Bioinformatics

First Semester

Course Unit MB 5010: Genomes, Genetics & Molecular Biology – 3 credits

(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:

At the end of this course unit student will be able to

- Describe organization of genomes, genes, intergenic regions in DNA and Mendellian inheritance
- Explain how information is stored in DNA, replicated and converted to functional proteins through mRNA
- Discuss how damaged DNA is repaired
- Discuss DNA variation, epigenetic changes and their consequences
- Give an overview of different forms of RNA

Course content:

Definitions of genome, transcriptome, proteome and introduction to Bioinformatics
Organisation of human genome and other genomes: Sizes of eukaryotic genomes**;** C value paradox**;** Packaging of DNA into chromosomes
Genes: Organisation of the eukaryotic genes; Genome sizes and gene numbers for selected eukaryotes; Human Gene catalogue, gene superfamilies, gene families; Genetic code**;** Pseudogenes and their evolution.
Intergenic regions: Repetitive DNA
Mendelian inheritance with examples: Autosomal inheritance; Recessive inheritance; Sex linked inheritance
Mitochondrial and chloroplast genomes -components and organisation
Prokaryotic genomes: Prokaryotic chromosomes; Genome size and gene numbers; Organisation of prokaryotic genes
Biological information storage in the cell and usage for life: DNA replication; DNA repair- need for and overview of different mechanisms involved; Transcription; Translation
Detecting gene expression: Overview of methodologies to detect expression of selected genes and global gene expression; Demonstration of mRNA expression analysis by RT-PCR; Demonstration of protein expression by Western blotting
Genome imprinting: Functional significance; Imprinting by methylation; Imprinting by changes to the chromatin structure
Genetic variations: Mutations and polymorphisms; Single nucleotide polymorphisms and the HAPMAP project; Copy number variations; Practical use of DNA variations
Other RNAs: rRNA in ribosomal function; species identification and phylogenetic analyses, Overview of non-coding RNAs

References:

Genomes 4 by T A Brown

Journal articles recommended

Evaluation Procedure:

By continuous assessments contributing 30% and end of course assessment contributing 70% of the final mark for the course

Course Unit MB 5020: Biocomputing I – 3 credits

(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:

At the end of this course unit student will be able to

- Identify the skills needed in Bioinformatics
- Execute commands in the Unix environment
- Identify the use of fundamental data structures
- Design and evaluate algorithms
- Design, develop and debug simple programs in Python

Course content:

Bioinformatics skills
Introduction to Unix
Working with remote machines: Connecting to remote machines with SSH
Data structures and Algorithms: Arrays, Stacks and Queues, Linked lists, Trees, Hashing; Recursive algorithms; Computational complexity- Big-O notation
Introduction to Programming: Python : Variables, values and Expressions **(**Values-Booleans, Integers, Floats, Strings**;** Expressions - Numerical, Logical and String operators); Subroutines and Functions (Assigning names, Defining functions, Using Modules); Collections (Sets, Sequences, Mappings, Streams, Collection-Related Expression patterns); Control Statements (Conditionals, Loops**,** Iterations**,** Exception handlers); Regular Expressions; Classes **(**Defining Classes, Class and Method Relationships)
Data management: Relational DBMS, XML

References:

Buffalo. V, Bioinformatics Data Skills, O' Reilly Media, Inc
Mitchell L. M, Bioinformatics Programming using Python, O' Reilly Media, Inc,

Evaluation Procedure:

By continuous assessments contributing 50% and end of course assessment contributing 50% of the final mark for the course

Course Unit MB 5030: Statistical methods in Bioinformatics I – 3 credits

(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:

At the end of this course unit student will be able to

- Visualize and summarize data
- Apply basic probability laws
- Identify basic probability distributions

- Recognize the statistical estimation techniques

- Identify linear models and basic analysis of variance techniques
- Perform basic data manipulation and statistical tests using R and Bioconductor

Course content:

Data display and descriptive statistics: Univariate data display; Descriptive statistics

Probability and probability distributions: Basic probability laws; Important discrete distributions; Important continuous distribution;Brief introduction to zero-inflated distributions; Characteristics of a random variable; Entropy and related concepts

Estimation and statistical inference: Classical estimation methods; Statistical hypothesis testing; Bayesian approach to hypothesis testing and estimation

Linear Models: Definition of linear models; One-way analysis of variance; Two-way analysis of variance; Checking assumptions; Robust tests

A Brief Introduction to R and Bioconductor

References:

Statistical Methods in Bioinformatics: An Introduction (Warren Ewens and Gregory Grant)
Applied Statistics for Bioinformatics using R (Wim P. Krijne)
R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics By Paul Teetor

Evaluation procedure:

By continuous assessments contributing 30% and end of course assessment contributing 70% of the final mark for the course

Second Semester

Course Unit MB 5040: Cell Biology & Biochemistry – 2 credits

(Lectures - 15 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:

At the end of this course unit student will be able to

- Outline the life processes and tree of life
- Describe different biomolecules and their structure, function and metabolism
- Give an overview of cellular components including biological membranes
- Outline how DNA is isolated and identified
- Discuss the function of enzymes, cell communications, signal transduction and transport across membranes in life processes
- Describe how cells proliferate, undergo programmed death
- Describe the molecular and genetic basis of cancer

Course content:

Overview of life and life processes

Tree of Life: Prokaryotes, Eukaryotes, Archae etc

Biomolecules: Carbohydrates, proteins, lipids, and nucleic acids structure, function, synthesis and degradation; DNA extraction and visualization on gel electrophoresis - practical demonstration
Bioenergetics: Enzymes as catalysts; Generation of energy by oxidative phosphorylation; Overview of metabolism and its regulation

Biological membranes and transport across biological membranes: Composition, organization and properties of biological membranes; Overview of different mechanisms used for transport across biological membranes
Cell communications and signal transduction: Neural and chemical communications; Overview of signaling mechanisms via ion channels, cell membrane, cytoplasmic and nuclear receptors
Cell cycle, apoptosis and cancer: Cell division, overview of cell cycle and its regulation
Overview of apoptosis; Visualization of cells and apoptosis; Cancer biology, genetics and genomics

References:
Lippincott's Biochemistry, Dennis R Ferrier, Lippincot, Williams and Wilkins, New York
Journal articles recommended

Evaluation procedure:
By continuous assessments contributing 30% and end of course assessment contributing 70% of the final mark for the course

Course Unit MB 5050: Sequence Analysis & Phylogeny – 3 credits

(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:
At the end of this course unit student will be able to

- Describe methods of DNA sequencing
- Compare sequences by different sequence aligning methods
- Apply different BLAST tools to search sequences over the databases
- Predict structural and functional properties of proteins
- Explain the necessary principles and methods of sequence analysis and phylogeny
- Demonstrate how popular sequence analysis tools work
- Plan experiments; identify appropriate tools/ techniques to analyze sequences; interpret results and draw conclusions
- Define, explain and critically review the basic concepts in phylogenetics and phylogenetic analysis
- Analyze molecular data using different phylogenetic methods and construct phylogenetic trees
- Evaluate and interpret the phylogenetic trees

Course content:

Overview of DNA sequencing: principles behind and procedures used
Basic sequence data analysis
Pairwise sequence alignment
Multiple sequence alignment
Similarity searching over the data bases: BLAST & FASTA
Basic phylogenetic analysis using DNA and protein sequences

Protein sequence analysis: properties, subcellular localization, transmembrane regions, functional properties etc.
Gene prediction: Gene prediction in prokaryotes and eukaryotes
Promoter and regulatory element prediction
Genome wide analysis: Comparative genomics; Gene expression analysis
Phylogeny: Definition and history; Distance computation; Phylogeny construction, analysis of significance of branches and clades; Methods in phylogenetic analysis; Evaluation methods in phylogenetic analysis

References:

Lesk A. M., Introduction to Bioinformatics, Oxford University Press
Jin Xiong, Essential Bioinformatics, Cambridge University Press
Krane D. E., Ramer M. L., Fundamental concepts of Bioinformatics
Felsenstein, J., Inferring Phylogenies, Sinauer Associates is an imprint of Oxford University Press
Lemey, P., Salemi, M. and Vandamme, A., The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny, Cambridge University Press
Hall, B., G., Phylogenetic Trees Made Easy: A How To Manual
Wiley, E., O., and Lieberman, B., S., Phylogenetics: Theory and Practice of Phylogenetic Systematics, Wiley-Blackwell

Evaluation procedure:

By continuous assessments contributing 50% and end of course assessment contributing 50% of the final mark for the course

Course Unit MB 5060: Statistical Methods in Bioinformatics II – 3 credits

(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:

At the end of this course unit student will be able to

- Identify cluster analysis techniques
- Perform principal component (PCA) and factor analysis
- Recognize Poisson processes and Markov chains
- Identify Hidden Markov Models (HMM) and algorithms
- Identify statistical techniques related to micro array data
- Perform PCA on micro array data.

Course content:

Cluster analysis and trees: Introduction to distance measures; Clustering methods; Introduction to principal component analysis (PCA) and factor analysis
Poisson processes and Markov chains: The homogeneous Poisson process and the Poisson distribution; The Poisson and the binomial distributions; The Poisson and the gamma distributions; The pure-birth process; Introduction to finite Markov chains; Transition probabilities and the transition probability matrix; Markov chains with absorbing states; Markov chains with no absorbing states; Stationary distributions; Simple random walk
Hidden Markov models: Introduction to HMM; Algorithms; Applications

Gene expression, microarrays and multiple testing: Introduction; The statistical analysis of microarray data: one gene; Differential expression – multiple genes; Comparing two groups by discriminant analysis; Principal components and microarrays

References:
Statistical Methods in Bioinformatics: An Introduction (Warren Ewens and Gregory Grant)
Applied Statistics for Bioinformatics using R (Wim P. Krijne)
R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics By Paul Teetor

Evaluation procedure:
By continuous assessments contributing 30% and end of course assessment contributing 70% of the final mark for the course

Course Unit MB 5070: Biocomputing II – 2 credits

(Lectures - 15 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:
At the end of this course unit student will be able to

- Explain theoretical basics of basic Machine learning techniques
- Evaluate different machine learning machine learning algorithms
- Select appropriate machine learning methods for data analysis
- Apply machine learning algorithms to biological data
- Identify the use of parallelization and distribution
- Evaluate the use of different parallelization algorithms

Course content:
Machine Learning foundations: Introduction to Bayesian modeling; Bayesian Inference and Induction

Machine Learning Algorithms: Dynamic programming; Gradient descent; EM/GEM Algorithms; Markov – chain Monte-Carlo Methods; Simulated Annealing; Evolutionary and Genetic Algorithms
Neural Networks: Priors and likelihoods; Learning algorithms
Hidden Markov Models: Prior information and initialization; Likelihood and basic algorithms; Learning algorithms
High Performance Computing: Parallel programming algorithms – MapReduce, Hadoop, Apache Spark, Scala

References:
Pierre Baldi and Soren Brunak, Bioinformatics- The Machine Learning Approach, Second edition.

Evaluation procedure:
By continuous assessments contributing 30% and end of course assessment contributing 70% of the final mark for the course

Third Semester

Course Unit MB 5080: Proteomics – 3 credits
(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:
At the end of this course unit student will be able to
- Describe different protein structures and outline in vitro methods used to study them
- Review different online databases available to study protein structure and function
- Use appropriate databases to solve a given problem
- Discuss the impact of structural proteomics on drug design
- Discuss the impact of structural proteomics on enzyme engineering

Course content:

Review of protein primary, secondary structures: Secondary structure prediction; Motifs and domains; Posttranslational modifications of proteins; Protein stability and dynamics
Protein folding
Tertiary structure determination and visualization: X-ray crystallography and NMR spectroscopy; Protein structure database and computer aided visualization using online tools (eg: Chime, Cn3D, RasMol, PyMol, Chimera, Swiss PDB viewer etc.); Structure Prediction and Modeling

Protein structure-function relationship: Protein-ligand interaction; Protein-protein interaction; Protein-DNA interaction
Subcellular localization of proteins
Use of structural information (Drug design, enzyme engineering etc.): Molecular modeling and molecular docking; Online tools for molecular modeling and molecular docking

References:
Lippincott's Biochemistry, Dennis R Ferrier, Lippincott, Williams and Wilkins, New York
Other references provided during the course

Evaluation procedure:
By continuous assessments contributing 30% and end of course assessment contributing 70% of the final mark for the course

Course Unit MB 5090: Genomic Data Science – 2 credits
(Lectures - 15 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:

At the end of this course unit student will be able to

- Explain the use of pipelines and workflows
- Use a workflow management systems to process and analyze data - Galaxy
- Build general Bioinformatics pipelines
- Use different tools to analyze large sets of genomic data
- Evaluate the results of different tools

Course content:

Workflows and Pipelines: General workflows and pipelines; Quality control with Galaxy; DNA sequencing data and RNA Sequencing data analysis
Bioconductor: Gene expression analysis; High throughput data analysis; Genomic data analysis
Tools for data analysis: Alignment and sequence variant detection tools; Transcriptomics
ADAM: Processing genomic data with ADAM

References:
https://www.coursera.org/specializations/genomic-data-science

Evaluation procedure:
By continuous assessments contributing 50% and end of course assessment contributing 50% of the final mark for the course

Course Unit MB 5100: Bioinformatics Algorithms – 3 credits

(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:
At the end of this course unit student will be able to

- Identify the use of different algorithms in Bioinformatics
- Describe principles of algorithm design
- Explain popular algorithms with their computational complexity
- Evaluate an algorithm
- Design efficient algorithms

Course content:

Exhaustive search
Greedy algorithms
Dynamic programming algorithms
Multiple alignments
Gene prediction
Pattern matching
Clustering and Trees
Hidden Markov Models

References:

Jones, Neil, and Pavel Pevzner. *An Introduction to Bioinformatics Algorithms*. Cambridge, MA: MIT Press, ISBN: 9780262101066.
Durbin, Richard, Sean Eddy, Anders Krogh, and Graeme Mitchison. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge, UK: Cambridge University Press, ISBN: 9780521629713.
Duda, Richard, Peter Hart, and David Stork. *Pattern Classification*. New York, NY: Wiley-Interscience, ISBN: 9780471056690.

Evaluation procedure:
By continuous assessments contributing 30% and end of course assessment contributing 70% of the final mark for the course

Course Unit MB 5110: Guided Independent Study – 5 credits

Learning outcomes:
At the end of this course unit student will be able to

- Develop online self -learning skills
- Identify a problem requiring bioinformatics solution
- Apply the knowledge gained during the programme to solve the problem identified
- Present the findings both orally and in writing

Course content:
Complete one of the recommended courses in Bioinformatics in a MOOC (e.g. Coursera or EdX)
Carryout a mini project in bioinformatics based on what has been learnt during the programme

Evaluation procedure:
By initial proposal writing and evidence of online learning (basically through a presentation based on the MOOC) contributing 25% and mini project evaluation and final report assessment contributing 75% of the final mark of the course.

Fourth Semester

Course Unit MB 6010: Systems Biology and Pathway Analysis – 3 credits

(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 30 h)

Learning outcomes:
At the end of this course unit student will be able to

- Define, describe concepts in systems biology and pathway analysis.
- Demonstrate using popular Data analysis and Visualization tools
- Interpret the results of network analysis data
- Identify suitable methods for data analysis
- Recognize the challenges faced in the field of systems Biology

Course content:

Experimental methods in systems biology

Cellular Networks: Network representation, Metabolic Networks, Protein Networks, Gene Regulatory Networks
Network building and analysis: Networks and Graphs; Properties of biological networks; Types of biological networks
Modelling methods for systems biology: Building Models; Linear and Nonlinear models
Systems biology tools and applications
Differential Equation Models
Integrated analysis in systems biology- Topology to Function

References:

Herbert M. Sauro, Systems Biology: Introduction to Pathway Modelling, Ambrosius Publishing and Future Skills Software
Coursera courses: (Introduction to Systems Biology and Networks in Systems Biology)

Evaluation procedure:
By continuous assessments contributing 40% and end of course assessment contributing 60% of the final mark for the course

Course Unit MB 6020: Special topics in Bioinformatics – 4 credits
(Lectures - 30 h, Practical classes/ Assignments/ Tutorials – 60 h)

Learning outcomes:
At the end of this unit students will be able to

- Identify novel applications
- Discuss upcoming techniques relevant to the field of bioinformatics
- Explain the fundamental concepts of selected special topics in bioinformatics
- Demonstrate accurate and efficient use of specific bioinformatics techniques relevant to data generated through novel applications
- Work on practical examples of the novel techniques
- Apply novel technique to solve given set of problems

Course content:
Next Generation Sequence analysis (Genome assembly, Genome annotation, RNAseq Etc.);
Other current topics

References:
Online reference materials provided during the course

Evaluation procedure:
By continuous assessments contributing 50% and end of course assessment contributing 50% of the final mark for the course

Course Unit MB 6030: Research Methodology – 3 credits
(Lectures - 15 h, Practical classes/ Assignments/ Tutorials – 60 h)

Learning outcomes:
At the end of this course unit student will be able to

- Discuss the terms research, scientific method and hypothesis
- Define a problem in an area of interest and produce a well-developed research proposal
- Select and use internet resources and software tools to support a research project
- Select and use an appropriate methodology to conduct the research concerned
- Interpret the results obtained in a meaningful way
- Communicate findings in a way suitable to research/ academic audience

Course content:
History and etymology of research; The scientific method; Research designs: Qualitative and quantitative approaches; Searching the literature; Data collection; Analyzing results; Technical communication; Research ethics; Specific issues pertaining to genetic and genomic research involving humans

References:
Social Science Research: Principles, Methods, and Practices. Anol Bhattacherjee. CreateSpace Independent Publishing Platform.
Research Methodology: Methods and Techniques. C. R. Kothari. New Age International Publishers, India.
Research Methods for Science. Michael P. Marder. Cambridge University Press, UK.
Research Design: Qualitative, Quantitative and Mixed Methods Approaches. John W. Creswell. SAGE Publications, India.

Evaluation procedure:
By continuous assessments contributing 60% and end of course assessment contributing 40% of the final mark for the course

Course Unit MB 6040: Research Seminar- 3 credits

Learning outcomes:
At the end of this course unit student will be able to

- Identify the most promising research papers in a given area of bioinformatics
- Describe major criteria used to evaluate scientific research papers
- Demonstrate the creative independence through presenting the papers/research and discussing research papers
- Critically present scientific literature

Course Content:
Review presentations/ group discussions on recent advanced research publications on relevant fields

Evaluation procedure:
Assessments of seminar presentations throughout the course contributing 100% of the final marks

Fifth & Sixth Semester

Course Unit MB 6050: Research project – 15 credits

Learning outcomes:

At the end of this course unit student will be able to

- Independently design a research project
- Implement the designed project
- Analyse results
- Discuss the findings
- Defend the study and its conclusions

Course content:

Literature review & project proposal presentation, project implementation & data analysis; Writing and submission of the Dissertation for examination, defending the Dissertation

Evaluation procedure:

Literature review and project proposal presentation, final presentation, Dissertation examination, Viva Voce contributing to 20%, 20%, 50%, 10% of the final mark respectively