MSc Computational Biology
Programme Structure

Course Name	Course code	Credit	Course Name	Course code	Credit
Semester 1			**Semester 3**		
Introduction to Life Sciences	COB 511	4	Computer Aided Drug Design	COB 531	4
Applied Mathematics & Scientific Computing	COB 512	3	Soft Computing Methods	COB 532	4
Fundamentals of Molecular Biology	COB 513	4	DSP in Bioinformatics	COB 533	4
Perl & Bio-Perl	COB 514	4	Mini Project	COB 534	4
Free and Open Source Software (E)	COB 501	3	Gene Annotation & Ontology (E)	COB 503	3
Total Credits (including one elective)		**18**	**Total Credits (including one elective)**		**19**
Semester 2			**Semester 4**		
Computational Genomics & Transcriptomics	COB 521	4	Project & Viva Voce	COB 541	6
Computational Proteomics & Metabolomics	COB 522	4	Seminar	COB 542	2
Computational Phylogenetics	COB 523	4	Systems Biology	COB543	4
Data Warehousing & Data Mining	COB 524	4	Computer Networks & Web Programming (E)	COB 504	4
Professional Studies (E)	COB 502	3	**Total Credits (including one elective)**		**16**
Total Credits (including one elective)		**19**			
Grand Credits		**72**			

Credits for cores in each semester: **16 – 19** & for electives: **3**

Electives

Free and Open Source Software	COB 501	3	Data Structure for Bioinformatics	COB 506	3
Professional Studies	COB 502	3	Grid Computing	COB 507	3
Gene Annotation & Ontology	COB 503	3	Micro-array & Image Processing Techniques	COB 508	3
Computer Networks & Web Programming	COB 504	4	Programming in R	COB 509	3
Computational Biology Algorithms	COB 505	3			

COB 511 Introduction to Life Sciences

Module I

Nature and scope of life science, Branches of life sciences, characteristics of life, Levels of Organization, Origin of life, Biochemical evolution-evolution of Proteins and Nucleotide, Genetic code. The cell as basic unit of life-Prokaryotic cell and Eukaryotic cell

Module II

Introduction to Molecular Biology: Macromolecules, Brief study of carbohydrates, proteins, Lipids, Nucleic acids, Amino acids, Genetic Material: DNA Replication, DNA Repair, Transcription, Translation, Mutation and Mutagenesis. Central dogma: genetic code, gene and operon, Structure of DNA and RNA.

Module III

Cell Membrane, Cell Organelles, Cell Structure and Function, Cell Energetics: Photosynthesis, Chemical reaction and energy, Metabolic pathways: Kreb's Cycle, Electron transport chain Glycolysis, Enzymes, ATP Structure and Formation. Cell Division: Mitosis, Meiosis, Chromosome, Genes and DNA.

Module IV

Introduction to Genetics: Basic concepts and Terminologies, Mendel's Experiment, Monohybrid and Dihybrid genetics, Test Crosses, Multiple Alleles, Polygenic Inheritance.

Module V

Introduction to Computational Biology: The digital language of computers, Binary Computing and DNA, ASCII representation of DNA, Algorithms in Computational Biology: Definition of Algorithms, Consensus, Matrices, Search algorithm in Bioinformatics, Global alignment (Needleman - Wunch), Local alignment (Smith-waterman) Optimal versus Heuristic Algorithms, BLAST, NCBI Resources and other Databases.

Reference

1. Campbell Reece, "Biology".
2. Eldon D. Enger, Frederick C. Ross, "Concepts in Biology" Tata McGraw-Hill Publishing Company.
3. Becker, Kleinsmith, "The World of the cell".
4. Gerald Karp, "Cell and Molecular Biology".
5. Benjamin Lewin, "Genes IX" Jones and Bartlett Publishers.
6. Wunschiers, "Computational Biology".
7. Gary Page, "Algorithms in Bioinformatics".
8. Fall, Maryland, " Computational Cell Biology".
9. Zvelebil & Baum, " Understanding Bioinformatics".
10. Stephen Misena & Stephen A. Krawetz, "Bioinformatics Methods and Protocols".

COB 512 Applied Mathematics & Scientific Computing

Module I

Introduction to Scilab/ MATLAB**:** Scilab/ MATLAB environment, Numbers, Variables, Special Constants, Operators, Built-in Functions, Concatenation, Command Line Editing, Matrices, Generating Matrices, Matrix Operations- addition, subtraction multiplication, random matrix, Transpose, Inverse

Module II

More Scientific Computing in Scilab/Matlab: Defining Polynomials, Roots of polynomials, degree, coefficients, eigen value, Derivatives and Integrals, Statistics: Statistical functions- Mean, Standard deviation, Variance. Permutation & Combination: Permutation, combination & applications in bioinformatics. Programming in Scilab/ MATLAB: Variables, Variable declaration, Programming environment, Program flow, control statements- - if Statements, select statements, for loops, while loops break statements

Module III

Differential Equations: Overview of differential equations, Partial derivatives , Partial derivatives of Higher orders, Homogeneous function,

Euler's Theorem on Homogeneous Functions, Introduction, Total Differential Coefficient, Tangent plane to a surface, Error determination, Jacobian method, Properties of Jacobians, Partial Derivatives of Implicit Functions by Jacobian, Taylor's series, maximum or minimum, Lagrange's method of undermined Multipliers. Scilab/ MATLAB implementation of Jacobians, Taylor and Langrange's methods.

Module IV

Numerical Linear Algebra: Review of matrix algebra – basic matrix operations, determinants, eigen values & eigen vectors. Application of eigen values – differential equations, dynamical systems. Solution of linear equations – Gauss elimination, Gauss-Jordan, Gauss-Jacobi & Gauss Seidel methods & SciLab implementation.

Module V

Bio-Statistics: Introduction to principles of statistical sampling from a population, Methods of least squares, Chi-square test, random sampling, systematic and random sampling, accidental and systematic errors, correlation and regression analysis, frequency distributions and associated statistical measures. Probability distributions – Normal, Binomial, Poisson and extreme value distributions.

Reference

1. Monty J. Strauss, Gerald L. Bradley, Karl J. Smith, 'Calculus', Pearsons Education.
2. James Stewart, 'Calculus', Thomson Brooks.
3. Bio-Mathematics & statistics
4. V. Sundarapandian, 'Numerical Linear Algebra', Prentice Hall India Pvt. Ltd.
5. Bernad Kolman, David R. Hill, 'Introductory Linear algebra with Applications', Pearson Edition.
6. The MathWorks, Inc., 'MATLAB Getting started Guide'.

COB 513 Fundamentals of Molecular Biology

Module I

What is Biotechnology? Scope & Importance, Role of microorganism in biotechnology – E.coli as a model organism. Physical Nature of DNA, Watson-Crick model of DNA. RNA splicing and RNA editing, Gene organization, Gene regulation – Lac operon, Structural gene, Operator, Promoter, Repressor. DNA Amplification – Principle & Applications of PCR

Module II

Recombinant DNA Technology: Steps in the Construction of rDNA, Appilications of rDNA, Gene Cloning, Applications of Gene Cloning. Genetic Engineering : Enzymes Useful for Genetic Engineering – Exonuclease, Endonuclease, Restriction endonuclease, Ligase, Reverse transcpitase, DNA Polymerase. Foreign DNA – Cloning vectors – Plasmids, cosmids, cDNA, Genomic Library. DNA sequencing.

Module III

Protein sequencing. Overview of proteome analysis- invitro and invivo, Protein separation, purification and characterization- protein sample preparation, Separation methods - Separation by electrophoresis- isoelectric focusing, Two dimensional, SDS-PAGE, western blotting, micro arrays, mass spectrometry, peptide mass fingerprinting,

Module IV

Genetic Engineering – cDNA construction General Principal of Gene Therapy – Somatic Gene therapy, Germline Gene therapy. Antisense RNA and other Oligonucelitides, DNA fingerprinting – Methodology and Applications.

Module V

Introduction to Biophysics – Structure of atoms, molecules & chemical bonds, Molecular interactions – Van der Waals, electrostatic, hydrogen bonding, hydrophobic interactions. Thermodynamics – entropy, enthalpy, free energy, chemical kinetics, Solutions – pH, electrolytes, acids, bases, buffer. Principle and applications of chromatography, gel filtration, ultracentrifugation. Biophysical techniques for analysis of biomolecules –

X-ray diffraction, NMR, Photometry & spectrophotometry. Cell Energetics.

Reference

1. Biotechnology. David P.Clark
2. A Text Book of Biotechnology. R.C.Dubey
3. Cell and Molecular Biology. Gerald Karp
4. Instant Notes on Biotechnology
5. Cloning, Gene Expression and Protein Purification. Hardin Edwards
6. Essentials of Biophysics. P.Narayanan
7. Bioinstrumentation. John G.Webster
8. Biophysics, An Introduction. Rodney Cotterill
9. Basic Biophysics. M.Daniel
10. Biophysics. Vasantha Pattabhi, N. Gautham

COB 514 Perl & Bio-Perl

Module I

Perl Basics: Evolution & Environment – Features of Perl; Scalar Data & Operators, Control Structures,. Lists & Arrays, Array Functions, Associate Arrays, Arrays & Data Containers, Hash

Module II

Functions: User defined functions – Built in Functions, References, Regular Expressions – Processing Text with R.Es., Strings & Sorting Smart Matching. Perl Modules

Module III

Files & Directories: File I/O, Directory Operations; Perl & Relational Data bases, Perl and the Web, CGI & HTML Forms – Perl scripting for CGI, Cookies & Sessions

Module IV

Perl & Graphics, Process Management, Bio Perl– Basics, Overview of Bio Perl objects, seq objects, BLAST parsing, Annotated data base sequences.

Reference

1. Mastering Perl for Bioinformatics, James D. Jisdall, O'Reilly
2. Programming with Perl, John P. Flyut, Course Technology Publications
3. Schwart, et.al, Learning Perl, O'Reilly

COB 521 Computational Genomics & Transcriptomics

Module I

Overview: Central Dogma - DNA-RNA-Protein, Chromosome-Genome-Genes, Gene Loci, Genetic Linkage map, Physical map, Markers, Gene sequencing, Base calling-shot gun & clone contig approach, Genome annotation using Magpie, Restriction enzymes - Restriction sites - endonuclease and exonuclease – fragment assembly.

Module II

DNA Databases: String view of DNA - Basic file formats- Composition of DNA - Chagraff's Rule, prokaryotic DNA vs eukaryotic DNA, Six reading frames codon-genetic code-transcription & translation-mRNA, sense/coding as anti-sense/template strands, rRNA, tRNA, upstream and downstream genomic DNA, complimentary DNA - introns and exons, alternative splicing, junk DNA, EST, Sequence Data Bases, Detailed study of GenBank of NCBI- typical Gen Bank(DDBJ+EMBL) entry and for DNA and RNA, concepts of similarity - homologous, orthologous and paralogous sequences, TIGR, BLAST & FASTA file formats.

Module III

Sequence Alignments: The concepts and need for sequence alignments, Concepts of homologs, orthologs and paralogs, Dot plots - dotlet software, Types of alignments - local and global, Pair-wise and multiple, Sequence similarity and distances - Hamming and Levenshtein distances, similarity scores, match, mismatch and gap scores, affine scores, PAM and BLOSUM matrices, the Needleman-Wunch algorithm for global alignment, Smith – Waterman algorithm for local alignment, Statistical significance of alignments - e values, BLAST & FASTA algorithms - using and interpreting BLAST results, Specificity and sensitivity, Multiple sequence alignment - need, complexity and scoring MSA - the SP

measure - n dimensional dynamic programming- heuristic methods, ClustalX software.

Module IV

Sequence Representation & Analysis: Basic gene statistics–base counts, word(n-mer)frequencies, Codon usage bias, tandem and inverted repeats, vector contamination analysis, experiments using Perl scripts, ORFs, gene finding, splice site recognition, transcription factor binding site identification, promoters, enhancers, silencers, regulating motifs in DNA, sequence profiles, SNPs, sequence logos, sequence chromatograms, RNA sequence data analysis-RNA structure prediction.

Reference

1. Dan E. Krane and Michael L. Raymer, 'Fundamental concepts of Bioinformatics', Pearson Education (low Priced Edition).
2. Claverie & Notredame, 'Bioinformatics - A Beginners Guide', Wiley-Dreamtech India Pvt Ltd, 2003.
3. K Mani & N Vijayaraj, 'Bioinformatics: A practical approach', Aparna Publications, Coimbatore.
4. Setubal &Meidanis, 'Introduction to Computational Molecular Biology', Thomson: Brooks/Cole, International Student Edition, 2003.
5. Pevnezer, 'Bioinformatics and functional genomics', John Wiley.

COB 522 Computational Proteomics & Metabolomics

Module I

Overview: Proteome and proteomics, amino acids- structure, properties,- hydrophobicity and hydrophilicity, EIIP, molecular weight, α and ß propensities etc, proteins as polypeptides, backbone flexibility- Φ and ψ angles, proteins as workhorse molecules of life, Classification of protein.

Module II

Protein Databases: Primary databases, uniProtKB/Swiss-Prot, Specialised protein sequence databases GOA, ENZYME, Secondary databases such as InterPro, Structure databases-PDB, Typical entries, Protein visualization tools, Swiss PDB viewer, Expasy, Emboss, proteomic tools such as AACompIdent, AACompSim, MultiIdent, ProtParam, Compute pI/Mw etc.

Module III

Protein Structures: Primary, secondary, tertiary & quaternary structures, Interatomic forces- covalent interaction, hydrogen bonds, hydrophobic and hydrophilic interaction, charge/dipole interaction, Vander waals forces, steric interaction. Determining protein structure with X-ray Crystallography & NMR spectroscopy. Secondary structures- alpha helices, beta sheets and turns, Protein Domains- ProDom, PFAM, In- silico structure prediction- Chou Fasman method- p(a), p(b) and p(turn) propensities, Garnier Osguthorpe and Robson(GOR) method. Threading- Homology modeling, CASP, Abinitio prediction. Molecular dynamics & conformational energy calculation. Protein function prediction, Introduction to softwares: JPred, 3DPSSM, PhD, 123D, Modeller, Procheck, Ramachandran plot.

Module IV

Protein Classification : SCOP- class, fold, super-family, family, FSSP, CATH- architecture, topology/fold, homologous, sequence family levels classification, Metabolomics: Metabolism, metabolic pathways- Glycolysis, Krebs cycle, metabolomite, metabolome, metabolomic separation and analysis techniques, metabolic profiling, fingerprinting, Metabolome informatics. Resources/databases of metabolomics, applications.

Reference

1. Structures of Life, Downloadable document from NIH site
2. Dan E. Krane and Michael L. Raymer, fundamental concepts of Bioinformatics, Pearson Education (low Priced Edition).
3. Claverie & Notredame, Bioinformatics - A Beginners Guide, Wiley-Dreamtech India Pvt Ltd, 2003.
4. Pevnezer, Bioinformatics and functional genomics, John Wiley
5. Julio Collodo, 'Gene Regulation and Metabolism'.

COB 523 Computational Phylogenetics

Module I

Overview of theory of evolution- History of taxonomy and phylogeny; Linne's classification systems- Darwin's theory of evolution, Whittaker's five kingdom system, Carl Woese's three domain system. Traditional Systematics/phylogeny - Molecular data as molecular fossils; Four major factors of evolution: mutation, natural selection, drift, migration, Molecular-clock-hypothesis;

Module II

Basic terminology and concepts of phylogenetics: Trees-branches, taxa, node, root node, clade/monophyletic group, monophyletic vs paraphyletic; tree topologies - bifurcating/dichotomy and multifurcating/polytomy, rooted vs unrooted trees; Converting unrooted to rooted trees-use of outgroups and mid-point rooting approach; Gene phylogeny vs species phylogeny - speciation events vs gene duplication events-orthologs, paralogs, xenologs; Phylograms vs cladograms; The newick format; Building consensus trees-combinatorial explosion of tree topologies;

Module III

Constructing molecular phylogenetic trees-I: Choice of molecular markers- nucleotide Vs amino acid sequences;-DNA/RNA sour ces-nuclear, mitochondrial, ribosomal, chloroplast; multiple sequence alignments- quality of alignments-effect of homoplasy-transition/transversion ratio; Correcting homoplasy-statistical models: Poisson, Gamma and p-distance methods, Jukes-cantor model, Kimura model.

Module IV

Constructing molecular phylogenetic trees-II: Methods for constructing phylogenetic trees - Classification of methods; Distance based methods: Clustering based methods-UPGMA and neighbor joining, Optimality based: Fitch-Margoliash and minimum evolution algorithm; Character based methods-Maximum Parsimony(MP) amd Maximum Likelihood(ML) methods; Comparison of methods; Evaluation of phylogenetic trees-reliability and significance; Boot strapping-parametric and non-parametric; Jackknifing;

Module V

Phylogenetic softwares & applications: Overview of major features of PAUP and Phylip,T- coffee. One case study using PAUP or Phylip. Applied phylogenetics: Human Population Genetics: Mutation history of Y chromosome, alleles, Evolution and Variation, Genetic Polymorphism; systematics, ecological, conservation genetic approaches, use of phylogeny in drug discovery, immunology, HIV research, forensics, identifying how enzymatic functions evolves, study of protein function and super families.

Reference:

1. Jin Xiong, "Essential bioinformatics", Cambridge University Press (Chapter 10 and 11).
2. Marketa Zvelebil, Jeremy O. Baum,"Understanding Bioinformatics", Garland Science (Taylor and Francis Group). New York,2008. (Chapter 7 and 8)
3. Rastogi et al, Bioinformatics Methods and Applications Practice Hell of India (Chapter 5).

COB 524 Data Warehousing & Data Mining

Module I

Overview And Concepts: Need for data warehousing, Basic elements of data warehousing, Planning And Requirements: Project planning and management, Collecting the requirements. Architecture And Infrastructure: Architectural components, Infrastructure and metadata.

Module II

Data Design And Data Representation: Principles of dimensional modeling, Dimensional modeling advanced topics, data extraction, transformation and loading, data quality. Information Access And Delivery: Matching information to classes of users, OLAP in data

warehouse, Data warehousing and the web. Implementation And Maintenance: Physical design process, data warehouse deployment, growth and maintenance.

Module III

Introduction: Basics of data mining, related concepts, Data mining techniques.Data Mining Algorithms: Classification, Clustering, Association rules. Knowledge Discovery : KDD Process. Web Mining: Web Content Mining, Web Structure Mining, Web Usage mining.

Module IV

Advanced Topics: Spatial mining, Temporal mining. Visualisation : Data generalization and summarization-based characterization, Analytical characterization: analysis of attribute relevance, Mining class comparisons: Discriminating between different classes, Mining descriptive statistical measures in large databases Data Mining Primitives, Languages, and System Architectures: Data mining primitives, Query language, Designing GUI based on a data mining query language,

Reference

1. M.H. Dunham, "Data Mining Introductory and Advanced Topics", Pearson Education.
2. Han, Kamber, "Data Mining Concepts and Techniques", Morgan Kaufmann
3. Pieter Adriaans, Dolf Zantinge , "Data Mining", Pearson Education Asia
4. M Berry and G. Linoff, "Mastering Data Mining", John Wiley.
5. W.H. Inmon, "Building the Data Warehouses", Wiley Dreamtech.
6. R. Kimball, "The Data Warehouse Toolkit", John Wiley.

COB 531 Computer Aided Drug Design

Module I

Drug Discovery: The Evolution and Process, Brief History of drug discovery, Chemistry driven drug discovery, Natural product Chemistry, Target directed drug discovery. Concept of Disease: Genetic factors, environmental factors. Therapeutic targets and drug targets. Conventional Therapeutic Drugs: Small molecule drugs, Peptides and Protein mediators, antibodies, enzymes, vaccines, DNA products, Cell Based Therapies, Tissue and Organ Transplantation. An overview of Ayurveda and other alternative system of medicine. Traditional drugs in Kerala, Arogya Pachha - a case study.

Module II

Modern Drug Discovery: Modern Drug Discovery Process- Target Selection, Target Validation, Lead finding (HTS), Lead optimization, Pre-clinical Development, Clinical Development, Regulatory Approval, Product Registration. Pharmacology- Pharmacodynamics: Mechanism of Drug Action (Receptor Pharmacology) Pharmacokinetics: Membrane Transport, Absobtion and Distribution of drugs, Metabolism and Excretion of Drugs, Kinetics of Elimination, Pharmacotherapeutics: Clinical Pharmacology, Chemotherapy, Pharmacy, Toxicology, Adverse Drug Effects.

Module III

Computer Aided Drug Discovery: File Formats and Molecular Visualisation Tools, Database Libraries: PDB, Drug Bank, APD, GLIDA, ORPHANET, Super Drug Database, ChemBank, ChemPDB, KEGG, NCI etc. Vritual Screening: Docking, Types of search methods in Docking, Docking Algorithms. The scoring function. Denovo Design, Structure based Drug Design.Protein- Ligand Docking and Virtual Screening using GOLD, Scoring Functions: GOLD Score, Chem Score, Protein and Ligand Preparation, Binding Site Definition, Constraints and Restraints. Structure based drug design – receptor based, Ligand based, Pharmacophore modeling, combinatorial library design. Denovo drug design – Incrimental construction, Fragment based design, stochastic optimization, denovo design tools. Introduction to Immuno-informatics:

Vaccination and Immunization, Conventional and Reverse Vaccinology, Epitope Prediction Databases – IGMT, IEDB, Vaccine Design Tools.

Module IV

An Overview of Molecular Mechanics and Molecular Dynamics: Molecular Mechanics: Quantum Mechanical simulations, Ab-Initio methods, Semi-empirical methods, Force Field, Classical Force field, Energy minimizations, Algorithms for Energy Minimization. Application of Molecular mechanics in Drug Design. Molecular Dynamics: Algorithms to integrate the equation of motion. Approximation in molecular Dynamics. Limitations of Molecular Dynamics. Simulated annealing. Molecular Dynamics in drug design.

Module V

QSAR – Introduction. 2D QSAR, 3D QSAR, 4D QSAR. QSAR –a brief history. Biological activity – Ki, IC_{50}, ED_{50}, Km. Force Field – Steric, Electrostatic. Molecular descriptors – Topological, Constitutional, Geometrical, Electrostatic descriptors. Molecule Conformation generation methods – Distance geometry.

Reference

1. H.P.Rang,'Drug Discovery and Development-Technology in Transition', published by Reed Elsevier India Private Limited.
2. H.John Smith,'Smith and Williams's Introduction to the Principles of Drug Design and Action' Fourth Edition, published by CRC Press, Taylor and Francis Group.
3. P.N.Bennett & M.J.Brown,'Clinical Pharmacology' Tenth Edition, published by Reed Elsevier India Private Limited.
4. Susanna Wu-Pong & Yon Rojanasakul,'Biopharmaceutical Drug Design and Development' Second Edition, published by Humana Press.
5. E.J.Corey, Barbara Czako & Laszlo Kurti, 'Molecules and Medicine', published by John Wiley & Sons, Inc.
6. V.Kothekar, 'Essentials of Drug Designing', published by Dhruv Publications.
7. K.D.Tripathi, 'Essentials of Medical Pharmacology', published by Jaypee Brothers Medical Publishers (P) Ltd.
8. K.Anand Solomon, 'Molecular Modelling and Drug Design', published by MJP Publishers.
9. Juan Alvarez & Brian Shoichet, 'Virtual Screening in Drug Discovery', published byCRC Press, Taylor & Francis Group.
10. N. Claude Cohen, 'Guidebook on Molecular Modeling in Drug Design', published by Academic Press.

COB 532 Soft Computing Methods

Module I

Introduction to Soft Computing: History - The need for tolerance for imprecision, uncertainty, approximate reasoning etc for low cost solutions; Biological motivation in soft computing – comparison of human brain and CPU.

Module II

Artificial Neural Networks: Historic evolution – Perceptron, Features of NNs – Fault Tolerance, Parallelism, Flexibility and Adaptivity, Learning etc, Disadvantages; supervised and unsupervised networks, Multi - layer Perseptrons, Back Propagation Algorithm (Derivation not required), Learning & Momentum Parameters, selecting Hidden modes, Training & Testing, Overview of other ANNs -Kohonen's networks, Boltzman Machine, ART, ANN applications bioinformatics, overview of Support Vector Machines.

Module III

GA, and ACO: Genetic Algorithms, Basic Concepts, Reproduction – Cross over – Mutation - Fitness Value- Optimization using GAs; Applications in bioinformatics; Ant Colony Optimization: Swarm Intelligence – Basic motivation. –artificial ants – the bridge crosiing problem – theory and applications.

Module IV

Fuzzy Logic: Fuzzy sets, Membership functions, Logical Operations, Fuzzification & Defuzzification – Linguistic Variables, Fuzzy rule - based reasoning; Applications in Bioinformatics; overview of Nero – Fuzzy Systems. Dimensionality Reduction: Overview of Dimensionality Reductions techniques, K-means clustering, Principal Component Analysis -Eigen values & Eigen vectors of co-variance matrix, choosing components

Module V

Hidden Markov Models: Markov processes and Markov Models, Hidden Markov Models, 3 basic HMM problems & corresponding algorithms – Application in bioinformatics

Reference

1. Sushmita Mitra, Yoichi Hayashi, "Bioinformatics with Soft Computing", IEEE Transactions on systems, Man and Cybermetix, Vol.56, No.5,September 2006.
2. Dorrigo and Stutzle,'Ant Colony Optimization', Prentice Hall publications
3. David E Goldberg, 'Genetic algorithm', Pearson Education.
4. Timo Koshi, 'Hidden Markov model for Bioinformatics', Kluwer Academic Publishers.
5. M Ganesh, 'Introduction to Fuzzy sets and Fuzzy logic', Phi Publications, Inc.
6. Martin Gollery, 'Hand book for Hidden Markov model for Bioinformatics', CPC Press.

COB 533 Digital Signal Processing in Bioinformatics

Module I

Signals and Systems: Signals – Types of Signals – Discrete, continuous, digital, periodic, symmetric etc. Examples of signals; Sampling and Quantisation; Representation of Signals – Mathematical, Graphical and Numeric; System Concepts – Linear and Non-linear Systems – Principle of superposition, LTI systems, Causal systems, Bounded and unbounded systems, energy, stability; Review of complex numbers – notion of phasors generating signals

Module II

Standard Signals: Impulse, Step, Ramp, Sinusoid, Noise etc. Basic operations on Signals Correlation, cross correlation and auto correlation, Impulse response theorem, convolution. Time domain and frequency domain – spectral analysis – Fourier theorem – Spectrograms.

Module III

Filters & Windows: Filtering –, digital filter kernels – point, local and global filtering, various types of filters – FIR & IIR -– Hi, Lo- Pass filters, edge detection, contrast enhancement, Kalman filters. Design of filters – filter performance. Need for windows – types of windows.

Module IV

2-D Signal Processing – Images as 2-D signals – Grey level and Color Images – Image file formats - Various filters for digital image processing – Video image processing – Image compression

Module V

DSP and Bio-sequences: Mapping bio-sequences to digital signals – various approaches – indicator sequences – distance signals – use of clustering to reduce symbols in amino-acid sequences – use of chaos game representation to generate signals - analysis of bio-sequence signals – case study of spectral analysis for exon location and RRM for macro molecular interaction analysis.

Reference

1. Lyons, R.G. (2004). Understanding digital signal processing. Prentice Hall (2nd ed.).
2. Stein, J. (2000). Digital signal processing - a computer science perspective. Wiley.
3. Salomon, D. (2002). A guide to data compression methods. Springer.

4. Anastassiou, D. (2000). Digital Signal Processing of biomolecular sequences. Technical report, Dept. of EE, Columbia University, 2000-20-041.
5. Anastassiou, D. (2001). Genomic Signal Processing. IEEE Signal Processing Magazine 18(4), 8-20.
6. Anastassiou, D. (2000). Frequency-domain analysis of bimolecular sequences. Bioinformatics 16, 1073-1081.
7. Vaidyanathan, P. P. and Yoon, B. J. (2002). Gene and exon prediction using allpass-based filters. In: Workshop on Genomic Signal Processing and Statistics (GENSIPS), Raleigh, NC.
8. Tiwari, S., Ramachandran, S., Bhattacharya, A., Bhattacharya, S. and Ramaswamy, R. (1997). Prediction of probable genes by Fourier analysis of genomic sequences. Comput. Appl. Biosci. 13, 263-270.

COB 534 Mini Project

Mini Project is envisaged as a preparation for the main project and the guide lines for the same are applicable mutatis mutandis**.**

COB 541 Project Work and Viva Voce

AIM: To expose student to industry-standard project practices, through a real-life project work under time and deliverable constraints, applying the knowledge acquired through various courses.

Students are required to carryout a six month individual project and submit a dissertation embodying the findings of the same. The project work is to be done preferably in an External organization of repute such as national R and D institutions or global IT companies. Only in exceptional cases will the student be permitted to work in the Department of study itself. Suggested areas for project work are • Algorithm development and Software implementation in genomics and proteomics (example: Gene finding, splice site recognition, transcription factor binding site identification, protein secondary structure prediction etc.) • Phylogenetic analysis • Micro array data analysis and algorithm development for the same. • Systems Biology Modeling. • Molecular modeling & protein structure prediction. • Soft Computing methods for sequence classification

Assessment Components

There shall be 6 components that will be considered in assessing a project work, with weightage as indicated:

- * Timely completion of assigned tasks as evidenced by Academic Diary
- * Individual involvement, and adoption of industry work culture 10%
- * Quality of Project Documentation (precision, stylistics etc) 10 %
- * Achievements of project deliverables 20%
- Effective Technical Presentation of Project work 10 %
- Viva 20%

For the starred components, internal and external supervisors shall separately award marks and the same shall be averaged. External Evaluation of the dissertation and conduct of viva-voce shall be done by a board of examiners appointed by the University. The Viva shall, in addition to evaluation of project work, also attempt to gauge overall professional development of the student and also the generic subject awareness and knowledge of the student, mainly through an oral examination.

COB 542 Seminar

Seminars are opportunities for self study of a selected topic and a formal presentation of the same. Skills acquired in various courses should be utilized here in addition to practicing serious focused reading and assimilation of science and technology literature. Each student shall have a supervisor. Presentations shall be made using Powerpoint/Presenter slide show. The student shall also submit a report of not less than 20 pages. Even though seminar report need not have any original content, students are to ensure that the report is in their own words and any verbatim inclusion of material is properly cited. The seminar reports should be preferably made in LaTeX and follow IEEE style or Bioinformatics (OUP) style

COB 543 Systems Biology

Module I

Simple synthetic networks, regulatory networks: determining their basic building blocks and structural properties.

Module II

Metabolic pathways: high-level structural properties; gleaning evidence of robustness in these pathways; reconstructing metabolic pathways from data.

Module III

Obtaining biological insights by combining diverse sources of data: functional annotation of proteins; improving the reliability of the data; refining our current knowledge about biochemical pathways.

Module IV

'Foundations for Synthetic Biology', 'BioBricks: a standard for physical DNA composition', 'PoPs and RiPs: transcriptional device standards for functional composition', 'Biological part characterisation and quality control', 'Designing a biological system from Biobricks', e-Cell project.

Reference

1. Voigt CA. Genetic parts to program bacteria. Curr Opin Biotechnol 2006 Oct; 17(5) 548-57. doi:10.1016/j.copbio.2006.09.001 pmid:16978856. PubMed HubMed [1]
2. An introduction to Systems Biology, Alon 2006
3. Endy D. Foundations for engineering biology. Nature 2005 Nov 24; 438(7067) 449-53. doi:10.1038/nature04342 pmid:16306983. PubMed HubMed [2]
3. Andrianantoandro E, Basu S, Karig DK, and Weiss R. Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2006; 2 2006.0028. doi:10.1038/msb4100073 pmid:16738572. PubMed HubMed [3]
4. Hasty J, McMillen D, and Collins JJ. Engineered gene circuits. Nature 2002 Nov 14; 420(6912) 224-30. doi:10.1038/nature01257 pmid:12432407. PubMed HubMed [4]
5. Drubin DA, Way JC, and Silver PA. Designing biological systems. Genes Dev 2007 Feb 1; 21(3) 242-54. doi:10.1101/gad.1507207 pmid:17289915. PubMed HubMed [5]
6. SyntheticBiology.org
7. BioBricks Foundation
8. Registry of Standard Biological Parts
9. international Genetically Engineered Machine Competition (iGEM) (Imperial Team 2006 , Imperial Team 2007)
10. Royal Society Summer Exhibition 2007: 'Building Living Machines'
11. 'Google News' on 'Synthetic Biology'
12. Engineering Biology: a talk with Drew Endy (Edge.org)
13. CellDesigner (v.3.5.2), free download.
14. PlasmaDNA, University of Helsinki, free download.
15. MediaWiki (OWW script), Open Source
16. Computational Systems Bioinformatics Methods and Biomedical Application, Ziabbu Zhou
17. Drubin DA, Way JC, and Silver PA. Designing biological systems. Genes Dev 2007 Feb 1; 21(3) 242-54. doi:10.1101/gad.1507207 pmid:17289915. PubMed HubMed [5]
18. SyntheticBiology.org
19. BioBricks Foundation
20. Registry of Standard Biological Parts
21. international Genetically Engineered Machine Competition (iGEM) (Imperial Team 2006 , Imperial Team 2007)
22. Royal Society Summer Exhibition 2007: 'Building Living Machines'
23. 'Google News' on 'Synthetic Biology'
24. Engineering Biology: a talk with Drew Endy (Edge.org)
25. CellDesigner (v.3.5.2), free download.
26. PlasmaDNA, University of Helsinki, free download.
27. MediaWiki (OWW script), Open Source

28. http://openwetware.org/wiki/Imperial_College/Courses/Spring2008/Synthetic_Biology/Syllabu
29. Computational Systems Bioinformatics Methods and Biomedical Application, Ziabbu Zhou.

COB 501 Free and Open Source Software

Module I

Introduction: definitions & history; motivations & evolution; FOSS licenses

Module II

Getting Started: evaluating projects; downloading & installing; using & studying FOSS

Module III

Collaborating: version control systems; collaboration & tracking systems; defects & feature requests; analysis & design; workflow; adding, building, & leveraging FOSS

Module IV

Case Studies: Apache, Drupal, Moodle, Mozilla, RedHat, TWiki

Module V

Conclusions: FOSS culture; community dynamics; commercial influences & business models; unresolved issues & future directions.

Reference

1. DiBona, C., S. Ockman, M. Stone, B. Behlendorf, S. Bradner, J. Hamerly, K. McKusick, et al. 1999. Open Sources: Voices from the Open Source Revolution. 1st ed. O'Reilly Media.
2. Dionisio, J., and K. D. Dahlquist. 2008. Improving the computer science in bioinformatics through open source pedagogy. SIGCSE Bulletin. 40(2):115-119.
3. Fogel, K. 2005. Producing Open Source Software: How to Run a Successful Free Software Project. O'Reilly Media.
4. Ghosh, R. 2005. CODE: Collaborative Ownership and the Digital Economy. The MIT Press.
5. Golden, B. 2004. Succeeding with Open Source. 1st ed. Addison-Wesley Professional.
6. Hann, I, J. Roberts, S. Slaughter, and R. Fielding. 2002. Why do developers contribute to open source projects? First evidence of economic incentives. In Proceedings of the 2nd Workshop on Open Source Software Engineering. Orlando, FL: ACM.
7. Jensen, C., and W. Scacchi. 2007. Role migration and advancement processes in OSSD projects: A comparative case study. In Proceedings of the 29th International Conference on Software Engineering (ICSE'07), 364-374. Minneapolis, MN: IEEE Computer Society.
8. Moody, G. 2002. Rebel Code: Linux and the Open Source Revolution. 1st ed. Basic Books.
9. O'Mahony, S., and F. Ferraro. 2007. The emergence of governance in an open source community. The Academy of Management Journal 50(5): 1079-1106.
10. Raymond, E. 2001. The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media.
11. Roberts, J., I. Hann, and S. Slaughter. 2006. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. Management Science 52(7): 984-999.
12. Rosenberg, D. 2000. Open Source: The Unauthorized White Papers. 1st ed. Hungry Minds.
13. Weber, S. 2005. The Success of Open Source. Harvard University Press.
14. Wu, C., J. Gerlach, and C. Young. 2007. An empirical analysis of open source software developers' motivations and continuance intentions. Information Management 44(3):253-262.

COB 502 Professional Studies

Module I

Research and Technical Writing: What is research? The process of research – various types of research – research methodology - Hypothesis – research writing – basic principles; publication process – peer review - Journal impact factors – popular journals in Computational Biology & Bioinformatics (brief overview of their scope), Professional Societies in the field – their role in research and knowledge dissemination, Open Access Publications, Concept of ethics – its application in Scientific Research and Academics. Solving ethical conflicts, moral reasoning & ethical theories, responsibilities and rights.

Module II

Laws & Policies: Concept of Intellectual Property – Copy rights & Patents – Overview of the Indian & International laws on IPR – Process of Patenting – Patent Search Techniques & Sources. International treatise – GATT, TRIPS; IT & BT laws & Policies :Overview of the scope & key provisions of IT & BT Acts of Govt. of India including IT Act 2000, 2008, and various BT Acts. Patent Amendment of 2005 and its impact on Indian Pharmaceutical Industries. Overview of Bioinformatics Policy of Govt. of India; Profile of Key Bioinformatics/CB/ Pharmaceutical Institutions and Industries in India and also internationally.

Module III

Proprietary & Open-Source Software: Concept of software copy rights & patents – Free software movement- history and evolution – popular open source software & utilities- O.S., Office Package, Scientific publishing Packages like Tex etc. Overview of features and scope only; Total Quality Management, 7 sigma principles, Quality Certification in Software industries – ISO and CMM.

Module IV

Knowledge & Life Skills: Critical Reasoning Skills: Basic ideas of inductive and deductive logic, introduction to common logical fallacies; Informatics skills: Advanced Internet Search skills, Issues related to IT and Health; Time Management; Productivity in work place; Team work, who is a professional? Social attitudes, beliefs and values.

Module V

Miscellaneous Topics: Drug Licensing and Drug Marketing; starting an Industry; Job opportunities in Bioinformatics and Computational Biology – skill Profiles; Copyrights and patents applied to software & life forms, Application of Bioinformatics to Biodiversity, Gene Therapy, Genetic Engineering, Human Genetics, Agriculture, Anthropology Etc; Future of Biotechnology – bioinformatics, bio-electronics, bio-chips, bio-computers, dna computing, systems biology, e-cell project

Reference

1. Poornima M. Charantimath, Total Quality Management, Pearsons Education.
2. Indian IT & BT Acts.
3. Richard Stallman, Free software- A perspective, Prajasakthi Book House.
4. Gerala M. Noich, Learning to Think Things through.

COB 503 Genome Annotation & Ontology

Module I

Analysis of the whole genome sequencing data: Processing and assembly of whole genome sequence data, Base-calling (PHRED), Vector and E-coli masking. Assembly using PHRAP, CAP3, Assessment of final data quality (Coverage, PHRAP score International guidelines for data quality) Types of Misassembles and their solution.

Module II

Analysis and submission of EST and GSS data: Processing and quality trimming of nascent sequences; Preparation of submission files; Clustering of ESTs (overview of clustering procedure, pros and cons of clustering).

Module III

Whole Genome annotation strategies: Basic overview of whole genome annotation strategies, strategies for Human and Arabidopsis genomes.

Module IV

Gene Ontology Consortium: Basics concept and importance. Brief description of ontology, Methods and tools for assignment of ontology-G.O. consortium –structure controlled vocabularies, Organizing Principles of GO-Cellular component, biological processes, molecular function; GO slims, formats for ontology data-XML and any SQL.

Module V

Data Mining: Basic Concept of Knowledge Discovery and Data Mining, Decision Trees, Association rules, k-mean clustering, Principal Component Analysis, Clustering of Gene Micro array images with MAGIC software.

Reference

1. Zvelebil & Baum, 'Understanding Bioinformatics'.
2. Primorose &Twyman, 'Principles of Gene Manipulation & Genomics'.
3. Arthur M. Lest, 'Data Base Annotation in Molecular Biology'.
4. Azuaje, Dopazo, 'Data Analysis and Visualization in Genomics & Proteomics'.

COB 504 Computer Networks & Web Programming

Module I

Overview of Computer Networking: History & development of computer network, network topologies, ISO reference model. MAC layer: Aloha Protocols, LAN—Ethernet, token ring and data link layer. Network layer: Routing algorithms, Congestion Control algorithms, multicast and mobile routing. Internetworking: Bridges, Switches, Repeaters and Routers. Transport Layer: Connection Management, Flow control and buffering. Application Layer: DNS, SNMP, MAIL, WWW, and FTP.

Module II

Programming in Java: Data types, Variables, Arrays, Operators, Classes & Methods, Inheritance Packages and Interfaces. Exception Handling - multithreaded programming, Applets - Database connectivity: JDBC API - loading database drivers, establishing a database connection, issuing dynamic SQL statements, processing a ResultSet.

Module III

Client-side Technologies: Introduction to Markup language :SGML, HTML, DHTML. Introduction to scripting languages:JavaScript, PHP, XML elements, attributes, Prolog, CDATA, XML parser, namespaces, BSML fundamentals, introduction to DTD, DTD for bioinformatics, XML schemas for bioinformatics.

Module IV

Server-side Technologies: J2EE: Architecture ,Servlets, Understanding the http protocol, Writing servlets, Servlet API, Writing servlets to receive requests and send responses, JSP - Server processing of JSP's, Java programs in JSP's, Applying MVC principles using JSP's, Remote Method Invocation and JavaBeans.., Creating a simple web service, web servers: Tomcat, Apache, Web sphere, Web logic etc in Windows.

Reference

1. A. S. Tennanbaum, 'Computer Network', 2nd Edition, PHI.
2. D. E. Comer, 'Internetworking with TCP/IP: Principles, Protocols and Architecture' Vol. I, 2nd edition, PHI.
3. Ethan Cerami, 'XML for Bioinformatics', Springer.
4. Patrick Naughton and Herbert Schildt, 'Java 2 - The Complete Reference 3/e', , TMH.
5. Walsh, Aaron E, 'J2EE a Professional Guide', Tata MC-Graw Hill.

COB 505 Computational Biology Algorithms

Module I

Introduction: Algorithm design paradims, algorithm efficiency, run time analysis, asymptotic notations; big O, omega, theta. Greedy Method: Overview, examples & analysis of algorithms, optimisation problems, minimum cost spanning tree, approximation problem, Knapsack, single source shortest path,prims algorithm, Kruskal's algorithm. Huffman tree.

Greedy method in Computational Biology applications- genome rearrangement, sorting by reversal, motif finding.

Module II

Decrease and Conquer: Overview and structure, examples- depth first search, breadth first search, insertion sort, generating permutations, topological sorting. Divide and Conquer- Overview and structure, examples- binary search, quick sort, Strassen matrix multiplication, recurrence relation. Computational Biology applications – space efficient sequence alignment, block alignment.

Module III

Transform and Conquer: Overview and structure, examples –pre-sorting, hashing- separate chaining, closed addressing-, merge sort. Dynamic Programming - Overview and structure, examples –shortest path, computing binomial coefficiaent, Warshall's algorithm, Floyd's algorithm, Dynamic Programming in Computational Biology applications – longest common sub-sequence, Manhattan tourist problem, local and global alignment, scoring alignment gap penalty alignment, gene prediction, and statistical approaches in gene prediction.

Module IV

Back tracking: Overview & structure, examples - 8-queens & knapsack problems. Branch & Bound: Overview & structure, examples – travelling sales man problem & 0/1 knapsack. Graph Algorithms in Computational Biology applications: Graph and Genetics, shortest superstring problem, SBH & Hamiltonian problem, SBH & Eulerian problem, identification and database search.

Reference

1. Neil C. Johns and Pavel A. Pevzner, 'An Introduction to Bioinformatics Algorithms', Ane Books
2. Horowitz , S. Sahini, and Rajasekharan, 'Fundamentals of Computer Algorithms', Galgotia Publications
3. 3. T. H. Coreman, Leiserson, Rivest and Stein, 'Introduction to Computer Algorithms', PHI.

COB 506 Data Structures for Bioinformatics

Module I

Review of C Programming Fundamentals: Fundamental data types & storage classes, Operators & expressions, conditional program execution, iterative statements, modular programming – functions, parameter passing, scope and life time –arrays, structures, unions, pointers, C processor, file operations.

Module II

Data Structure Fundamentals:Basic terminology, elementary data organisation, Arrays: Definition, linear arrays - representation, traversal, insertion & deletion – two dimensional arrays – representation, traversal, insertion, deletion – bubble sort, linear search, binary search, various string operations- traversal, searching, replacing, string alignment, - multi-dimensional arrays.

Module III

Stack & Queues**:** Basic concepts, array representation basic operations, applications, recursion, infix, prefix & postfix notations, infix postfix expressions & conversion. Queues basic concepts & operations, Priority queues & Dequeues. Linked Lists: Singly linked lists**-** basic concepts, representation, operations – traversal, searching, insertion, deletion, header node. Doubly linked list: operations – traversal, searching, insertion, deletion, applications – queue, stack, polynomial addition.

Module IV

Trees & Graphs: Tree- Definition, Basic terminology, binary tree, array & linked list representation complete & binary trees, extended binary tree. Fundamentals operations –traversal, searching, insertion deletion, algorithm complexity. Types of trees & operations & algorithm complexity– Binary search tree, height balanced tree, AVL tree, red-black tree, threaded tree. Graph: Definition, basic terminology, representation

–sequential & linked, Types of graphs –multi graph, direct graph, etc connected components, spanning tree. Shortest path, topological sorting.

Module V

Files & Sorting: Sorting- Insertion sort, selection sort, shell sort, quick, merge sort, heap sort, complexity, external sort. File Structure - Physical storage media, file organization, record blocks, sequential, indexed, indexed sequential, hashing methods, primary & secondary indices, B tree and B+ tree files. Biological data storage, searching & retrieval using various file management methods.

Reference

1. Horowitz and Sahini. 'Fundamentals of Data Structure', Galgotia Publications
2. Tannenbaum, 'Data Structures', PHI
3. R. L. Kruse, B. P. Leary, ' Data Structures and Program Design in C', PHI.

COB 507 Grid Computing

Modue I

Grid Computing: Introduction-Definition and Scope of Grid Computing. Grid Computing Initiatives.

Module II

Grid Computing Organizations and their roles – Grid Computing analog- Grid Computing road map.

Module III

Grid Computing Applications Merging the Grid Sources – Architecture with the Web Devices.

Module IV

Architecture .Technologies: OGSA – Sample use cases – OGSA platform components – OGSA Basic Services.

Module V

Grid computing tool kits: Globes GT 3 Toolkit – Architecture, Programming model, High level services – OGSI. Net middleware solutions.

Reference

1. C.S.R. Prabhu, Grid & Cluster Computing, PHI 2008
2. Joshy Joseph & Craig Fellenstein, "Grid Computing", PHI, PTR – 2003
3. Ahmar Abbas, "Grid Computing: A Practical Guide to technology and Applications", Charles River media – 2003.

COB 508 Micro Array & Image Processing Techniques

Module I

Overview of Microarray: Basic concepts - Concept of gene expression- Comparative Genomics, Types of Micro arrays: DNA Microarray, Tissue Microarrays, Antibody Microarrays, Protein Microarrays. DNA Microarray. Protocol. Making microarrays - Spotted Microarrays -Insitu synthesised Oligonucleotide arrays - Affymetrix Technology - Inkjet array synthesis - Using microarrays - Sample preparation and labeling - Hybridisation - Washing - Image Acquisition, Computer design of Oligonucleotide probes.

Module II

Basis of image processing: Digital representation of images, resolution, Quantization, image file formats, basic image handling in MATLAB or SciLab, Basic point operations in images, Concept of filtering, high and low pass filtering, histogram stretching

Module III

Advanced Image Processing: Segmentation and pattern recognition in images, Spectral and wavelet analysis of images- Basic idea of spectral decomposition-Fourier theorem- Discrete Fourier transforms, Fast Fourier transform, MATLAb/ Scilab implementation. Image Compression methods- Concept of wavelets, Image Comparison using Haar Wavelets

Module IV

Microarray Image Processing: Image Formats - Feature Extraction - Normalisation - Measuring and Quantifying Microarray Variability - Analysis of Differentially Expressed Genes, experiments Using MAGIC

Module V

Data Mining: Basic Concept of Knowledge Discovery and Data Mining, Decision Trees, Association rules, k-mean clustering, Principal Component Analysis, Clustering of Gene Micro array images with MAGIC software.

Reference

1. Dov Stekel, Microarray Bioinformatics, Cambridge University Press (Indian Edition).
2. Helen C Causton et. Al., Microarray Gene Expression Data Analysis, Blackwell Publishing.
3. Sorin Draghici, Alexander Kuklin , Data Analysis Tools for DNA Microarrays, CRC Press.
4. Dhammika Amaratunga, Javier Cabrera, Exploration and Analysis of DNA Microarray and Protein Array Data, Wiley-IEEE.
5. Daniel P Berrar, Werner Dubitzky, Martin Granzow, A Practical Approach to Microarray Data Analysis, Springer
6. Richard J. Roiger and Michael W. Geat, Data Mining-a Tutorial based primer, Pearson Education
7. Jason T L Wang, Mohammed J Zaki, Hannu T T Toivonen, Dennis Shasha, Data Mining in Bioinformatics, 2004
8. Arun K Pujari, Data Mining Techniques, Universities Press
   MAGIC Software Documentation (freely downloadable)

COB 509 Programming in R

Module I

Introduction: Installing R; R as a deluxe calculator, creating objects and assigning values.

Module II

Graphics; simple plotting, advanced plotting, using color in plots, using subscripts and superscripts in graph labels, interactive graphics, saving graphical output, loops.

Module III

Working with data sets: Data structures, moving to and from files, character data, generating random values.

Module IV

Statistical distributions: The use of statistical distributions in epidemiology, statistical analyses, writing functions, closing a session.

References

1. ' R Programming for Bioinformatics', by Robert Gentleman. ISBN-13: 978-1420063677
2. Crawley, M.J. 2007. The R Book. Wiley.
3. http://www.apsnet.org/Education/AdvancedPlantPath/Topics/RModules/Doc0/
4. Gentleman, R., Carey, V.J., Huber, W., Irizarry, R.A., and Dudoit, S., eds. 2005. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. Springer, New York.
5. Murrell, P. 2006. R Graphics. Chapman & Hall/CRC, Boca Raton.
6. Venables, W.N., Smith, D.M., and the R Development Core Team. 2007. An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. Version 2.6.1. http://cran.r-project.org/doc/manuals/R-intro.pdf.