

**BHARATHIAR UNIVERSITY, COIMBATORE 641046**  
**CENTRE FOR COLLABORATION OF INDUSTRY AND INSTITUTION(CCII)**  
**PG DIPLOMA IN GENOME INFORMATICS**

**(For the CCII students admitted from the academic year 2014-15 and onwards)**

**Course Description**

The course focuses on casting contemporary problems in systems biology and functional genomics in computational terms and providing appropriate tools and methods to solve them. Topics include genome structure and function, transcriptional regulation, and stem cell biology in particular; measurement technologies such as microarrays (expression, protein-DNA interactions, chromatin structure); statistical data analysis, predictive and causal inference, and experiment design. The emphasis is on coupling problem structures (biological questions) with appropriate computational approaches.

**1. Admission Eligibility : Minimum 50 % Marks in the following UG / PG Courses**

1. B.Sc. (Botany - Plant Biology & Bio-Technology)
2. B.Sc. (Advanced Zoology & Bio-Technology)
3. B.Sc. (Bio-Informatics)
4. B.Sc. (Bio-Technology)
5. B.Sc( Micro – Biology)
6. B.Sc (Micro – Biology with Nano Technology)
7. B.Sc / B.C.A / B.Sc (IT) / B.Sc (CT) with Biology knowledge in +2
8. B.E (CSE) / B.Tech with Biology knowledge in +2
9. B.Tech – Bio- Informatics
10. B.E / B.Tech – Bio-Technology
11. M.Sc. (Botany - Plant Biology & Bio-Technology)
12. M.Sc. (Bio-Informatics)
13. M.Sc. (Bio-Technology)
14. M.Sc. Industrial Biotechnology

**DURATION: ONE YEAR (TWO SEMESTERS)****SCHEME OF EXAMINATIONS:**

SEMESTER	SUBJECT	UNIVERSITY EXAMS	
		Duration hrs.	Max. Marks
<b>FIRST</b>			
PAPER I	CELL BIOLOGY AND GENETICS	3	100
PAPER II	ESSENTIALS OF LINUX AND BASIC PERL PROGRAMMING	3	100
PAPER III	COMPUTATIONAL GENOMICS	3	100
PRACTICAL -I	IN-PLANT TRAINING	-	100
<b>SECOND</b>			
PAPER IV	NEXT GENERATION SEQUENCING DATA ANALYSIS	3	100
PRACTICAL- II	ESSENTIALS OF LINUX AND BASIC PERL PROGRAMMING	3	100
PRACTICAL- III	COMPUTATIONAL GENOMICS	3	100
	PROJECT	-	100
	Total		800

## PAPER 1 – CELL BIOLOGY AND GENETICS

### Unit – I :

**Cell biology:** Cell concept, structural organization of plant and animal cells, cell membrane and cell wall - Cell cycle: Cell division and regulation, Apoptosis

### Unit- II :

Genetic code: A brief account -Nucleic acids: Structure and properties of different forms of DNA and RNA -DNA replication -Basic concepts of signal transduction- Gene expression and regulation in prokaryotes.

### Unit – III

**Genetics:** Mendel's laws of inheritance and their chromosomal basis, extrachromosomal inheritance -Organization and packaging of genetic material of prokaryotes and eukaryotes, repetitive DNA, structure of chromatin - euchromatin and heterochromatin, chromosome organization and banding patterns, structure of gene - intron, exon and their relationships, overlapping gene .

### Unit – IV:

Molecular mechanism of general recombination, homologous and site-specific recombination, gene conversion

### Unit – V

Types of mutation, mutagens, molecular mechanisms of mutation, site-directed mutagenesis, detection and isolation of mutants, transposons in mutation, repair mechanisms

### Reference:

1. Alberts et. al. 2002, Molecular Biology of the Cell. Garland.
2. Lewin 2004, Genes VIII. Pearson.
3. Lodish et. al. 2004, Molecular Cell Biology. Freeman.
4. Karp 2002, Cell and Molecular Biology. John Wiley.
5. Pollard & Earnshaw 2002, Cell Biology. Saunders.
6. Tobin & Morcel 1997, Asking about Cells. Saunders.
7. Watson et. al. 2004, Molecular Biology of the Gene. Pearson.
8. Atherly et. al. 1999, The Science of Genetics. Saunders.
9. Griffiths et. al. 2004, An Introduction to Genetic Analysis.
10. Hartl & Jones 1998, Genetics - Principles & Analysis. Jones & Bartlett.

## PAPER 2 – ESSENTIALS OF LINUX AND BASIC PERL PROGRAMMING

**Unit – I – Linux :** Fundamentals of the Linux OS, Difference Between Linux and Windows, Difference Between Linux and Unix, GNU, Interesting Facts about Linux (b) Root: Who/why/what is root (c) Basic commands: mkdir, touch, ls, pwd, cd, chmod, df, du, dd, adduser, sort, passwd, rm/rmdir, date, tar, gzip, top (d) Editors: Vi Editor. (e) GNU/LINUX OS Installation (f) Basic System configuration and Administration (g) Understanding Files and Directories in Linux, File Structure and hierarchy, File Permissions

**Unit – II** :Software Installation In Linux: RPM, make. File manipulations using awk and sed, file redirection and piping, command line execution of various bioinformatics programs; shell scripting basics and examples of bioinformatics workflow implemented with unix shell programming

**Unit - III**

**Basic PERL programming** - Introduction to Perl - Get Started -Perl Commands -Basic Data Types –Operators -Control Flow.

**Unit – IV** :Functions & Subroutines -Basic Input and Output -Files & File Handling -Regular Expressions.

**Unit – V** :Patterns - Data Structures –References -Basic Formatting and Reporting

**References**

1. Programming Perl, 3rd Edition Larry Wall, Tom Christiansen, John Orwant. ISBN 0596000278
2. Perl Cookbook Tom Christiansen, Nathan Torkington. ISBN 1565922433
3. Advanced Perl Programming SramSrinivasan. ISBN 1565922204
4. Learning Perl, 3rd Edition Randal L. Schwartz, Tom Phoenix. ISBN 0596001320
5. Learning Perl/Tk Nancy Walsh. ISBN 1565923146
6. Linux complete command reference, Dean Miller

**PAPER - 3 : COMPUTATIONAL GENOMICS**

**Unit – I**

**Introduction to Bioinformatics:**

Definition and History of Bioinformatics, Internet and Bioinformatics, Introduction to Data Mining, Applications of Data Mining to Bioinformatics Problems and Applications of Bioinformatics. Diversity of Genomes: Prokaryotic and Eukaryotic Genomes: Structure, Organization, Gene Families

**Unit – II**

**Bioinformatics databases : Primary nucleotide sequence databases** – EMBL – GeneBank – DDBJ - **Secondary nucleotide sequence databases** - UniGene – SGD -EMI Genomes - Genome Biology -**Protein sequence databases** - SwissProt/TrEMBL - PIR **Sequence motif databases**- PFAM – PROSITE - **Protein structure databases** - Protein Data Bank – SCOP – CATH - **Other biological databases**

**Unit – III :**

**Sequence alignment and database searching** - Single sequence alignments Pairwise alignments - **Scoring matrix** - PAM - BLOSUM -Gap Penalty - **Dynamics programming** - Needleman-Wunsch - Smith-Waterman - **Heuristic methods** - FASTA BLAST - **Statistics of**

**sequence alignment score** : E-Value - P-Value - **Multiple sequence alignments** – ClustalW – Profile - Profile-sequence alignment -Profile-profile alignment PSI-BLAST - Hidden Markov Models – SAM - **Hidden Markov Models** - Markov chains and hidden Markov models - Parameter estimation for HMMs: maximum likelihood, expectation maximization - Connection between pair HMMs and alignments

#### **Unit – IV :**

**Applications of HMMs and Alignments** - Comparative Genomics , The human genome: chromosomes, repeats, genes, and SNPs - Synteny mapping and whole-genome alignment, Finding genes and other functional elements - **Phylogenetics** Sequence-based taxonomy, Models, assumptions, and interpretations, From multiple alignment to phylogeny, Neighbor joining, Maximum likelihood and parsimony, Computer tools for phylogenetic analysis, DISTANCES, GROWTREE, PAUP, PHYLIP

#### **Unit – V:**

**Genomic Sequencing** - Sequencing methods: Sanger sequencing, cloning, shotgun sequencing, Next generation sequencing technologies (illumina, roche/454, ion/torrent, PacBioetc) and its application. **Microarrays and Motifs** - Microarray technology, analysis, and gene clustering -Gene regulation, regulatory motifs, and computational motif finding

#### **References**

1. Bioinformatics Sequence and Genome Analysis - David W. Mount
2. Introduction to Computational Genomics, Nello Cristianini, Matthew W. Hahn Cambridge University Press, 2006, (ISBN-13: 9780521856034 | ISBN-10: 0521856035)
3. Frontiers in Computational Genomics - Publisher: Caister Academic Press ISBN-10: 0-9542464-4-6, ISBN-13: 978-0-9542464-4-0

## **PAPER - 4 NEXT GENERATION SEQUENCING DATA ANALYSIS**

#### **Unit – I**

Introduction to high performance computing, clusters and cloud computing and its role in next generation sequencing data analysis - **DNA-SEQ Assembly**: Reference guided and de novo based whole genome assembly -Post assembly correction – scaffolding -Repeat analysis - Gene finding and annotation- Gene ontology and pathway studies - Domain analysis

#### **Unit – II**

**Reference mapping**: whole genome reference alignment - Coverage analysis SNP and InDel calling - Structural variation and copy number variation calling

#### **Unit – III**

**RNA-SEQ** - Transcriptome assembly and differential gene expression analysis - Different statistical approach for expression studies - Gene set enrichment analysis

**Unit – IV:**

**CHIP-SEQ** - Experimental design - Mapping software - Peak calling, and non-peak based analysis.

**Unit – V:**

Analysis of enriched areas - Viewing ChIP- Seq data in genome browsers

**References**

1. Mark Yandell& Daniel Ence A beginner's guide to eukaryotic genome annotation Nature Reviews Genetics 13, 329-342 (May 2012) | doi:10.1038/nrg3174
2. Daniel R. Zerbino and Ewan Birney Velvet: Algorithms for de novo short read assembly using de Bruijn graphs March 18, 2008, doi: 10.1101/gr.074492.107
3. Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with bowtie 2. Nat Methods 9(4): 357–359
4. Givan SA, Bottoms CA, Spollen WG Computational analysis of RNA-seq Methods Mol Biol. 2012;883:201-19. doi: 10.1007/978-1-61779-839-9\_16
5. Ma W, Wong WH The analysis of ChIP-Seq data Methods Enzymol. 2011;497:51-73. doi: 10.1016/B978-0-12-385075-1.00003-2.