

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

GENOMICS AND PROTEOMICS

CODE: 19BI/PC/GP24

CREDITS : 4

L T P : 3 0 2

TOTAL CONTACT HOURS: 65

OBJECTIVES OF THE COURSE

- To provide an insight into the complete genome sequences of a few organisms as well as the Human genome through Comparative and Functional genomics
- To acquaint knowledge on functional genomics techniques such as microarrays, EST, SAGE and interpret data obtained through high throughput expression studies
- To develop an understanding of the entire protein complement of a cell through analytical approaches, Data mining and other software tools

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Gain an insight of the basic and advanced concepts and applications of sequencing technologies
- Understand the mechanisms of genomics and proteomics and exploit the same in the growing field of omics
- Apply functional genomics techniques to analyse data for biological system
- Implement techniques and database search to analyze complex protein samples
- Analyze the proteomic interactions in complex diseases

Unit 1 (13 Hours)

Genomics

- 1.1 Understanding a Genome sequence, Locating the genes in a Genome Sequence, Gene location by Sequence Inspection, Experimental Techniques for Gene Location, Determining the Functions of Individual Genes
- 1.2 Genome Sequencing technologies - Conventional Sequencing techniques (Maxam Gilbert and Sanger Sequencing), Whole Genome Shotgun Sequencing, Genome assembly, Genome annotation and Genome databases
- 1.3 Rates and patterns of Nucleotide substitution, Molecular Clocks, Local Clocks, Computer Analysis of a Gene Function, Assigning Gene Function by Experimental Analysis

Unit 2 (12 Hours)

Comparative Genomics

- 2.1 Comparative Genomics - Genome Sequencing Projects, Variations at the Level of individual Nucleotides, Duplications, Comparisons at the Chromosome Level: Synteny, Genomes of Chimpanzees and Humans

- 2.2 Phylogenetic Analysis - Relationship of Phylogenetic Analysis to Sequence Alignment, Genome Complexity and Phylogenetic Analysis, Maximum Parsimony Method, Distance Methods, Gene Prediction by ORF analysis
- 2.3 Gene mapping - pedigree analysis, Application of DNA markers - RFLPs, SNPs, Physical mapping - Restriction mapping, Fluorescent *in situ* hybridization, Assessing genomic variations

Unit 3 **(15 Hours)**

Functional Genomics

- 3.1 Transcriptomes and analysis - Micro Array technology, SAGE, Applications of Microarrays In Medicine, Databases – GEO, MAML
- 3.2 ESTs Generation, EST Clustering and Assembly, EST databases (DB-EST, UNIGene)
- 3.3 KEGG and Metabolic Pathways, Regulatory Networks, Sequence based and structure-based approaches to assign gene functions, Role of databases in function assignment, Structural changes in sequences by the influence of polymorphisms (dbSNPs)

Unit 4 **(13 Hours)**

Proteomics

- 4.1 Introduction to Proteomics - Proteins structure, Organization of protein structure, structural conformation of proteins, three dimensional structures of proteins
- 4.2 Analytical tools in Proteomics - 1D and 2D-gel electrophoresis, Mass Spectrometry - ESI, MALDI etc., Software for Matching MS Data with Specific Protein Sequences, Peptide sequencing by tandem mass spectrometry
- 4.3 Preparative IEF, HPLC, Tandem LC/ MS-MS, Protein Digestion Techniques

Unit 5 **(12 Hours)**

Application of Proteomics

- 5.1 Proteomic interactions - Yeast Two-Hybrid, Mammalian Screen Methods and Co-Immuno Precipitation techniques
- 5.2 Protein-Protein Interactions and Protein Complexes, Databases and proteomic tools
- 5.3 Protein Interaction Networks and Protein Pathways, Mapping Protein modifications

BOOKS FOR STUDY

Arthur Lesk M. *Introduction to Genomics*. New York: Oxford university press, Third edition, 2017.

Brown, T. A. *Genomes -3*. USA: John Wiley and Sons inc., 2006.

Leland Hartwell, Michael L. Goldberg and Janice Fischer. *Genetics: From Genes to Genomes*. USA:McGraw-Hill Publishing Company. 2018

Daniel C. Leibler. *Introduction to Proteomics: Tools for New Biology*. USA: Humana Press, 2002.

Srivastava Sudhir. *Informatics in Proteomics*. USA: Taylor & Francis Group, 2005.

BOOKS FOR REFERENCE

Brown P. O and Botstein D. *Exploring the new world of the genome with DNA microarrays*. USA: Nat. Genet, 1999.

Collado Vides Julio and Ralf Hofstadter. *Gene Regulation and Metabolism – Post Genomic Computational Approaches*. India: Ane Books, 2004.

Dale, Jeremy W and Malcolm von Schantz. *From Genes to Genomes – Concepts and Applications of DNA Technology*. USA: John Wiley and Sons, 2012.

Arthur Lesk M. *Introduction to Genomics*. New York: Oxford university press, 2008.

Griffiths, A.J.F, Miller, J.H, Suzuki, D.T. Lewontin, R. C. and Gelbart, W.M. *An Introduction to Genetic Analysis*. USA: W.H. Freeman, 1996.

Hunt Stephen P and Livesey Fredrick J. *Functional Genomics -A Practical Approach*. Great Britain: Oxford University Press, 2000.

Golemis and Erica. *Protein-Protein Interaction*. USA: CSHL, 2005.

Lesk Arthur M. *Introduction to Protein Science: Architecture, Function and Genomics*. New York: Oxford university press, 2016.

Mount David W. *Bioinformatics: Sequence and Genome Analysis*, USA: Cold Spring Harbor Lab., 2005.

Pennington S and M. J. Dunn. *Proteomics: From Proteins Sequence to Function*. Germany: Springer Publications, 2001.

Palzkill and Timothy. *Proteomics*. USA: Kluwer Academic Publishers, 2013.

JOURNALS

Genomics, Proteomics & Bioinformatics

Journal of Data Mining in Genomics & Proteomics

Human Genomics and Proteomics

Journal of Proteomics and Genomics

WEB RESOURCES

<http://www.oncolink.org/resources/article.cfm?id=326>

<http://www.nature.com/nature/journal/v422/n6928/full/nature01510.html>

<http://proteomics.cancer.gov/whatisproteomics>

<http://www.isaaa.org/resources/publications/pocketk/15/default.asp>

PATTERN OF ASSESSMENT

Continuous Assessment Test: Total Marks: 50 Duration: 90 minutes

Theory:

Section A – 15 x 1 = 15 Marks (All questions to be answered)

Section B – 2 x 5 = 10 Marks (2 out of 4 to be answered)

Practical:

Section C – 5 x 5 = 25 Marks

Other Components: Total Marks: 50

Assignment/Test/Seminars

End Semester Examination: Total Marks: 100 Duration: 3 hours

Theory:

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 2 x 15 = 30 Marks (2 out of 4 to be answered)

Practical:

Section C – 5 x 10 = 50 Marks

Questions comprising the following:

Genome databases of plants, animals and pathogens

Clusters of Orthologous Groups (COGs)

Gene Prediction by ORF analysis, Gen scan, UCSC Genome Browser

DNA markers - dbSNP, Restriction mapping

Transcriptomes analysis - Micro Array data analysis, GEO

EST Clustering databases - DBEST, UNIGene

Metabolic pathway database – KEGG, PharmGKB

Protein classification and structure analysis - CATH, SCOP

Protein Motif and Domain search - PROSITE, PDBeMotif

Protein - protein interaction analysis - DIP, STRING, BIND

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

MOLECULAR BIOLOGY

CODE: 19BI/PC/MB24
: 4

CREDITS

L T P : 4 1 0

TOTAL TEACHING HOURS: 65

OBJECTIVES OF THE COURSE

- To understand the general principles of gene organization and expression
- To explore the various levels of gene regulation and protein function
- To analyse the various genetic and molecular changes occur in a normal cell

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Represent and illustrate the structural organization of genes and the control of gene expression
- Explore the prokaryotic and eukaryotic protein synthesis mechanism
- Conceptualize mechanisms of signal transduction, cell cycle and cell death
- Link the concepts of cell and molecular biology to a better understanding of diseases, including cancer

Unit 1 (15
Hours)

Structure and Organisation of Genes and Chromosomes

- 1.1 DNA-Structure and Conformations, Histones and Non-Histones, **Chromosomes - Structure and Function of Chromosomes**
- 1.2 Cell division - Mitosis and meiosis, Cell cycle regulation, Check points
- 1.3 Organisation of Genomes –Coding Sequences, Repetitive Sequences, transposons

Unit 2 (15
Hours)

Replication and Transcription

- 2.1 DNA replication, repair and recombination, DNA damage and repair mechanisms in prokaryotes and eukaryotes, Mutations
- 2.2 Transcription: Eukaryotes and Prokaryotes, Genetic code, Transcriptional Control by Regulatory Proteins, Steroid Hormone Receptors - Heat Shock Genes- Homeotic Genes
- 2.3 Mechanisms Modifying Transcriptional Control – DNA Methylation, Histone Modification, Post Transcriptional Regulation

Unit 3 (12
Hours)

Translation

- 3.1 RNA- Types, structure and functions, Ribosomes – Structure and Assembly
- 3.2 Translational Regulation - Regulation of gene expression in Prokaryotes (Operon) and Eukaryotes, Gene Silencing
- 3.3 Genetic Control of Vertebrate Immune System

Unit 4 (10
Hours)

Organelle Genome

- 4.1 Mitochondrion genome – Organisation and Function
- 4.2 Chloroplast genome – Organisation and Function
- 4.3 Transcription and Translation in Mitochondria

Unit 5 (13
Hours)

Cell Signalling and Cancer

- 5.1 Cell signalling – Signalling molecules, Receptors - Hormones receptors, cell surface receptor, signal transduction pathways, regulation of signalling pathways
- 5.2 Cancer Biology- Characteristics of Cancer, Genetic basis of cancers, Proto-oncogene, Oncogenes, Tumor Suppressor Genes, Tumor Metastasis
- 5.3 Oncogenesis - Cancer Immunotherapy, Regulation of Cell Death, Apoptosis

BOOKS FOR STUDY

Harvey Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, Anthony Bretscher, Hidde Ploegh. *Molecular Cell Biology*. USA: W. H. Freeman, Eighth edition, 2016.

Wolfe, Stephen L. *Molecular and Cellular Biology*. USA: Wadsworth, 2005.

Watson, James, D. *Molecular Biology of the Gene*. USA : The Benjamin Cummings Publishing Company, 2007.

BOOKS FOR REFERENCE

Cooper, Geoffrey M. and Robert E. Hausman. *The Cell, A Molecular Approach*. USA: Sinauer Associates, 2004.

Harvey Lodish, Arnold Berk, Chris A. Kaiser, Monty Krieger, Matthew P. Scott, Anthony Bretscher, Hidde Ploegh and Paul Matsudaira. *Molecular Cell Biology*. USA: W.H.freeman, 2008.

Watson, James, D. *Molecular Biology of the Gene*. UK: Pearson, Seventh edition, 2017.

Darnell, James, Harvey Lodish and David Baltimore. *Molecular and Cell Biology*, Scientific American Books, USA: W.H. Freeman, 2004.

Karp and Gerald. *Cell and Molecular Biology- Concepts and Experiments*, USA : John Wiley, 2013.

Lewin and Benjamin. *Genes IX*, UK :Oxford University Press, 2009.

Roitte, Ivan M., Brostoff, Jonathan and Male, David K. *Immunology*. Philadelphia: J.B. Lippincott, 1990.

Purvis, William K, David Sadava, Craig Heller and Gordan H. Orians. *Life: The Science of Biology*. USA : Sinauer, 2004.

JOURNALS

Journal of Molecular Biology

Molecular Biology

Journal of Genetics and Genomics

BMC Cell Biology

WEB SOURCES

www.cellbio.com

www.molbiolcell.org

www.sciencedirect.com

<http://www.nature.com/scitable/topic/cell-biology-13906536>

http://www.biology.arizona.edu/cell_bio/cell_bio.html

<http://ghr.nlm.nih.gov/>

PATTERN OF ASSESSMENT

Continuous Assessment Test: Total Marks: 50 Duration: 90 minutes

Section A – 10 x 1 = 10 Marks (All questions to be answered)

Section B – 2 x 10 = 20 Marks (2 out of 4 to be answered)

Section C – 1x 20 = 20 Marks (1 out of 2 to be answered)

Other Components: Total Marks: 50

Assignment/Test/Seminars

End Semester Examination: Total Marks: 100 Duration: 3 hours

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 4 x 10 = 40 Marks (4 out of 7 to be answered)

Section C – 2 x 20 = 40 Marks (2 out of 4 to be answered)

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

MOLECULAR BIOLOGY PRACTICAL

CODE: 19BI/PC/P122

CREDITS : 2

L T P : 0 0 3

TOTAL HOURS : 39

OBJECTIVE OF THE COURSE:

- To identify subcellular structures, organelles and understand their functions
- To provide practical experience of the various techniques involved in Molecular Biology and Biochemistry
- To perform a range of molecular techniques used for the isolation, estimation, purification of biomolecules

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Utilize laboratory skills to enhance understanding of cell structure and function while participating in a group environment
- Develop responsible conduct of laboratory skills appropriate to the field of cell and molecular biology
- Apply the molecular biology techniques to biotechnological approaches

Unit 1 (8 Hours)

- 1.1 Cell Fraction and Extraction of cell organelles
- 1.2 Isolation of Sub-Cellular Organelles and Particles –Mitochondria and Chloroplast

Unit 2 (10 Hours)

- 2.1 Extraction of DNA from Onion, Extraction of RNA from Yeast
- 2.2 Estimation of DNA and RNA
- 2.3 Estimation of Proteins by Lowry's Method

Unit 3 (7 Hours)

- 3.1 Estimation of Mitochondria by Assessing The Marker Enzyme
- 3.2 Denaturing Proteins and Identification of Amino Acids by Thin Layer Chromatography

Unit 4 (7 Hours)

- 4.1 Isolation of Plasmid DNA (Demo)
- 4.2 Amplification of DNA by PCR

Unit 5 (7 Hours)

- 5.1 Electrophoretic Techniques: Agarose Gel Electrophoresis, SDS PAGE (Demo)

5.2 Southern Blotting (Demo)

BOOKS FOR REFERENCE:

Wilson, K; Walker, J. *Principles and techniques of Biochemistry and Molecular Biology*. USA: Cold Spring Harbor Laboratory Press, 2010.

Sambrook, J; Russel, DW. *Molecular Cloning*. USA: Cold Spring Harbor Laboratory Press, 2001.

Sadasivam, S. and Manickam, A. *Biochemical Methods*. India: New Age International, 2009.

Wilson, K; Walker, J. *Principles and techniques of Biochemistry and Molecular Biology*. USA: Cold Spring Harbor Laboratory Press, Eighth edition, 2010.

Swati Agarwal, Suphiya Khan. *Advanced Lab Practices in Biochemistry & Molecular Biology*. India: I K International Publishing House, 2018.

PATTERN OF ASSESSMENT

End Semester Examination **Total Marks: 100** **Duration: 3 hours**

Spotters 4 in number each carrying 5 marks totalling 20 marks

Any two experiments each carrying 30 marks each—10 marks for procedure, 10 marks for the result and 10 marks for the conduct of the experiment

Viva - 10 marks

Record - 10 marks

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2024)

RESEARCH METHODOLOGY

CODE: 19BI/PC/RM24
: 4

CREDITS

L T P : 4 1 0

TOTAL TEACHING HOURS :

65

OBJECTIVES OF THE COURSE

- To describe and express the role and importance of research in basic and applied sciences
- To facilitate writing of research proposals / projects and apply for grants in the field of bioinformatics
- To understand the analytical tests to be applied for research

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Better understanding of the research methods
- Design an action plan of research
- Acquire skills of writing a research manuscript
- Application of statistical study in research
- Understand the ethics in writing research work

Unit 1
Hours)

(15

Types of Data and research problem identification

- 1.1 Data Collection, Sampling. Sources of Data Primary, Secondary and Tertiary Sources Classification and Presentation of Data
- 1.2 Documents, Types of Documents, Archives, Chronologies
- 1.3 Definition of Research and Research Methodology. Principles and Practice of Research. Exploring the Broad Area – Using the Library and Online Resources. Identifying The Research Problem

Unit 2
Hours)

(15

Scientific Communication

- 2.1 Literature Review - Its Relevance and Importance in Directing Research. Citations – Types Of Citations, Bibliography and End Matters, Editing and Proof Reading
- 2.2 Action Plan, Design and Pilot Study Undertaking a Research Project, Writing a Research grant Proposal, writing papers and posters, Format of thesis
- 2.3 Paper critiquing- the Purpose and the Methodology of Paper Critiquing

Unit 3 (10
Hours)

Writing well

- 3.1 Writing for non- native audiences, usage of simple sentences, untangle long noun phrases, make complete sentences.
- 3.2 Use of punctuations- comma, colon, semicolon, dash and periods.
- 3.3 Creating non-textual information- acquiring, processing and printing illustrations. Concepts of mind maps.

Unit 4 (12
Hours)

Bioethics

- 4.1 Introduction. Intellectual Property Rights (IPR) and Patents, TRIPS
- 4.2 Case studies on Patents (Basmati, Turmeric and Neem), ethics in science practicals
- 4.3 Plagiarism and Common Errors in Scientific Writing. Misconduct in science

Unit 5 (13
Hours)

Tools for research

- 5.1 Use of Encyclopaedias, Research Guides, Handbook etc., Academic Databases for Computer Science Discipline.
- 5.2 Use of tools / techniques for Research: methods to search required information effectively, Reference Management Software like Zotero/ Mendeley,
- 5.3 Software for paper formatting like LaTeX/MS Office, Software for detection of Plagiarism

BOOKS FOR STUDY

Gopalan, R. *Thesis Writing*. India: Vijay Nicole Imprints Private Limited, 2005.

Gurumani, N. *Research Methodology for Biological Sciences*. India MJ Publishers, 2010.

BOOKS FOR REFERENCE

Pence, G.E. *Classic Cases in Medical Ethics*. India: McGraw-Hill, 2004.

Kothari C R. *Research Methodology, Methods and Techniques*. India: Wishwa Prakashan, 2009.

JOURNALS

The Journal of Communication

International Association for Media And Communication Research

Indian Journal of Science Communication

WEB RESOURCES

<http://www.palgrave.com/studentstudyskills/page/choosing-appropriate-researchmethodologies/>

<https://explorable.com/research-methodology>

PATTERN OF ASSESSMENT

Continuous Assessment Test: Total Marks: 50 Duration: 90 minutes

Section A – 10 x 1 = 10 Marks (All questions to be answered)

Section B – 2 x 10 = 20 Marks (2 out of 4 to be answered)

Section C – 1x 20 = 20 Marks (1 out of 2 to be answered)

Other Components: Total Marks: 50

Assignment/Test/Seminars

End Semester Examination: Total Marks: 100 Duration: 3 hours

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 4 x 10 = 40 Marks (4 out of 7 to be answered)

Section C – 2 x 20 = 40 Marks (2 out of 4 to be answered)

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

SOFT SKILLS

CODE:19BI/PK/SS22
: 2

CREDITS

L T P : 2 0 0

TOTAL TEACHING HOURS : 26

OBJECTIVES OF THE COURSE

- To empower and create opportunities for self-development
- To instill confidence and face challenges
- **To** develop self-motivation, raised aspirations and belief in one's own abilities

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Connect and work with others to achieve a set task
- **Assessing** the requirements of a task, identifying the strengths within the team
- **Create** awareness of one's place and role within a community through volunteering and conservation opportunities
- **Responsible** for one's self, learning self-reliance and independence
- **Develop Employability skills**, time and resource management, conflict resolution, teaching and mentoring others

Unit 1

Hours)

(6

Behavioural Traits

- 1.1 Self Awareness
- 1.2 Communication Skills – Verbal and Non Verbal
- 1.3 Leadership Qualities
- 1.4 Etiquette and Mannerisms
- 1.5 Experiential Learning – Based on Activities

Unit 2

Hours)

(5

Team Work

- 2.1 Interpersonal Skills
- 2.2 People Management
- 2.3 Creative Thinking
- 2.4 Critical Thinking

2.5 Experiential Learning – Based on Activities

Unit 3 (5 **Hours)**

Time Management

- 3.1 Importance of Time Management
- 3.2 Planning and Prioritizing
- 3.3 Organizing skills
- 3.4 Action Plan
- 3.5 Experiential Learning – Based on Activities

Unit 4 (5 **Hours)**

Conflict Resolution

- 4.1 Reasons for Conflict
- 4.2 Consequences of Conflict
- 4.3 Managing Emotions
- 4.4 Methods of Resolving Conflicts
- 4.5 Experiential Learning – Based on Activities

Unit 5 (5 **Hours)**

Career Mapping

- 5.1 Goal Setting
- 5.2 Career Planning
- 5.3 Resume Writing
- 5.4 Handling Interviews
- 5.5 Experiential Learning – Based on Activities

BOOKS FOR REFERENCE

Khera, Shiv. *You Can Win*. India: Macmillan India Ltd, 2002.

Mishra, Rajiv K. *Personality Development: Transform Yourself*. India: Rupa and Co, 2004.

Newstrom, John W. and Scannell, Edward E. *Games Trainers Play: Experiential Learning*. India: Tata McGraw Hill, 1980.

Anjali Ghanekar and Ghanekar. *Communication and Soft Skill Development*. India: Everest Publishers, 2016

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

ALGORITHMS FOR BIOINFORMATICS

CODE: 19BI/PC/AL34

CREDITS : 4

L T P : 4 1 0

TOTAL TEACHING HOURS : 65

OBJECTIVES OF THE COURSE

- To develop a quantitative understanding of how living things are built
- To raise the awareness of the impact of algorithms on the efficiency of the system
- To develop skills to analyse algorithms related to Bioinformatics

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Know the main problems in the field of bioinformatics and computational molecular biology
- Analyse the correctness of algorithms and how it works
- Describe the divide-and-conquer paradigm and explain when an algorithmic design situation calls for it.
- Apply the algorithms and design techniques to solve problems
- Analyse the complexities of various problems in different domains

Unit 1 (10 Hours)

Introduction

- 1.1 Algorithms and Complexity. Definition, Biological Algorithms versus Computer Algorithms, Fast versus Slow Algorithms Big-O Notation
- 1.2 Algorithm Design Techniques Exhaustive Search Branch-and-Bound Algorithms Greedy Algorithms
- 1.3 Dynamic Programming Divide-and-Conquer Algorithms Machine Learning Randomized Algorithms

Unit 2 (10 Hours)

Restriction Mapping

- 2.1 Impractical Restriction Mapping Algorithms, Practical Restriction Mapping Algorithm
- 2.2 Regulatory Motifs in DNA Sequences Profiles: The Motif Finding Problem Search Trees
- 2.3 Finding a Median String. String matching algorithm

Unit 3 (15 Hours)

Sequence Alignment

- 3.1 Longest Common Subsequences - Global Sequence Alignment- Local Sequence Alignment
- 3.2 Graph Algorithms- Graphs and Genetics- DNA Sequencing Shortest Superstring Problem
- 3.3 DNA Arrays as an Alternative Sequencing Technique. Sequencing by Hybridization

Unit 4 (15 Hours)

Clustering and Evolutionary Trees

- 4.1 Gene Expression Analysis. Hierarchical Clustering -k-Means Clustering- Clustering and Corrupted Cliques
- 4.2 Evolutionary Trees - Distance-Based Tree Reconstruction Reconstructing Trees from Additive Matrices. Evolutionary Trees and Hierarchical Clustering Character-Based Tree Reconstruction
- 4.3 Secondary Structure Prediction methods, Artificial Neural Networks

Unit 5 (15 Hours)

Pattern Matching

- 5.1 Combinatorial Pattern Matching. - Identical, Similar and Distant Repeats Finding methods. Exact Pattern Matching
- 5.2 Keyword Trees and Suffix Trees. Heuristic Similarity Search Algorithms
- 5.3 BLAST: Comparing a Sequence against a Database

BOOKS FOR STUDY

Neil C Jones and Pavel A. Pevzner. *An Introduction to Bioinformatics Algorithms*. USA: MIT press, 2011.

Pavel A. Pevzner. *Computational Molecular Biology- An algorithmic approach*. USA: MIT press, 2004.

BOOKS FOR REFERENCE

Alfred V. Aho, John E. Hopcroft and Jefferey D.Ullman. *Data Structures and Algorithms*. London: Addison Wesley,1983.

Clark, John and Derek Allan Holton. *A First Look at Graph Theory*. Singapore: Singapore Publishers,1995.

Horowitz, Ellis, and Sartag Sahni. *Fundamentals of Computer Algorithms*. New Delhi: Galgotia Publications,1994.

Jeffrey J. McConnell. *Analysis of Algorithm*. New Delhi: Narosa Publishing House, 2002.

Thomas H. Cormen, Charles E. Leiserson and Ronald L. Rivest. *Introduction to Algorithms*. New Delhi: Prentice Hall of India, 1990.

JOURNALS

Algorithms for Molecular Biology
Journal of Computational Intelligence in Bioinformatics
International Journal of Bioinformatics Research and Applications
Developments in Bioinformatics Algorithms

WEB RESOURCES

http://www.comp.nus.edu.sg/~ksung/algo_in_bioinfo/
<http://bioinformaticsalgorithms.com/>
http://bix.ucsd.edu/bioalgorithms/presentations/Ch08_GraphsDNAseq.pdf
<http://www.ait-budapest.com/advanced-algorithms-for-bioinformatics>

PATTERN OF ASSESSMENT

Continuous Assessment: **Total Marks: 50** **Duration: 90 minutes**

Section A: $5 \times 10 = 50$ (7 questions to be set)

Other Components:

Seminars/Assignments/Problem solving

End Semester Examination: **Total Marks: 100** **Duration: 3 Hours**

Section A: $10 \times 10 = 100$ (12 questions to be set)

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

MOLECULAR MODELING AND COMPUTER AIDED DRUG DESIGN

CODE: 19BI/PC/MC34

CREDITS : 4

L T P : 4 1 0

TOTAL TEACHING HOURS : 65

OBJECTIVES OF THE COURSE

- To provide clear concepts on bond angle, bond stretching, bond distance and role on different types of bonds in interactions
- To provide a theoretical background to the various methods of molecular modelling, mechanics and interaction
- To develop and understand the mechanism of drug design using computers

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Perform protein structure prediction and understand energy simulation methods and its importance in drug action
- Gain insight on the molecular dynamics and Monte Carlo simulation methods
- Understand the concept of molecular interactions and QSAR studies
- Apply the knowledge gained to find new targets and design drug to treat diseases
- Learn the concept of drug development

Unit 1 (13 Hours)

Molecular Mechanics

- 1.1 Concepts in Molecular Modeling - Molecular Representations, Coordinate Systems, Potential Energy Surfaces
- 1.2 Molecular Mechanics, Force fields - Bond Length, Bond Angle and Torsion Angle potential
- 1.3 Non- bonded interactions-Van der Waals and Electrostatic Potential – Hydrogen Bonding Terms

Unit 2 (12 Hours)

Energy Minimization Methods

- 2.1 Energy Minimization- Derivative and Non-derivative Energy Minimization Methods
- 2.2 Calculation of Simple Thermodynamic Properties, Computer Simulation, Boundaries, Monitoring the Equilibration, Long Range Forces
- 2.3 Analyzing the Results of Simulation and Estimating Errors

Unit 3 (15 Hours)

Molecular Dynamics and Monte Carlo Simulations

- 3.1 Molecular Dynamics Using Simple Model, Molecular Dynamics with Continuous Potentials
- 3.2 Molecular Dynamics at Constant Temperature and Pressure, Incorporating Solvent effects into Molecular Dynamics, Conformational Changes From Molecular Dynamics Simulation
- 3.3 Monte Carlo Simulation of Molecules, Calculation of Chemical Potential- Simulating Phase Equilibria by Gibbs Ensemble Monte Carlo Method

Unit 4 (12 Hours)

Molecular Modeling and Structure Analysis

- 4.1 Protein Structure prediction - Secondary Structure Prediction, Homology modeling
- 4.2 Threading and *ab initio* method, Tools for Structure prediction; Protein structural visualization; Geometry optimization and Loop refinement
- 4.3 Structure validation tools - Ramachandran Plot.

Unit 5 (13 Hours)

Molecular Docking

- 5.1 Molecular Docking -Structure Based Drug Design - Target Discovery and Validation, Active Site Prediction, Lead identification and Optimization, De Novo Drug Design
- 5.2 Molecular Descriptors - QSAR, 3D Pharmacophore identification and mapping

5.3 Ligand-based drug designing approaches: Lead Designing, High Throughput Screening (HTS), Chemical libraries, ADME prediction

BOOKS FOR STUDY

N. Claude Cohen. *Guidebook on Molecular Modelling In Drug Design*. California: Academic Press, 2006.

Andrew R. Leach. *Molecular Modeling: Principles and Applications*. USA: Prentice Hall, 2007.

Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to applications*. USA: Academic Press, 2002.

Claudio N. Cavasotto. *In Silico Drug Discovery and Design: Theory, Methods, Challenges, and Applications*. USA: Taylor & Francis Group, 2017

BOOKS FOR REFERENCE

Charifson P S. *Practical Application of Computer Aided Drug Design*. New York: Dekker, 1997

Alan Hinchliffe. *Molecular Modelling for Beginners*. USA: John Wiley & Sons, 2008

Sivasamy Ramasamy. *Molecular Modeling*. India: LAMBERT Academic Publishing, 2015

Luca Monticelli, Emppu Salonen. *Biomolecular Simulations: Methods and Protocols*. USA: Humana Press, 2016.

JOURNALS

Journal of Molecular Modeling

Journal of Molecular Graphics and Modelling

Journal of Computer-Aided Molecular Design

Current Computer Aided-Drug Design

WEB RESOURCES

<http://accessengineeringlibrary.com/browse/computer-aided-drug-design-and-delivery-systems>

<http://www.southernresearch.org/life-sciences/lead-discovery-and-optimization/medicinal-chemistry/computational-chemistry>

<http://www.ch.ic.ac.uk/local/organic/mod/>

http://www.chemcomp.com/MOE-Molecular_Modeling_and_Simulations.htm

PATTERN OF ASSESSMENT

Continuous Assessment Test: Total Marks: 50 Duration: 90 minutes

Section A – 10 x 1 = 10 Marks (All questions to be answered)

Section B – 2 x 10 = 20 Marks (2 out of 4 to be answered)

Section C – 1x 20 = 20 Marks (1 out of 2 to be answered)

Other Components: **Total Marks: 50**
Assignment/Test/Seminars

End Semester Examination: **Total Marks: 100** **Duration: 3 hours**

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 4 x 10 = 40 Marks (4 out of 7 to be answered)

Section C – 2 x 20 = 40 Marks (2 out of 4 to be answered)

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

PYTHON AND R PROGRAMMING PRACTICAL

CODE: 19BI/PC/P232

CREDITS : 2

L T P : 0 0 3

TOTAL HOURS: 39

OBJECTIVE OF THE COURSE

- Demonstrate how to locate and download files for data analysis involving genes and medicine
- Select datasets, open files and pre-process data using Python and R language
- Develop and write R scripts to replace missing values, normalize data, discretize data, and sample data

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Relate the necessity for programming in biology
- Handling biological concepts with Python and R scripts
- Apply programming to analyse genomic sequences
- Gain efficient programming skills
- Perform Microarray data analysis using R language

Unit 1 (7 Hours)

Basics of Python

1.1 Counting letters in DNA strings

1.2 Write a Python program related to Bioinformatics transcription

Unit 2 (8 Hours)

Biopython

2.1 Biopython- using Bioseq –Sequence reading and writing

2.2 Biopython using Bio.Genbank – reading entries

Unit 3 (8 Hours)

Basics of R

3.1 Creating vectors and dataframes

3.2 Plots – simple and advanced plots

Unit 4 (8 Hours)
Bioconductor
4.1 Bioconductor packages- bioclite, biostrings
4.2 Bioconductor packages- edge r

Unit 5 (8 Hours)
Data Analysis
5.1 Data Manipulation and visualization
5.2 Microarray data analysis – Limma

BOOKS FOR STUDY

Robert Gentleman, *R programming for Bioinformatics*, CRC Press, 2016

Jason Kinser. *Python for Bioinformatics*. Massachusetts: Jones and Barlett Publishers, 2009

PATTERN OF ASSESSMENT

End Semester Examination **Total Marks: 100** **Duration: 3 Hours**

Part - A - Programs -	3X 20- 60 marks
Part- B Programs and error handling	2X10 - 20 marks
Viva -	10 marks
Record -	10 marks

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

PYTHON AND R PROGRAMMING PRACTICAL

CODE: 19BI/PC/P232

CREDITS : 2

L T P : 0 0 3

TOTAL HOURS: 39

OBJECTIVE OF THE COURSE

- Demonstrate how to locate and download files for data analysis involving genes and medicine
- Select datasets, open files and pre-process data using Python and R language
- Develop and write R scripts to replace missing values, normalize data, discretize data, and sample data

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Relate the necessity for programming in biology
- Handling biological concepts with Python and R scripts
- Apply programming to analyse genomic sequences
- Gain efficient programming skills
- Perform Microarray data analysis using R language

Unit 1 (7 Hours)

Basics of Python

1.3 Counting letters in DNA strings

1.4 Write a Python program related to Bioinformatics transcription

Unit 2 (8 Hours)

Biopython

2.3 Biopython- using Bioseq –Sequence reading and writing

2.4 Biopython using Bio.Genbank – reading entries

Unit 3 (8 Hours)

Basics of R

3.3 Creating vectors and dataframes

3.4 Plots – simple and advanced plots

Unit 4 (8 Hours)

Bioconductor

4.3 Bioconductor packages- bioclite, biostrings

4.4 Bioconductor packages- edge r

Unit 5 (8 Hours)

Data Analysis

5.3 Data Manipulation and visualization

5.4 Microarray data analysis – Limma

BOOKS FOR STUDY

Robert Gentleman, *R programming for Bioinformatics*, CRC Press, 2016

Jason Kinser. *Python for Bioinformatics*. Massachusetts: Jones and Barlett Publishers, 2009

PATTERN OF ASSESSMENT

End Semester Examination	Total Marks: 100	Duration: 3 Hours
Part - A - Programs -	3X 20- 60 marks	
Part- B Programs and error handling	2X10 - 20 marks	
Viva -	10 marks	
Record -	10 marks	

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

PYTHON AND R PROGRAMMING

CODE: 19BI/PC/PR34

CREDITS : 4

L T P : 4 1 0

TOTAL TEACHING HOURS : 65

OBJECTIVES OF THE COURSE

- Demonstrate how to locate and download files for data analysis involving genes and medicine
- Select datasets, open files and pre-process data using Python and R language
- Develop and write R scripts to replace missing values, normalize data, discretize data, and sample data

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Relate the necessity for programming in biology
- Handling biological concepts with Python and R scripts
- Apply programing to analyse genomic sequences
- Gain efficient programming skills
- Perform genomic data analysis

Unit 1 (12 Hours)

Introduction to Python

- 1.1 Installation of Python, Variables, types, strings, Jupiter notebooks
- 1.2 Objects, functions, control structures, operators, Numpy and Scipy
- 1.3 Fasta files, Parsing DNA and protein information, Gene locations splices, extracting all gene locations

Unit 2 (12 Hours)

Biopython

- 2.1 Getting started and installation, Coding DNA, proteins, extracting translations
- 2.2 Modules- Bio Import, Bio Seq, Bio Align
- 2.3 Plot ABI traces, Retrieve and Annotate Entrez gene

Unit 3 (15 Hours)

R programming

- 3.1 Introduction to R, Installing R, Loops
- 3.2 R as a Deluxe Calculator, Creating Objects and Assigning Values
- 3.3 Graphics: Simple Plotting, Advanced Plotting, Using Color in Plots, Using Subscripts and Superscripts in Graph Labels, Interactive Graphics, Saving Graphical Output, Loops

Unit 4 (13 Hours)

Gene Expression Data Analysis

- 4.1 Feature selection models, Data Preprocessing, Normalization- methods
- 4.2 Data reduction, data sampling, Heatmaps
- 4.3 Classification based on analogy, rules, probabilities, statistics and prediction with R

Unit 5 (13 Hours)

Bioconductor

- 5.1 Introduction, Bioconductor Packages
- 5.2 Expression set, data annotation biomaRt
- 5.3 Applications of R in Phylogenetics and Sequence analysis

BOOKS FOR STUDY

Robert Gentleman, *R programming for Bioinformatics*, CRC Press, 2016

Jason Kinser. *Python for Bioinformatics*. Massachusetts: Jones and Barlett Publishers, 2009.

Mitchell L Model. *Bioinformatics Programming Using Python*. USA: O'Reilly Media Publication, 2009.

BOOKS FOR REFERENCE

Mark Lutz. *Learning Python*. USA: O'Reilly Media Publication, 2009.

Martin C Brown. *Python: The Complete Reference*. Osborne: McGraw-Hill Media, 2001

Gentleman R, Carey V.J, Huber W, Irizarry, RA, and Dudoit, S. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. New York: Springer, 2008.

JOURNALS

The Python Papers Source Codes

The Python Papers Anthology

Python Journal

The R Journal

WEB RESOURCES

www.sthurlow.com/python/

www.learnpython.org

www.codecademy.com/en/tracks/python

<https://docs.python.org/2/tutorial/>

www.pyschools.com/

<http://cran.r-project.org/doc/Rnews/>

PATTERN OF ASSESSMENT

Continuous Assessment Test: Total Marks: 50 Duration: 90 minutes

Section A – 10 x 1 = 10 Marks (All questions to be answered)

Section B – 2 x 10 = 20 Marks (2 out of 4 to be answered)

Section C – 1x 20 = 20 Marks (1 out of 2 to be answered)

Other Components: **Total Marks: 50**
Assignment/Test/Seminars

End Semester Examination: **Total Marks: 100** **Duration: 3 hours**

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 4 x 10 = 40 Marks (4 out of 7 to be answered)

Section C – 2 x 20 = 40 Marks (2 out of 4 to be answered)

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

SUMMER INTERNSHIP

CODE: 19BI/PN/SI32

CREDITS: 2

OBJECTIVES OF THE COURSE

- To enable students to gain experiential learning in the field of Bioinformatics
- To acquire hands on training in Bioinformatics Softwares

The Summer Internship program is for a minimum period of three weeks. The students are expected to have regular attendance in their respective Institutes and submit a report to the Department reporting the experiments they have observed/conducted. The students are expected to give a seminar presentation in the third semester of the work they have observed/conducted.

Guidelines for Evaluation

The maximum marks for the Summer Internship is 50 and is divided into the following:

- a) Log Book (20 Marks)
- b) Seminar presentation (15 Marks)
- c) Attendance (15 Marks)

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

ADVANCES IN BIOINFORMATICS

CODE: 19BI/PC/AB44

CREDITS : 4

L T P : 4 1 0

TOTAL TEACHING HOURS: 65

OBJECTIVES OF THE COURSE

- To develop a quantitative understanding of recent and emerging fields of Bioinformatics
- To provide Hands on experience of handling the genomic and proteomic datasets
- To provide a better understanding of data and its applications in Bioinformatics

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Students will be able to analyse the raw reads of sequences
- Understand the analysis of gene expression
- Experiential knowledge on gene editing techniques
- Perform genomic data analysis
- Learn the skills of cancer genomic data analysis

Unit 1 (10 Hours)

Next Generation Sequencing

- 1.1 Introduction to Next-generation sequencing. History and Future of DNA Sequencing
- 1.2 Introduction to Linux commands and Different Platforms and Applications
- 1.3 Different file formats – FASTQ, SAM, BAM, GFF, Databases and tools – UCSC genome, Galaxy, SRA, NCBI refseq, ENA, FastQC, Bowtie

Unit 2 (12 Hours)

Metagenomics

- 2.1 Metagenomics – Introduction and biological background, case studies of recent research
- 2.2 Alpha and Beta diversity of metagenomic studies
- 2.3 Analysis of metagenome data and logical steps for metagenome analysis

Unit 3 (15 Hours)

Transcriptomics

- 3.1 Introduction and Biological background, case studies of recent research
- 3.2 Quantifying RNA: RNA seq and other techniques. Generating expression table
- 3.3 Logical steps for analysing RNA seq data – differential expression and factor regression analysis

Unit 4
Epigenetics (15 Hours)

- 4.1 Gene regulatory dynamics from analysis of regulatory sequence motifs, transcription factor-DNA interaction,
- 4.2 Local chromatin dynamics and epigenetic modifications, RNA dynamics at the level of transcription and post-transcriptional processing,
- 4.3 3D dynamics of chromatin and the resulting gene regulatory dynamics on daily and development time scales

Unit 5 (13 Hours)
Crispr- Cas 9

- 5.1 Introduction to Crispr, cas9, selection of targets from sequences
- 5.2 Targeted mutagenesis – guide RNA design, recognition sequences
- 5.3 Repair and data analysis of the edited genome, Therapeutic applications

BOOK FOR STUDY

Head, Steven R., Ordoukhanian, Phillip, Salomon, Daniel R, *Next Generation Sequencing Methods and Protocols*, Springer, 2018

Eija Korpelainen, Jarno Tuimala, Panu Somervuo, Mikael Huss, Garry Wong, *RNA-seq Data Analysis: A Practical Approach*, Taylor and Francis publishers, 2017

BOOKS FOR REFERENCE

Takashi Yamamoto. *Targeted Genome Editing Using Site-Specific Nucleases: ZFNs, TALENs, and the CRISPR/Cas9 System*, 2015

Jennifer Doudna, Prashant Mali, *CRISPR-Cas: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 2016

Richard Cummings, J. Pierce, *Handbook of Glycomics*, Academic Press, 2009
ISBN: 9780123736000

WEB RESOURCES

<http://www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practicalcourse/what-you-will-learn/what-next-generation-dna->

<http://www.personal.psu.edu/iua1/courses/2014-BMMB-852.html>

<https://www.illumina.com/science/technology/next-generation-sequencing.html>

<https://bitesizebio.com/21193/a-beginners-guide-to-next-generation-sequencing-ngs-technology/>

<https://edu.t-bio.info/courses>

JOURNALS

Next generation sequencing

PATTERN OF ASSESSMENT

Continuous Assessment Test: Total Marks: 50 Duration: 90 minutes

Theory:

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 5 x 2 = 10 Marks (2 out of 4 to be answered)

Section C – 2 x 10 = 20 Marks

Other Components: Total Marks: 50

Assignment/Test/Seminars

End Semester Examination: Total Marks: 100 Duration: 3 hours

Theory:

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 4 x 10 = 40 Marks (2 out of 4 to be answered)

Section C – 2 x 20 = 40 Marks (2 out of 3 to be answered)

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

DISSERTATION

CODE: 19BI/PC/DS47

CREDITS: 7

Dissertation : 50

Viva : 50

The Dissertation shall contain at least 50 pages and shall be typed with double spacing. The format for the thesis is as follows:

1. Cover page shall contain
 - a) Title of the dissertation
 - b) Name of the Candidate
 - c) Department of Bioinformatics
Stella Maris College (Autonomous), Chennai – 86
 - d) Month, Year

2. The dissertation shall contain
 - a) Contents page
 - b) i. Certificate page
ii. Acknowledgement page
 - c) At least 5 Chapters including an introduction, Review of Literature, Materials and Methods, Result and Discussion and Summary
 - d) List of figures / list of abbreviations (if needed) shall be given as an appendix
 - e) Bibliography shall be given in alphabetical / chronological order at the end.

3. Each candidate may prepare 3 hard copy and one soft copy of the thesis, one copy for her and submit 2 copies to the Head of the department 15 days before the commencement of the fourth semester examination.

4. The candidate may be advised that the dissertation will be valued and given credit on the criteria of
 - a) Motivation towards the chosen area / formulation of the problem
 - b) Methodology and Analysis
 - c) Capacity to interpret the results obtained

5. The Controller of Examination is requested to arrange for the valuation of the Dissertation as well as the conduct of the Viva – Voce at the college where the candidates take examinations, within two weeks of the last date of examination for M.Sc. Degree. The panel of examiners will consist of an external examiner and

the guide. The guidelines for the Viva-Voce examiners would be that a) They will satisfy themselves that this is a work of the candidate as certified by the department b) The thesis is in the given form and c) The candidate has clear understanding of the concepts, discussed in the thesis.

The Department should certify as follows:

This is to certify that the dissertation in the broad area _____ titled _____ is submitted by _____ - at the elective level for the degree of Master of Science (Bioinformatics) during the year _____

sd/

sd/

Head of the Department

Guide

6. A) Guidelines for evaluation

The maximum mark for the dissertation is 100 divided into four components

i.	Style, format and neatness in presentation	15
ii	Chapterization, logic and reasoning	10
iii	Methodology – Analysis and interpretation	25
iv	Viva	50

B) There will be double valuation for the dissertation by the guide and an external examiner who will conduct the viva – voce. The norms for evaluation will be same as applicable for theory papers.

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

ADVANCES IN BIOINFORMATICS PRACTICAL

CODE: 19BI/PC/P442

CREDITS : 2

L T P : 0 0 3

TOTAL HOURS: 39

OBJECTIVES OF THE COURSE

- Demonstrate how to locate and download files for data analysis involving genes and medicine
- Select datasets and pre-process data using
- Develop and replace missing values, normalize data, discretize data, and sample data

COURSE LEARNING OUTCOMES

On successful completion of the course, the student will be able to

- Analyse genomic sequences
- Handle raw data
- Understand the steps of data assembling
- Learn RNA sequence analysis
- Understand and design sg RNA for genome editing

Unit 1	(7 Hours)
Basics of NGS	
1.1 Introduction to UNIX commands and Virtual machine	
Unit 2	(8 Hours)
Metagenomics	
2.1 Analysis of metagenomic raw data using galaxy	
Unit 3	(8 Hours)
Transcriptomics	
3.1 Cancer data analysis using Webmev	
Unit 4	(8 Hours)
RNA seq analysis	
4.1 RNA seq analysis using Biojupies	
Unit 5	(8 Hours)
Crispr cas 9	
5.1 Small guide RNA design – Chop Chop, primerX	

BOOKS FOR STUDY

Head, Steven R., Ordoukhanian, Phillip, Salomon, Daniel R, *Next Generation Sequencing*

Methods and Protocols, Springer Protocols, Humana Press, 2018.

PATTERN OF ASSESSMENT

End Semester Examination	Total Marks: 100	Duration: 3 Hours
Data analysis any two methods	2X40= 80 marks	
Viva -	10 marks	
Record -	10 marks	

STELLA MARIS COLLEGE (AUTONOMOUS), CHENNAI - 600 086
M.Sc. DEGREE: BIOINFORMATICS

SYLLABUS

(Effective from the academic year 2019 -2020)

BIG DATA ANALYSIS

CODE: 19BI/PI/BD44

CREDITS : 4

L T P : 4 1 0

TOTAL TEACHING HOURS : 65

OBJECTIVES OF THE COURSE

- To develop a quantitative understanding of how Data Science in Bioinformatics plays a role in current decade
- To understand the various aspects of data science and applying them in health care
- To obtain adequate knowledge of machine learning approaches

OUTCOMES OF THE COURSE

On Successful completion of the course, the student will be able to

- Describe the Big Data landscape including examples of real world big data problems
- Explain the V's of Big Data and impacts of data collection, monitoring, storage, analysis and reporting
- Identify what are and what are not big data problems and be able to recast big data problems as data science questions
- Gain skills of Hadoop technology
- Learn to get value out of bigdata

Unit 1 (12 Hours)

Introduction to Data Science

- 1.1 Introduction to data science, Case Studies: Data Science in Biomedicine and Healthcare
- 1.2 Sequence Processing, Medical Image Analysis, Natural Language Processing
- 1.3 Network Modelling and Probabilistic Modelling

Unit 2 (15 Hours)

Big Data

- 2.1 What is big data? What makes big data valuable Example of Big Data
- 2.2 Where Does Big Data Come From? Machine-Generated Data and Advantages
- 2.3 Big Data Generated by People, organization of Generated Data, integrating the data

Unit 3 (13 Hours)

Characteristics of Big Data

- 3.1 Characteristics of big data Volume, Variety, Velocity
- 3.2 Characteristics of Big Data – Veracity, Valence and Value
- 3.3 Getting value out of big data using a 5-step process to structure your analysis

Unit 4 (12 Hours)

Data Science: Getting Value out of Big Data

- 4.1 Building a Big Data Strategy, How does big data science happen? Five Components of Data Science
- 4.2 Steps in the Data Science - Acquiring Data , preprocessing and Exploring Data
- 4.3 Analyzing Data, Communicating Results, Turning Insights into Action

Unit 5 (13 Hours)

Big data systems and Hadoop

- 5.1 What is a Distributed File System? Scalable Computing over the Internet, Programming Models for Big Data
- 5.2 Introduction to Hadoop systems, The Hadoop Distributed File System: A Storage System for Big Data, YARN: A Resource Manager for Hadoop
- 5.3 MapReduce: Simple Programming for Big Results, When to Reconsider Hadoop? Cloud Computing: An Important Big Data Enabler

BOOKS FOR STUDY

Peter Guerra and Kirk Borne, Ten Signs of Data Science Maturity, O'Reilly media Pvt ltd, 2016,

Tom White, *Hadoop: The Definitive Guide*” Third Edition, O’reilly Media, 2012.

Seema Acharya, Subhasini Chellappan, *Big Data Analytics*, Wiley 2015.

BOOKS FOR REFERENCE

Howard Wen, *Big Ethics for Big Data*, O'Reilly Media

Michael Mineli, Michele Chambers, Ambiga Dhiraj, *Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses*, Wiley Publications, 2013.

Judith S.Hurwitz, Alan Nugent, Fern Halper, Marcia Kaufman , *Big Data for Dummies*, 2015

JOURNALS

Journal of Bigdata, Springer
Big Data Research, Elseiver

WEB RESOURCES

<https://www.coursera.org/learn/big-data-introduction/home/welcome>
<https://www.coursera.org/learn/bioconductor?action=enroll&authMode=login>

PATTERN OF ASSESSMENT

Continuous Assessment Test: Total Marks: 50 Duration: 90 minutes

Section A – 10 x 1 = 10 Marks (All questions to be answered)

Section B – 2 x 10 = 20 Marks (2 out of 4 to be answered)

Section C – 1 x 20 = 20 Marks (1 out of 2 to be answered)

Other Components: Total Marks: 50

Assignment/Test/Seminars

End Semester Examination: Total Marks: 100 Duration: 3 hours

Section A – 20 x 1 = 20 Marks (All questions to be answered)

Section B – 4 x 10 = 40 Marks (4 out of 7 to be answered)

Section C – 2 x 20 = 40 Marks (2 out of 4 to be answered)