Course Title with Credit Load
M.Sc. in Bioinformatics

Course Code	Course Title	Credits (L+P)
	Major: 20 credits (12 credits of core plus 8 credits of optional)	
BI 501	Introduction to Bioinformatics & Computational Biology*	2+1
BI 502	Statistical Genomics*	2+1
BI 503	Genome Assembly and Annotation	1+1
BI 504	Biomolecular Modelling and Simulation*	2+1
BI 505	Transcriptomics and Metagenomics	2+1
BI 506	Biological Data Management*	2+1
BI 507	Biological network modelling and analysis	2+1
BI 508	Computer programming in bioinformatics	2+1
BI 509	Machine Learning Techniques in bioinformatics	2+1
	Minor (8 credits)	
	Molecular Biology and Biotechnology	
	Biochemistry	
	Genetics and Plant Breeding	
	Microbiology	
	Any other related discipline	
	Supporting (6 credits)	
	Statistics	
	Mathematics	
	Computer Science	
	Any other appropriate discipline	
	Common courses	5
BI	Seminar	0+1
BI 500	Research	0+30
	Total	70

*Core courses

Courses Contents
M.Sc. in Bioinformatics

I. **Course Title** : **Introduction to Bioinformatics and Computational Biology**

II. **Course Code** : **BI 501**

III. **Credit Hours** : 2+1

IV. **Aim of the course**

To provide theoretical and practical knowledge about handling and processing of genomic data, optimization and data mining techniques used in bioinformatics.

V. **Theory**

Unit I (15 Lectures)

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

Unit II (5 Lectures)

Preprocessing of gene expression data; Data Normalization techniques, Data quality control: Modelling of errors, Imputation etc; High-throughput screening.

Unit III (6 Lectures)

Optimization Techniques: concept and applications, Simulated Annealing, Genetic Algorithms: *Ab initio* methods for structure prediction; Information theory, entropy and relative entropy.

Unit IV (6 Lectures)

Foundations for Machine learning Techniques: Unsupervised and Supervised Learning, Cross Validation Techniques, Markov Model, Bayesian Inference: concepts and applications, Hidden Markov Model and applications, Introduction to WEKA package.

VI. **Practicals**

Database Similarity Searches, Multiple sequence alignment, Genome databases, Structural databases, Derived databases, Gene annotation, Gene prediction software. Analysis of DNA microarray experiments, Expression profiling by microarray/gene chip, Proteomics, Pattern recognition, Hidden Markov Models, Gibbs Sampling, Analysis of single and multiple DNA or protein sequences.

VII. **Suggested Reading**

- Baldi, P. and Brunak, S. 2001. *Bioinformatics: The Machine Learning Approach*. MIT Press.
- Baxevanis, A.D. and Francis, B.F. 2004. *Bioinformatics: A Practical Guide to the Analysis of*

Genes and Proteins. John Wiley.

- Wang JTL, Zaki MJ, Toivonen HTT and Shasha D. 2004. Data Mining in Bioinformatics.Springer.
- Amaratunga D and Cabrera J. 2004. *Exploration and Analysis of DNA Microarray and Protein Array*. John Wiley.
- Gupta GK. 2006. *Introduction to Data Mining with Case Studies*. Prentice Hall of India, New Delhi.
- Han J and Kamber M. 2006. *Data Mining: Concepts and Techniques*. Morgan Kaufman.
- Hand DH, Mannila P Smyth. 2001. *Principles of Data Mining*. Prentice Hall of India, New Delhi.

I. Course Title : Statistical Genomics

II. Course Code : BI 502

III. Credit Hours : 2+1

IV. Aim of the course

This course builds the basic understanding of statistical methods used in genetics and genomics.

VI. Theory

Unit I (14 Lectures)

Fundamentals of Population genetics: Hardy –Weinberg law, Effect of systematic forces on changes in gene frequency; Principles of Quantitative genetics: Values, Means and Variances, Detection and Estimation of Linkage, Inbreeding, Selection, Genetic Parameter Estimation, Variance component estimation, BLUP, G x E interaction, Path Analysis

Unit II (10 Lectures)

Molecular Marker based classification: similarity measures, clustering methods, bootstrapping; QTL mapping: Detection and Estimation of QTL, Single Marker Analysis, Interval Mapping and MQM;

Unit III (8 Lectures)

Design and Analysis of Expression Data; Genome Selection; Genome Prediction, Genetic Markers, Association Mapping; Genome Wide Association Analysis

VII. Practicals (16 Lectures)

Population genetics: Hardy-Weinberg law, Estimation of linkage, Inbreeding, Selection, Genetic parameter estimation, Variance component estimation, BLUP, Path analysis, Molecular marker based classification, Estimation of QTL, Single marker analysis, MQM, Analysis of gene expression data, Genome selection and Genome prediction.

VIII. Suggested Reading

- Xu, Shizhong. 2013. *Principles of Statistical Genomics*. Springer
- Ben Hui Liu. 1997. *Statistical Genomics: Linkage, Mapping, and QTL Analysis*.
- Sorensen D and Gianola D. 2002. *Likelihood, Bayesian and MCMC Methods in Genetics*. Springer.
- Ben HL and Leming MS. 2013. *Statistical Genomics and Bioinformatics*.

I. Course Title : Genome Assembly and Annotation

II. Course Code : BI 503

III. Credit Hours : 2+1

IV. Aim of the course

The primary objective of this course is to develop practical understanding of techniques and tools used in genome assembly with emphasis on issues and challenges of its structural and functional annotation.

V. Theory

Unit I (6 Lectures)

Types and methods of genome sequence data generation; Shot gun sequencing method; Problems of genome assembly, Approaches of genome assembly: Comparative Assembly, DE novo Assembly; Read coverages; Sequencing errors, Sequence Quality Matrix, Assembly Evaluation; Challenges in Genome Assembly.

Unit II (5 Lectures)

Various tools and related methods of genome assembly: MIRA, Velvet, ABySS, ALLPATHS-LG, Bambus2, Celera Assembler, SGA, SOAP*denovo,* etc.

Unit III (5 Lectures)

Basic concepts of genome annotation; Structural and Functional Annotation; Identification of open reading frame (ORF) and their regularization, Identification of gene structure, coding regions and location of regulatory motifs

VI. Practicals (16 Lectures)

Genome assembly methods for data from various sequencing platform, Sequencing error determination, Sequence quality matrix; Various tools for genome assembly: MIRA, Velvet, ABySS, ALLPATHS-LG, Bambus2, Celera Assembler, SGA, SOAP *denovo*, etc. Structural and functional Genome annotation.

VII. Suggested Reading

- Jung, S., Paul, Gordon, M.K., Sensen, C. W. 2012. *Genome Annotation*. Chapman and Hall/ CRC
- Venter, J. C., 2000. *Annotation of the Celera Human Genome Assembly*. Celera.
- Mark Menor. 2007. *Multi-genome Annotation of Genome Fragments Using Hidden Markov Model Profiles*
- Carson Hinton Holt. 2012. *Tools and Techniques for Genome Annotation and Analysis*
- Alistair G. Rust, Emmanuel Mongin and Ewan Birney Loraine A.E and Helt G.A. 2002. *Genome annotation techniques: new approaches and challenges. Drug Discovery Today*. 570-576 p.
- Weizhong Li and Adam Godzik. 2002. *Discovering new genes with advanced homology detection. Trends in Biotechnology*, 20: 8, 315-316 p.

I. Course Title : Bio-molecular Modelling and Simulation

II. Course Code : BI 504

III. Credit Hours : 2+1

IV. Aim of the course

The course aimed to develop understanding of bio molecular modelling techniques and simulation.

V. Theory

Unit I (8 Lectures)

Methods for 3D Structure Prediction: Homology modeling of protein 3D structures – approaches to loop building, energy considerations and evaluation of the accuracy of the model. *ab initio* approach to 3D structure prediction; Threading approach to 3D structure prediction. A Comparison of protein structure prediction methods: CASP

Unit II (8 Lectures)

Basic principles of modeling, modeling by energy minimization technique, concept of rotation about bonds, energy minimization by basic technique for small molecules, Ramachandran plot, torsional space minimization, energy minimization in Cartesian space, molecular mechanics-basic principle

Unit III (8 Lectures)

Basic concepts of Simulation Modelling: Units and derivatives, Force field and energy landscape, Truncation of non-bonded interactions, Introduction to solvation, Periodic boundary condition, Wald summation, implicit solvent model and continuum electrostatics, Monte Carlo simulation on parallel computers. Replica-exchange simulations, Restraint potentials, Free energy calculations, Membrane simulations

Unit IV (8 Lectures)

Energy Minimization: Concept of energy minimization - hypersurface, local and global energy minima, statement of problem. Derivative minimization methods - first derivative methods: the steepest descents method, line search in one dimension, arbitrary step approach, conjugate gradients minimization. Second derivative method – the Newton-Raphson method. Applications of energy minimization.

VI. Practicals: (16 Lectures)

Protein structure databases: PDB, MODBASE, Structure visualization – Rasmol and PyMol, Structural analysis- classification, CATH, SCOP, Protein geometry – bond length, bond angle, torsion angle, calculation of surface area, volume and radii: Swiss PDB Viewer. Small molecule generation - peptides and nucleic acids: ISIS draw / ChemSketch, Selection of query sequence, template selection: pdbBLAST, Comparative 3D structure prediction – SWISSMODEL, Model generation - building side chains and loops using Modeller, Threading, *ab initio* modeling, Structure validation - generation and analysis of Ramachandran plot using PROCHECK, WHATCHECK via SAVS server, Force field calculation and energy minimization, Structure refinement - loop building, removing non-bonded contacts, adding missing side chains via WhatIf interface, Scoring structural similarity - 3D structure alignment - RMS superimposition – VMD, Molecular dynamics simulation using Tinker. Simulation dynamics, Monte carlo simulation on parallel computers. Replica exchange simulation, free energy calculation. Docking

VII. Suggested Reading

- Schlick T. 2010. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Science.
- Gunsteren WF, Weiner PK, Wilkinson AJ. 1997. *Computer Simulation of Biomolecular Systems: Theoretical and experimental application*. Springer.
- Martin JF. 2007. *A Practical Introduction to the Simulation of Molecular Systems*. Cambridge University Press.
- Leach AR. 2001. *Molecular Modeling: Principles and Applications*. Prentice Hall. 784p.

- Bourne PE and H Weissig. 2003. *Structural Bioinformatics*. Wiley-Liss. 650 p.
- Marx D and Hutter J. 2009. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge University Press. 578p.

I. Course Title : Transcriptomics and Metagenomics
II. Course Code : BI 505
III. Credit Hours : 2+1

IV. Aim of the course

The course aims to teach basic concepts of metagenomics and various techniques used in the analysis of metagenomic data

V. Theory

Unit I (8 Lectures)

Microarrays, RNA-seq, Chip-Seq, EST-clustering, differential expression analysis

Unit II (6 Lectures)

Taxonomic and genetic annotation of high throughput sequence data, microbial diversity analyses, analyses of microbial community composition and change and metabolic reconstruction analyses.

Unit III (9 Lectures)

Comparison between Metagenomics and AL, EC, Comparison between LCS and Metagenomics, Symbiotic Evaluations: SANE, Comparison between SANE and Metagenomics, Horizontal Gene Transfer: Microbial GA.

Unit IV (9 Lectures)

Metagenome Sequencing, Single Cell Analysis, Host-Pathogen Interaction; Shotgun metagenomics; High-throughput sequencing; Comparative metagenomics; Community metabolism; Metatranscriptomics.

VI. Practicals (16 Lectures)

Meta genome annotation, Analyses of microbial community composition and change and metabolic reconstruction analyses; Metatranscriptomics; Comparative metagenomics. Microarray data analysis; RNA-seq, chip-seq, EST-clustering.

VII. Suggested Reading

- Diana marco. 2010. *Metagenomics: Theory, Methods and Applications*. Ceister academic press
- Streit WR and Daniel R. 2010. *Metagenomics: Methods and Protocols*. Springer protocols.
- Yeh WK, Yang H, McCarthy JR. 2010. *Enzyme Technologies: Metagenomics, Evolution, Biocatalysis and Biosynthesis*.wiley
- Muthukumar V. 2003. *Metagenomics for the Identification of Plant Viruses*. ProQuest.

I. Course Title : Biological Data Management
II. Course Code : BI 506
III. Credit Hours : 2+1

IV. Aim of the course

The course aims at teaching database management system and familiarizing with the techniques of data sources, data curation and integration of data sources

V. Theory

Unit I (6 Lectures)

Database Management System (DBMS): Need for DBMS - File system vs Database system, Advantages of DBMS - DBMS Architecture – DBMS services - Data abstraction - Overview of Data Models: Hierarchical Model - Network Model - Entity-Relationship (E-R) Model: Symbols - Components of E-R Model: Entities, Attributes, Relationships - Relational Model, Object-oriented Model.

Unit II (8 Lectures)

Overview of Relational Database Objects – Relation – Tuple - Cardinality – Attribute – Degree - Domain - Primary key – Foreign key - Relational data structure – Relational Data Integrity and Constraints: Domain constraints, Entity integrity, Referential Integrity, Operational constraints - Codd's Rules – Normalization: 1NF, 2NF, 3NF, BCNF, 4NF and 5NF.

Unit III (8 Lectures)

Structured Query Language (SQL): Overview of SQL – SQL Data types and Literals – SQL Commands: Data Definition Language (DDL), Data Manipulation Language (DML), Data Querying Language (DQL), Data Control Language (DCL), Data Administration Statements (DAS), Transaction Control Statements (TCS), SQL Operators: Arithmetic, Comparison, Logical and Set Operators – SQL Query, Nested Query - SQL Aggregate functions.

Unit V (10 Lectures)

Curation of genomic, genetic, proteomic data, High-throughput screening, array, qPCR data sets; Quality management of data: tools and techniques. Biological data sources, Data granularity, Schema modelling, architecture, query design, extraction, transformation and loading, Long term data management, storage and security. Bio-chip information system, visualization and reporting, Risk factors for data quality management. Un-structured or noSQL database; AI and BIG data Analytics

VI. Practicals (16 Lectures)

Understanding the data sources, Data granularity, Data modeling and architecture, development of database, Storage, Security, Visualization and reporting.

VII. Suggested Reading

- Kozak K. 2010. *Large scale data handling in biology*. Ventus Publishing ApS. ISBN 978-87-7681-555-4.
- Harold, E. and Means W.S. 2004. *XML in a Nutshell*, Third Ed. O'Reilly, Sebastopol, CA
- Witten, I.H. and Frank E. 2005. *Data Mining: Practical Machine Learning Tools and Techniques* (WEKA), 2nd Ed. San Francisco, Morgan Kaufmann,
- Lodish, H. *et al*. 2000. *Molecular cell biology*. New York: Freeman & Co.
- Kaneko K. 2006. *Life: An Introduction to Complex Systems Biology*. Springer.

I. Course Title : **Biological Network Modelling and Analysis**
II. Course Code : **BI 507**
III. Credit Hours : 2+1

IV. Aim of the course

This course aims to develop basic understanding of system biology through biological network modelling and its analysis.

V. Theory

Unit I (12 Lectures)

Introduction to biological networks, Graph theoretic modelling and analysis of biological networks, Discrete Dynamic modelling (Boolean networks, Petri nets), Continuous dynamic modelling (ODEs, stochastic simulation, etc.)

Unit II (12 Lectures)

Probabilistic modelling (Probabilistic Boolean networks, Bayesian networks, Mutual Information), Network inference from experimental data, Genome-scale modelling and network integration

Unit III (8 Lectures)

Evolution of molecular networks, Network-guided GWAS studies, FBA and epistasis detection, protein function prediction

VI. Practicals (16 Lectures)

Biological networks, Graph theoretic modelling and analysis of biological networks, Discrete Dynamic modeling; Continuous dynamic modeling; Probabilistic modeling; Genome-scale modelling and network integration; Evolution of molecular networks, Network-guided GWAS studies, FBA and epistasis detection, protein function prediction.

VII. Suggested Reading

- Junker BH. 2008. *Analysis of Biological Networks.*
- Koch I Reisig, W. Schreiber F. 2010. *Modeling in Systems Biology: The Petri Net Approach.*
- Ramadan EY. 2008. *Biological Networks: Modeling and Structural Analysis.*
- Laubenbacher R. 2007. *Modeling and Simulation of Biological Networks.*

I. Course Title : **Computer Programming in Bioinformatics**
II. Course Code : **BI 508**
III. Credit Hours : **2+1**

IV. Aim of the course

To learn programming skills for parsing biological data, parallel computing, database connectivity and web-interface.

V. Theory

Unit I (7 Lectures)

BioJava- Packages, Data Import, Manipulation; Python- Basic Syntax, Loops, Functions; BioPython.

Unit II (7 Lectures)

Bioperl: Introduction, Modules: SeqIO, SearchIO, Seq Feature, Finding introns, Alignments, LiveSeq and Tree.

Unit III (12 Lectures)

OpenMP: Clauses, Worksharing constructs, Synchronization constructs, Environment variables, Global Data, Runtime functions, Message Passing Interface (MPI): Introduction and programming, Point to point communications, Collective communications, Advanced MPI1 concepts, MPI2 introduction, Hybrid (openMP + MPI) programming.

Unit IV (6 Lectures)

Compute Unified Device Architecture (CUDA): Introduction and Programming, GPU computing.

VI. Practicals (16 Lectures)

BioPerl programing using bioperl modules such as SeqIO, SearchIO, LiveSeq and Tree; OpenMP programming on Work sharing and Synchronization constructs, Environment variables and global data; MPI programming on Point to point communications and Collective communications; Compilation of OpenMP and MPI programs; Execution of OpenMP and MPI programs; Use of high performance computing, computing resources and job scheduling.

VII. Suggested Reading

- Tisdall J. 2001. *Beginning Perl for Bioinformatics.* O-Reilly.
- Schwartz RL, Phoenix T, Foy BD. 2008. Learning Perl. O-Reilly.
- Orfali R and Harkey H. 1999. *Client/Server Programming with JAVA and CORBA.* John Wiley.
- Sriram Srinivasan. 1997. *Advanced Perl Programming.* O-Reilly.
- Bunce T and Descartes A. 2000. *Programming the Perl DBI.* O-Reilly.
- Mitchell L Model. 2010. *Bioinformatics Programming Using Python*, O'Reilly media, Cambridge, Bal HP 2003. *Perl Programming for Bioinformatics*, Tata McGraw Hill.

I. Course Title : **Machine Learning Techniques in Bioinformatics**
II. Course Code : **BI 509**
III. Credit Hours : **2+1**

IV. Aim of the course

The purpose of the course is to explain various machine learning techniques and its applications on biological data.

V. Theory

Unit I (10 Lectures)

Introduction to statistical learning theory, Empirical Risk Minimization, Structural Risk Minimization; Classification: Decision tree, Bayesian, Rule based classification, ANN, SVM, KNN; Case based reasoning and Applications in Bioinformatics.

Unit II (12 Lectures)

Clustering: Partition Methods, Heirarchical methods, Density based methods, Grid based clustering, Model based clustering, clustering of high dimensional data, constraints based clustering, Analysis of MD trajectories, Protein Array data Analysis.

Unit III (10 Lectures)

Dimensional Reduction Techniques, Methods of Feature Selection, Resampling Techniques, Elements of Text Mining and Web Mining, Soft Computing and Fuzzy logic system and application in bioinformatics.

VI. Practicals (16 Lectures)

Decision tree, classification techniques: ANN, SVM, KNN, Case based reasoning and its applications on biological data. Clustering techniques; Clustering of high dimensional data; Dimensional reduction techniques; Resampling techniques; Text mining and Web mining. Soft Computing and Fuzzy logic system & application in bioinformatics.

VII. Suggested Reading

- Witten, H.I., Frank, E. and Hall, M.A. 2011. *Data Mining: Practical Machine Learning Tools and Techniques.*
- Hastie, T., Tibshirani, R., Friedman, J.H. 2009. *The Elements of Statistical Learning: Data Mining Interface and Prediction.*
- Clarke, S.B., Fokoue, E. and Zhang, H.H. 2009. *Principles and Theory for Data Mining and Machine Learning.*