Annexure No.	45 FA
SCAA Dated	29.02.2008

BHARATHIAR UNIVERSITY , COIMBATORE-641 046
(With effect from 2008-09)
PG DIPLOPMA IN BIOINFORMATTICS PROGRAMME
(School of Distance Education)

Subject and Paper	Exam Duration	Total marks
Paper-I : Fundamentals of Biological Systems	3	100
Paper-II : Computational methods for Sequence analysis.	3	100
Paper-III : Programming for Bioinformatics	3	100
Paper-IV : Genomics & Proteomics	3	100
Practical-I : Biological Databanks and Sequence *	3	100

* Practical examination to be conducted at the end of the academic year.

Paper I

FUNDEMENTALS OF BIOLOGICAL SYSTEMS

Subject description   :
Some basic aspects of Molecular Biology and Genetics that are relevant to the course are included in this paper.
Goals:
To understand the basic structure of cell, mechanism and regulation of biological processes fundamental to genome structure and biochemistry.
Objectives:
Students completing this paper should be able to understand concepts of molecular biology that are basic to bioinformatics.

Unit I :
Biology of cells: Cells as a unit of life, structure of prokayotic and eukaryotic cells. An overview of organells (Mitochondria, chloroplasts, ER, Golgi, ribosomes, lysosomes and peroxysomes, nucleus and nucleolus). Differences and similarities in plant and animal cells. Cellular membrane: structure, transport, channels, carriers, receptors, endocytosis, membrane potentials.

Unit II:
DNA replication; Transcription and Translation.
Cell-cell interactions and signal transductions: Intercellular junctions, signaling by hormones and neurotransmitters; receptors, G-proteins, protein kinases and second messangers.Protein traffic in cells.

Unit III:
Cell Cycle and regulation – Mitosis, Meiosis.
Mutation – Types of mutations, types of mutagenic agents and their molecular mechanism; DNA repair; Chromosomal types and structure; Mechanism by which genome undergoes changes, recombination, mutation, inversion, duplication, and transposition.

UNIT-IV
Molecules of Life: Introduction to carbohydrates-Monosacharides and their derivatives, Disacharides, Polysacharides.

Proteins –Structure of aminoacids, Different levels of organization-Primary, secondary tertiary and Quarternary structures.

Nucleic acids – Purines, pyrimidines, Nucleosides and Nucleotides, Different structural form of DNA, denaturation and renaturation of DNA

Lipids-Structure and function of Fatty acids, Triacylglycerols, sphingolipids, steroids and glycerophospholipids.

Water, small molecules-Alkaloids, glycosides, phenols, oligopeptides, Flavonoids, and terpenoids

UNIT-V

Enzymes: Units of Activity,coenzymes and metal cofactors, temperature and pH effects, Michaelis – Menten kinetics, inhibitors and activators, active site and mechanism of enzyme action, Isoenzymes, allosteric enzymes.

Metabolism of glucose: glycolysis, TCA cycle, glycogenesis, glycogenolysis and gluconeogenesis, pentophosphate shunt, ETC. Digestion of protein and protein metabolism, nitrogen balance: transamination, oxidative deamination and urea cycle. Lipid metabolism: beta oxidation. Interconnection of pathways, metabolic regulations.

REFERENCES:
1. Lehninger, A. L. 1984. **Principles of Biochemistry.** CBS publishers and distributors, New Delhi, India
2. Horton, Moran, Ochs, Rawn, Scrimgeour **Principles of Biochemistry** Prentice Hall Publishers.
3. David. E. Sadava **Cell Biology: Organelle Structure and Feunction** Jones & Bartlett publishers.
4. Shanmughavel, P. 2005. **Principles of Bioinformatics,** Pointer Publishers, Jaipur, India.

Paper II

COMPUTATIONAL METHODS FOR SEQUENCE ANALYSIS

Subject description :
This paper describes how to acquire information from biological databases, use of computational approaches to analyze this information, and interpret the results as a guide to experiments in biology.
Goals: The goal of this course is to introduce the main principles of bioinformatics. The coverage will include concepts like sequence alignments, phylogenetic trees, and structure prediction.
Objectives: Understand Genomic data acquisition and analysis, comparative and predictive analysis of DNA and protein sequence, Phylogenetic inference etc

UNIT-I

Introduction to bioinformatics, Classification of biological databases, Biological data formats, Application of bioinformatics in various fields. Introduction to single letter code of aminoacids, symbols used in nucleotides, data retrieval- Entrez and SRS.

UNIT-II

Introduction to Sequence alignment. Substitution matrices, Scoring matrices – PAM and BLOSUM. Local and Global alignment concepts, Dot plot. Dynamic programming methodology: Needleman and Wunsch algorithm. Smith–Waterman algorithm. Statistics of alignment score. Multiple sequence alignment. Progressive alignment. Database search for similar sequences using FASTA and BLAST Programs.

UNIT-III

Evolutionary analysis: distances, Cladistic and Phenetic methods. Clustering Methods. Rooted and unrooted tree representation. Bootstrapping strategies, Use of Clustal and PHYLIP.

UNIT-IV

Gene finding methods. Gene prediction: Analysis and prediction of regulatory regions. Fragment assembly. Genome sequence assembly, Restriction Mapping, Repeat Sequence finder.

UNIT-V

Concepts of secondary structure prediction of RNA and Protein. Probabilistic models: Markov chain, Hidden Markov Models-other applications.

REFERENCES

1. **Bioinformatics – Concepts, Skills, Applications".** S.C. Rastogi, Namita Mendiratta, Parag Rastogi.
2. **Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins.** Andréa's D. Baxevanis, B.F. Francis Ouellette.
3. **Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids.** Richard Durbin et al.
4. **Computer Methods for Macromolecular Sequence Analysis**. Doolittle R.F. (Ed.) (Methods in Enzymology, VOl. 266).
5. Shanmughavel, P. 2005. **Principles of Bioinformatics,** Pointer Publishers, Jaipur, India.
6. **DNA and Protein Sequence Analysis. A Practical approach.** Bishop M.J. Rawlings C.J. (Eds.).
7. **Introduction to Bioinformatics.** Teresa. K. Atwood and David J. Parry-Smith.

Paper III

PROGRAMMING FOR BIOINFORMATICS

Subject description   :

This subject presents the fundamentals of programming techniques, namely sequence of execution, Selection of blocks to be executed, repetition of execution etc with the help of C programming language.

Goals:

To make the students to learn problem solving, execution of programs, thinking the problems in procedure manner and apply the concepts

Objectives:

On successful completion of the course the students should have:
Understood basic of approaching a problem to be computerized
                Learnt the various techniques of writing codes to be executed

UNIT-I:
Programming in C
Introduction, Data types, Operators, Expressions, Control Flow, Structures, Input and Output, Functions, Pointers and References, String Processing, File Handling

UNIT-II
Programming in C++
Basic concepts of OOPS-Introduction to C++,C vs C++-data types, variables, constants, operators and statements in c++- Functions in c++- function prototype-definition-inline functions-overloaded functions.

UNIT- III
Programming in PERL
Introduction, Basic Operators and Control Structures, Scalars, Lists, Hashes, File Manipulation, Pattern Matching and Regular Expressions, Subroutines, Text and String Processing

UNIT-IV
Python Programming
Overview, Data structures, Control Flow, Modules, Basic I/O, Exception Handling, Regular Expressions, File Manipulation, Classes, Standard library

UNIT-V
BioPERL Programming
General Bioperl classes, Sequences (Bio::Seq Class, Sequence Manipulation), Features and Location Classes (Extracting CDS), Alignments (AlignIO), Analysis (Blast, Genscan), Databases (Database Classes, Accessing a local Database)

REFERENCES:
1. The C Programming Language, B.W.Kernighan and D.M. Ritchie 2nd Edition. Prientice Hall of India.
2. Programming Perl – Larry Wall, Tom Christiansen & John Orwant 3ed 2000- O' Reilly
3. Programming Python – Mark Lutz – 2nd Ed., O' Reilly
4. E. Balagurusamy - **"Programming in C++ "** - Tata Mc. Graw Hill Edition
5. Byron Gottfried, - "Programming with C" (Schaum's Outline Series) - Tata
6. McGrawHill Publishing Company - 1998.
7. Object oriented programming with c++ -Robert Laffore -Waite series.
8. Programming Perl - Tom Christiansen, Larry. Wall Orielly Publications

Paper IV

GENOMICS & PROTEOMICS

Subject description :
This paper deals with genome map, comparative genomics, structural genomics, functional genomics and regulation.
Goals: To make the students to familiar with genome map, comparative genomics, structural and functional genomics.
Objectives:
Understand the genome architecture and extracting information like gene function, gene regulation, protein evolution and targets for drug designing

UNIT – I
Definition of Genome, Genome sequencing, Genome map: Types of Genome maps and their uses, High and low-resolution map, Map elements, Polymorphic markers, Types of maps: Cytogenetic, Linkage map, Transcript map, Physical map, Comparative map, Integrated map, STS content maps, Map repositories: NCBI – Entrez Human genome map viewer, OMIM – Online Mendelian Inheritance in Man, Linkage map resources: CEPH reference pedigree, CHLC – Cooperative human linkage center, Radiation hybrid map resources. Practical uses of genome maps: Locating genomic regions, Target identification, Arrangement of genes, SNP diagnosis, Positional specific cloning,

UNIT – II
Genome Anatomies The anatomy of the Eukaryotic Genome –The special features of metaphase chromosomes, where are the genes in the genome? Families of genes, pseudogenes – Eukaryotic organelle genomes, Repetitive DNA content of the human genome.
Transcriptomes and Proteomes
Genome Expression in outline; The RNA content of the Cell– the Transcriptome – yeast and human; The Protein content of the cell - the link between the Transcriptome and the Proteome.

UNIT – III
Annotation of the Genome: Structural annotation (Locating coding regions and other structural elements of the gene). Various approaches in gene prediction: ORF prediction, Gene prediction in prokaryotes and eukaryotes, Hidden Markov Model, Pattern discrimination, Evaluation of gene prediction methods, Prediction of promoter sequences, Functional annotation: (Prediction of gene function), Employing the similarity in the sequence, gene family and metabolic pathway. Employing the conserved domain, Profile and motif comparison, EST Comparison. Analysis of Human Genome.

UNIT – IV
Comparative Genomics:
Purpose and Methods of comparison, Tools for genomic comparison: Applications of Comparative Genomics, Reconstruction of metabolic pathway, Predicting regulatory elements, Identifying targets, examination of domain function, analysis of conserved strings.

Genome projects and Model Organism research -Yeast; Drosophila; C. elegans; and Mouse – a comparative analysis. Comparative genomics as an aid to gene mapping and in the study of human diseases.

UNIT – V
Functional Genomics:
Gene expression analysis by cDNA micro arrays, SAGE, Strategies for generating ESTs and full length inserts; EST clustering and assembly; EST databases (DBEST, UNIGENE); Expression and regulation of entire set of genes, Sporulation Vs Vegetative condition in yeast and *Bacillus.*

REFERENCES
1. **Active Conversation of Non-coding Sequences revealed by three way species comparisons.** Inna Dubchak et al. 2000. Genome Research. 10, 1305–1306.
2. **Bioinformatics A Practical Guide to the Analysis of Genes and Proteins**. Ed. Andreas D. Baxevanis and B.F. Francis Ouellette. A. John Wiley & Sons, Inc., Publications (For mapping and comparative Genomics and COG and other database repositories).
3. **Bioinformatics Sequence and Genome Analysis**. 2001. David W. Mount. Cold Spring Harbor Laboratory Press.
4. Shanmughavel, P. 2005. **Principles of Bioinformatics,** Pointer Publishers, Jaipur, India.
5. Shanmughavel, P. 2006. **Trends in Bioinformatics,** Pointer Publishers, Jaipur, India.
6. **Comparative genetics.** Ann Gibbons, 1998. Science. 281: 1432 – 1434.
7. **Encyclopedia or *Escherchia coli* genes and Metabolism.** Peter D. Karp et al. 1996. Eco-Cyc: Nucleic Acids Research. 10: 86-90.
8. **Structural Genomics and its importance for Gene Function Analysis.** Jeffrey et al. 2000. Nature Biotechnology. 18:283 – 287.
9. **The COG database: New developments in phylogenetic classification of Proteins from complete genomes.** Roman Tatusov et al. 2001. Nucleic Acids Research. 29:22-28.
10. **The Comprehensive Microbial Resource.** Jeremy D. Peterson et.al. 2001. Nucleic Acids Research. 29: 123 – 125.
11. **The Molecular Biology Database Collection: Updated Compilations of Biological Database Resources.** Baxevanis A.D. 2001. Nucleic Acids Research. 29 p 1-10.
12. **Genomes.** T.A. Brown, 2001. Taylor and Francis Group.

PRACTICAL – I - BIOLOGICAL DATABANKS
AND SEQUENCE ANALYSIS

- Biological Databanks Sequence Databases, Structure Databases, Specialized Databases
- Data retrieval tools and methods
- Database file formats
- Molecular visualization
- Gene structure and function prediction (using GenScan, GeneMark)

- Sequence similarity searching (NCBI BLAST)

- Protein sequence analysis (ExPASy proteomics tools)

- Multiple sequence alignment (Clustal)

- Molecular phylogeny (PHYLIP)

- Analysis of protein and nucleic acids sequences,

- Sequence analysis using EMBOSS or GCG Wisconsin Package

REFERENCE:
Bioinformatics a Practical Approach by K. Mani and N. Vijayaraj, Aparna Publications, Coimbatore.

BHARATHIAR UNIVERSITY
COIMBATORE- 641 046

Model Question Papers for SDE courses

FUNDAMENTALS OF BIOLOGICAL SYSTEMS

Time: 3 hours   Max Marks: 100

Answer any Five Questions (5 X 20 = 100)
Each questions carry equal marks

1.Describe about the various components of cell membrane
2.Discuss about the basic mechanism of initiation of transcription
3.Explain the structure of an antibody with a neatly labelled diagram
4.How cell cycle is regulated
5.Discuss about prokaryotic genome organization
6.Describe M.M equation and explain its significance
7.Give an overview of the immune system
8.Discuss the role of co-enzymes in enzyme catalysis

COMPUTATIONAL METHODS FOR SEQUENCE ANALYSIS

Time: 3 hrs   **Total: 100 marks**

Five out of Eight question (5 X 20 = 100)

1. What are the various secondary structure databases? Write its uses.
2. What is Bioinformatics? Write about the applications of   bioinformatics in various fields.
3. Write about sequence similarity search using BLAST and FASTA.
4. What is multiple sequence alignment? Write its applications.
5. Write the uses of Clustal and PHYLIP?
6. What are the methods used for evolutionary analysis.
7. Explain the various aspects of genome sequence assembly.
8. What are the applications of probabilistic models used for secondary structure prediction.

GENOMICS & PROTEOMICS

Time: 3 hours   Max Marks: 100

Five out of Eight question (5 X 20 = 100)

1.Give an account of mouse genome database and snapshot the data content in MGD
2.How will you predict gene function employing the similarity in sequence?
3.Discuss about sequence repeats and its importance
4.Discuss ontological comparison
5.Explain about high and low resolution map
6.Discuss about hidden Markov with neat-labelled diagram
7.Explain about genomic organization of homosapiens and plasmodium
 8.Explain gene clustering with one example.

PROGRAMMING FOR BIOINFORMATICS

Time: 3 hours   Max Marks: 100

Answer any Five Questions (5 X 20 = 100)
Each questions carry equal marks

1. Explain in detail about unions in detail.
2. Write a program to create a union called gene with two members name, count.
3. Create an array of structures called a with 10 members and manipulate them, and display them one by one .
4. How is pointer arithmetic handled. Explain in detail.
5. Write a program which copies an array of integers to another array using pointers.
6. Explain error handling in file I/O .
7. Write a program to store records and display them randomly.
8. Explain about arithmetic and relational operators with examples.