hashes.

Unit 2

Conditionals and Loops: Conditional Statement; if, if…else, if and if-else, unless statement, Loops: while, for, until, do..while, do..until and foreach loop, last next, redo, continue and case switch statement.

Input and Output: Creating a file, Reading Data from a file, Writing data to a file, Closing a file, Managing Files and Directories.

Regular Expressions and Pattern Matching: Regular Expression, Pattern Matching, Meta Character, Simple Pattern, Matching Group of Characters, Matching multiple instances of Characters, Pattern Building, Pattern and Variable, Pattern and Loops, Using Pattern for Search and Replace, Matching Pattern over multiple Lines etc.

Unit 3

Function and Subroutines: Built-in Functions, Defining and calling subroutines, Returning Values from Subroutines, Using Local Variables in Subroutines, Passing Values into Subroutine, Perl References, Perl module and their uses.

Unit 4

Applications of Perl in Bioinformatics: Concatenating DNA Fragments, Transcription: DNA to RNA, Reading Protein Files, Finding Motifs, Simulating DNA, Generating Random DNA, Analysing DNA, Translating DNA to Proteins, Reading DNA from Files in FASTA format, Separating Sequence and Annotation, Parsing Annotation, Parsing PDB files, Parsing BLAST output, Bio-perl
.

Books recommended:

- James Tisdall, “Beginning Perl for Bioinformatics”, O’Reilly & Associates.
- James Tisdall, “Mastering Perl for Bioinformatics”, O’Reilly.
- Cynthia Gibas & Per Jambeck, “Developing Bioinformatics Computer Skills”, O’Reilly & Associates,
- Rex A. Dawyer, “Genomic Perl”, Cambridge University Press
- Learning Perl, 3rd Edition , Author: Randal L. Schwartz and Tom Phoenix, O’Reilly

BI-508: OPTIMIZATION, MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE

Unit 1:

Optimization techniques: Concepts of local and global optimum, constrained and unconstrained optimization, gradient, Hessian of a function, first order and second order necessary conditions for local minimizers, method of steepest descent, steepest descent for quadratic function, Newton's method, LevenbergMarquardt algorithm, and Lagrange multipliers methods.

Unit 2:

Genetic algorithms: Introduction to Genetic algorithms, exploitation and exploration, general structure of the basic genetic algorithm, different coding schemes, fitness function, chromosome, population, basic genetic operators, different selection mechanism, mutation, single point and multi-point crossover, termination criteria, elitist strategy.

Basic concepts of machine learning: Basic notions of learning,, introduction to learning algorithms, incremental learning, supervised learning, unsupervised learning, reinforcement learning, instance based learning and analytical learning.

Unit 3:

Clustering and classification: Introduction to clustering, maximum likelihood decomposition and its application to normal mixtures, k-means clustering algorithms, hierarchical clustering algorithms. Introduction to classifier design, linear discriminant analysis – two category and multi category cases, perceptron criterion and its minimization, decision trees (ID3, C4.5), impurity functions, pruning methods, rule extraction from decision trees, nearest neighbour classifier, k-nearest neighbour classifier, Bayes decision rule, lose function, minimum error rate classification, Bayes classifier with multivariate normal density.

Introduction to neural networks: Introduction to neural networks, introduction to biological neural network, motivation for artificial neural network (ANN), significance of massive parallelism and characteristics of (ANN), various types of architectures.

Unit 4:

Feed forward neural network: Layered networks, perceptron and motivation for multilayered perceptron (MLP), back-propagation learning , use of Levenberg-Marquardt method in training of MLP's, online vas batch learning, issues relating to initialization, termination, choice of architecture. Radial basis function networks and related training issues, properties of MLP and RBF.

Self-organizing maps and Recurrent neural networks: Self-organizing feature map, recurrent neural networks, Hopfield network and its applications, simulated annealing.

Books recommended:

- R. O. Duda, P. E. Hart and D. G. Stork, Pattern classification, John Wiley Sons, Second Edition,
- D. E. Goldberg, Genetic algorithms in search, optimization and machine learning, Pearson Education (Paper back)

- Christopher M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press,
- E. K. P. Chong and S. H. Zak, An introduction to optimization, John Wiley & Sons
- Tom Mitchell, Machine Learning, McGraw Hill, 1997

BI-514: COMPLEX ALGORITHMS IN BIOINFORMATICS

Unit 1

TSP; Weight matrices: Sequence weighting, pseudo count correction for low counts, Gibbs sampling, and Psi-Blast

Unit 2

Dynamic programming: Needleman-Wunsch, Smith-Waterman, and alignment heuristics; Data redundancy and homology reduction: Hobohm and other clustering algorithms

Unit 3

Hidden Markov Models: Model construction, Viterbi decoding, and posterior decoding, and Baum Welsh HMM learning

Unit 4

Artificial neural networks: Architectures and sequence encoding, feed-forward algorithm, and back propagation; BCO; ACO; Genetic Algorithm

Recommended books:

- Bioinformatics: Problem Solving Paradigms, (2008),by Sperschneider, Volker, Springer International Publisher
- Bioinformatics Algorithms: Techniques and Applications, (2008) Ion Mandoiu, Alexander Zelikovsky, Wiley Science.
- Bioinformatics Algorithms, by Moguel Rocha, Pedro G. Ferreira, Academic Press Elsiever.

BI-515: PYTHON PROGRAMMING FOR BIOINFORMATICS

Unit 1

Introduction and Overview: A brief history of python – Unique features –Installation of Python and IDE - Lexical structure of python – Introduction of variables and data types with examples.

Unit 2

Statements and control structure: Introduction to python interpreter and interactive mode – Statement Read and Print commands – Evaluating expressions - Decision, Boolean Logic and Repetition structures syntax with examples in biological application.

Unit 3

Functions and Regular expressions: Defining and Calling a function - Fruitful functions (return value, parameters, local and global scope, function composition, recursion) – Examples in sequence analysis using function - Introduction to Modules. Regular Expression: Importance of patterns in biology – String manipulation using regular expressions (Extraction, splitting and matching).

UNIT 4

Tuples and Dictionaries: Introduction to Lists – List slicing – Finding items in Lists with operator – Copying and Processing Lists – List built-in methods – Two Dimensional lists. Tuples: Basic tuple operations – creation, concatenation, repetition, slicing, immutable and deletion. Dictionaries: creation, accessing and processing - Dictionary methods.

UNIT 5

Files and Exception Handling: File objects – File built-in methods and attributes - Reading and writing files - command line arguments. Exception Handling: Errors and exceptions, Detecting and Handling Exceptions.

Books recommended:

- Bioinformatics Programming Using Python (2009) by Mitchell L Model, O'Reilly Media, Inc.
- Python for Bioinformatics, by Bassi Sebastian, CRC Press, Taylor & Francis group
- Python Programming for Biology, (2015) by Tim J. Stevens, Wayne Boucher, Cambridge University Press.
- Bioinformatics Programming in Python: A Practical Course for Beginners (2008)

  Ruediger-Marcus Flaig, Wiley-Blackwell

BI-524: STATISTICAL METHODS IN BIOINFORMATICS

Unit 1:

Review: Mean, Median, Mode Standard Deviation, Variance and Correlation (Emphasis to be placed on hands on approach with real data sets) Probability limits and distribution theorems.

Sampling Distributions:Statistic Distribution of sample mean Sample variance – (application of Central limit theorem)

Estimation Theory: Biased and unbiased estimator Confidence interval: population mean, proportion Variance

Unit 2:
Maximum Likelihood Estimation: Discrete and Continuous distributions Likelihood function Log-likelihood functions (use of package recommended).

Inference – Test of hypotheses: Formulation of Hypothesis: Simple and Composite Type I and Type II errors Power of a test Significance of a test P-value, Testing, Chi-Square, t-test and F-test, Chi-Square goodness of fit, Test of diversity based on entropic method, Non-parametric: Mann-Whitney test.

Unit 3:
Simple Linear Regression and Correlation: Linear regression model Least squares methods Estimating model parameters Residual sum of squares.

Analysis of Variance: Single factor ANOVA Multi-comparison ANOVA

Unit 4:
Statistical Methods for Stochastic Processes: Testing of independence Testing of Markov property Association test.

Analysis of Stochastic process: Long Repeat Scan Statistics Analysis of Patterns.

Books recommended:

- Statistical methods in Bioinformatics, (2001), **Ewens**, Warren J., **Grant**, Gregory R., Springer International Publications.
- Statistical Methods in Bioinformatics, An Introduction, 2004, (second edition) Warren J. Ewens and Gregory R. Grant Springer, Heidelberg.
- Statistics For Bioinformatics (2016) by Julie Thompson , Elsevier Science
- Applied Bioinformatics: An Introduction (2008) by Selzer Paul Maria Et Al, Springer

BI-525: BIOSAFETY AND SCIENTIFIC COMMUNCIATIONS

Unit 1
Biosafety and risk assessment issues, regulatory framework, National biosafety policies and law, The Cartagena protocol on biosafety, WTO and other international agreements related to biosafety; Cross border movement of germplasm; Risk management issues-containment.

Unit 2
General principles for the laboratory and environmental biosafety; health aspects; toxicology, allergenicity, antibiotic resistance etc.Impact on environment; gene flow in natural and artificial

ecologies; Sources of gene escape, tolerance of target organisms, creation of superweeds/superviruses etc.

Unit 3

Ecological aspects of GMOs and impact on biodiversity; Monitoring strategies and methods for detecting transgenics; Radiation safety and non-isotopic procedures; Benefits of transgenics to human health, society and the environment.

Unit 4

The WTO and other international agreements; Intellectual properties , copyrights, trademarks, trade secret, patents, geographical indications, etc.; Protection of plant variety and farmers right act; Indian patent act and amendments, patent filing; Convention on biological diversity; Implications of intellectual property rights on the commercialization of biotechnology products.

Scientific Literature: Structure, Type (Research article, review article, Short communication, Book review). Abstracting methodology. Concept of Impact factor; Citation Index

Books Recommended

- Beier, F.K., Crespi, R.S. and Straus, T. Biotechnology and Patent protection-Oxford and IBH Publishing Co. New Delhi.
- Intellectual property rights and Bio-Technology (Biosafety and Bioethics), Anupam Singh, Ashwani Singh, NPH, New Delhi
- Sasson A, Biotechnologies and Development, UNESCO Publications.
- Singh K, Intellectual Property rights on Biotechnology, BCIL, New Delhi
- Regulatory Framework for GMOs in India (2006) Ministry of Environment and Forest, Government of India, New Delhi
- Cartagena Protocol on Biosafety (2006) Ministry of Environment and Forest, Government of India, New Delhi

Books recommended:

- W. J. Ewens and G. R. Grant, “Statistical Methods in Bioinformatics”: An introduction Springer-
- Verlag 2001
- J. L. Devore, “Probability and Statistics”, (Fifth Edition), Thomson Asia.
- S. M. Ross, “A First course to Probability”, 6th Edition, Pearson Education

Semester-4

BI-509: MOLECULAR MODELING AND DRUG DISCOVERY

Unit 1:

Introduction to drug designing: Different approaches to drug designing, Basic principle of similarity and complimentarity, High throughput vs rational drug designing, Use of computer modelling technique to drug designing.

Small molecular structures: Different coordinate systems and transformations amongst them, Basic Principle 2D and 3D Graphics and use of molecular graphics packages (e.g. RasMol, RasTop, Qmol, MolMol), Cambridge Structural Data base, Building small molecules using chemical information, Use of Builders and Sketchers (eq. ISIS Draw, HyperChem).

Unit 2:

Basic principles of target structure based (rational) drug designing: Target identification and validation, Active site analysis, Basic principle of Docking a ligand in active site of a target, Virtual Screeing, Different methods of scoring and Lead optimization.

Unit 3:

Basic concepts in quantitative structure activity relationship (QSAR): Objective of QSAR, Development of Hansch QSAR equation, QSAR Descriptors, Regression analysis.

Pharmacophore model building: Ligand-based pharmacophore modeling, Structure-based pharmacophore modeling.

Unit 4:

Forces stabilizing Protein & Nucleic Acid structures**,** Energy minimization and Molecular Dynamics methods**:** Theory, Application to Proteins, Applications to nucleic acids.

Books recommended:

- Molecular Modeling: Basic principles and applications. Holtje HD, sippl W, Rognan D and Folkers G. Wiley-VFH 2nd Edition.
- Quantum Biology: S. P. Gupta, New Age publishers
- olecular modelling and drug design. Andrew Vinter and Mark Gardner and Boca Raton, CRC Press.
- Molecular Similarity in drug design, Dean PM, Chapman and Hall.

BI-526: HUMAN GENETICS AND HUMAN GENOME PROJECT

Unit 1

Genetic mapping of Mendelian characters: Recombinants, Non-recombinants, Genetic markers, Two point mapping, Multipoint mapping, Fine mapping using extended pedigrees and ancestral haplotypes. Identifying Human disease genes: Principles and strategies in identifying disease

genes, Positional cloning, Use of chromosomal abnormalities, confirming a candidate gene, various ways of identifying disease genes.

Unit 3

Mapping and identifying genes conferring susceptibility to complex diseases: Deciding whether a non-Mendelian character is genetic: the role of family, twin and adoption studies, Linkage analysis of complex characters, Association studies and linkage disequilibrium, Identifying the susceptibility alleles, Examples that illustrate the varying success of genetic dissection of complex diseases.

Unit 4

Molecular Pathology: Rules for nomenclature of mutations & databases of mutations, Loss of function mutations, Gain of function mutations, Molecular pathology from gene to disease, Molecular pathology from disease to gene, Molecular pathology of chromosomal disorders.

Unit 5

Human Genome Project (HGP) – an overview of the project, goals of the project, major scientific strategies & approaches used in HGP, expected scientific & medical benefits of this project, about the organizations behind this project. How Human genome was mapped – physical mapping, genetic mapping, gene ontology, gene annotation.

Recommended books:

- The Human Genome, 3rd Edition,(2010) by Julia Richards R. Scott Hawley, Academic Press
- Human Genes and Genomes, (2013), by Leon Rosenberg, Dione Rosenber, Academic Press.
- Human Genome Project, (2013) by Kelly E. Happe, New York University Press.
- Genomics: The Science and Technology behind the Human Genome Project (1999), by Charles R. Cantor, Cassandra L. Smith, John Wiley & Sons, Inc.

BI-527: NGS DATA ANALYSIS

Unit 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

Unit 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.

Unit 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.

Unit 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.

Unit 6:
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 ChIPSeq Data Analysis.

Books recommended

- Next-Generation Sequencing Data Analysis by Xinkun Wang, (2016) CRC Press, Inc. Subs. of Times Mirror 2000 Corporate Blvd. NW Boca Raton, FL United States.

- Statistical Analysis of Next Generation Sequencing Data (2014) Editors: Datta, Somnath, Nettleton, Dan (Eds.), Springer International Publications.
- Computational Methods for Next Generation Sequencing Data Analysis, Editor(s): Ion Măndoiu , Alexander Zelikovsky (2016), John Wiley & Sons.
- Next-generation sequencing data analysis (2016), Xinkun Wang, CRC Press, Boca Raton, Florida, Oxford publications.
- Big Data Analytics in Genomics, **Wong**, Ka-Chun (Ed.), Springer International Publications

BI-550: SEMINAR (Core)

BI-560: DISSERTATION (Core)

Each Student will have to submit an allotted Dissertation, which would be based on research works and will submit a report on which Viva-Voce will be conducted.