

# **POST GRADUATE (P.G.) CERTIFICATE COURSE IN PHYLOGENOMICS**

## **(Syllabus for the session 2014-15 onwards)**

**Duration:** 1 Year (Weekend) - 10 credits

**Intake:** 20 Seats

**Course fee:** Rs. 10,000/-

### **About Phylo-Genomics**

Phylogenomics is a new way of looking at biological information. It refers to the intersection of several important aspects of modern biology such as molecular biology, systematics, population biology, evolutionary biology, computation, and informatics, with genome-level information as the source for testing hypotheses and for interpretation of data. Because the amount of information from genomes is orders of magnitude greater than previously available, novel approaches and new skills are needed by biologists to make sense of these data. In order to understand the biological information in a phylogenomic context, we first need to understand the nature of biological information and why and how we organize it. The technology explosion has caused a parallel explosion of data production and storage. Understanding the nuances of computing therefore becomes an integral part of understanding phylogenomics. But we also need to have a good handle on the important molecular and evolutionary questions facing modern biology in order to formulate the right questions.

### **Career prospects**

The course gives an added value to the professionals of biological sciences research both academic level and at industry level. The jobs are available at scientist level in pharma industries, vaccine development, clinical research projects, academic research, etc.

### **Eligibility**

Any of the following (minimum 50% marks required in qualifying examination):

MBBS / B.D.S. / BAMS / B.H.M.S / B.Pharmacy / B.Tech-Biotech / BVSC / B.Sc.(Nursing) / M.Sc./ M.Pharm/ B.Sc (with minimum one year relevant industry/ academic research work experience).

Note: Any candidate pursuing post-graduate course in M.D. University, Rohtak, may also co-opt for this course along with their regular course.

### **Course Duration**

Post Graduate Certificate course in Phylo-genomics will be covered within six months beginning with a foundation in bioinformatics and computational biology, followed by modules on structural and functional phylo-genomics. Medium of instruction will be English. The course is designed for postgraduates and those working in industry / academics and is scheduled on weekends only.

## Teaching Methodology

- Theory and practical lectures will be conducted over weekends at the venue by faculty approved by Centre for Bioinformatics, M.D. University, Rohtak.
- Projects: Each student will select the topic for his/her project within eight weeks of joining the course. Thus the projects can be taken by a group of students (not more than 5). Students will send their proposal to the Course Coordinator, who will help to nominate the guides. The project should involve about 10 days of research work.
- Note: This being a post graduate certificate course the students are expected to gain knowledge through interaction with the faculty, use of library /internet and self-study of course material.

## Evaluation and Examination

Students will be evaluated through periodic internal examination, evaluation of assignment and project work. At the end of the course students will appear for final examination conducted by Centre for Bioinformatics, M.D. University, Rohtak.

### Program Specific Outcomes

The students completing this Certificate course will be able to:

- PSO1** Serve the global scientific society with classification, identifying pathogen, answering biological question, forensic and bioinformatics.
- PSO2** Revolutionize the field of evolutionary biology, making it possible to ask and resolve questions tackled only using phylogenomics.
- PSO3** Develop phylogenetic tree, showing the inferred evolutionary relationships among various biological species and their phylogeny-based upon similarities and differences in their physical or genetic characteristics.
- PSO4** Accurately predict gene function based on gene sequence.
- PSO5** Understand the nuances of computing.

The students will be awarded as per absolute grading system, detailed below:

Interval of Marks	Grade	Grade Points
> or = 80 but <or =100	O (Outstanding)	10
> or = 70 but <80	A+(Excellent)	9
> or = 65 but <70	A (Very Good)	8
> or = 55 but <60	B+ (Good)	7
> or = 50 but <55	B (Above Average)	6
> or = 45 but <50	C (Average)	5
> or = 40 but <45	P (Pass)	4
Less than 40	F (Fail)	0
	Ab (Absent)	0

*Note: A student obtaining Grade F shall be considered failed and will be required to reappear in the examination.*

### Course syllabus outline & Scheme of examination

S.No.	Paper code	Paper title	Internal evaluation	End term evaluation	Credits	Total marks
1.	CPG 101	Basic bioinformatics & Biostatistics	10	40	2	50
2	CPG 102	Comparative Genomics	10	40	2	50
3	CPG 103	Structural phylogenomics	10	40	2	50
4	CPG 104	Tools & Techniques for phylogenetics	10	40	2	50
5	CPG 105	Lab-course	-	50	-	50
6	CPG 106	Project work	-	50	2	50
<b>Total</b>			<b>40</b>	<b>260</b>	<b>10</b>	<b>300</b>

**POST GRADUATE (P.G.) CERTIFICATE COURSE IN PHYLOGENOMICS**  
(Syllabus for the session 2014-15 onwards)

**CPG 101: Basics Bioinformatics and Biostatistics**

**MM: Th 80 + IA:20**

**Time: 3 Hours**

**Students completing this course will be able to:**

- CO1** Access the world of Bioinformatics, different types of biological data and databases.
- CO2** Access and explain the tools and techniques of analyzing DNA and protein sequences.
- CO3** Describe & solve questions based on hidden markov model, monte carlo method.
- CO4** Analyze in detail with reference to genetics following techniques & solve questions based on them:  $\chi^2$  -test, t-test.

**Note: Nine questions are to be set in all and the candidates are required to attempt five questions including compulsory question. Question no. I is compulsory covering the whole syllabus. Out of remaining eight questions, two questions are to be set from each unit. Candidate is required to attempt four questions, selecting one question from each unit.**

**UNIT I**

**Bioinformatics resources**

Biological databases, Basic classification – Sequence & Structure; Generalized & Specialized; Primary & Secondary, with example databases (Genbank, EMBL, DDBJ, INSDC, Swiss Prot, PIR, PDB, NDB, BLOCKS, Pfam, ProSITE, etc.); Literature databases.

**UNIT II**

**Bioinformatics tools**

Information retrieval system (Entrez, SRS); Sequence alignment tools (BLAST, FASTA, CLUSTAL-W/X, MUSCLE, TCOFFEE), Variants of BLAST (BLASTn, BLASTp, PSI-BLAST, PHI-BLAST, etc)

**UNIT III**

**Biostatistics**

Clustering: Unsupervised Learning In Large Biological Data, Measures of Similarity, Clustering, Assessment of Cluster Quality, Statistical Network Analysis For Biological Systems And Pathways, Boolean Network Modeling, Bayesian Belief Network, Modeling of Metabolic Networks , Hidden Markov Model, Monte Carlo Method.

## UNIT IV

### Statistical Bioinformatics

Samples and Sampling Distribution, Standard Error, significance level, Degrees of freedom, Tests of significance, tests for proportion, t and F tests Confidence Intervals, Contingency tables of  $\chi^2$  (Chi square) tests of goodness of fit and homogeneity.

Correlation: Simple, Partial and Multiple Correlation, Methods of averages and least squares, polynomial fitting, Regression Analysis. Analysis of variance for one and two way classification

#### Suggested Readings:

1. Essential Bioinformatics by Jin Xiong.
2. Bioinformatics : Sequence and Genome Analysis by David W. Mount.
3. Bioinformatics: Concepts, Skills and applications by R. C. Rastogi.
4. Introduction to Bioinformatics by Singhal and Singhal.
5. Bioinformatics: A Practical Guide to analysis of Genes and Genomes by Andreas D. Baxevanis and b. F. Francis Ouellette.
6. Bioinformatics Basics by Hooman H. Rashidi, Lukas K Buehler.
7. Bioinformatics: A modern approach by Vittal R. Srinivas.
8. Bioinformatics: Database and Systems by Stanley Latovsky.
9. Basic Biostatistics By B. Burt Gerstma
10. Biostatistics: The Bare Essentials By Geoffrey R. Norman, David L. Streiner
11. Fundamentals of Biostatistics By Bernard Rosner

# POST GRADUATE (P.G.) CERTIFICATE COURSE IN PHYLOGENOMICS

(Syllabus for the session 2014-15 onwards)

**CPG 102: Comparative Genomics**

**MM: Th 80 + IA:20**

**Time: 3 Hours**

**Students completing this course will be able to:**

- CO1** Explain the principle of comparative genomics, gene family evolution, gene network and epigenetics.
- CO2** Explain genome comparison, genome alignments (BLAST2, MUMmer and VISTA).
- CO3** Discuss gene order, comparative genomics (viruses, microbes, pathogens) and comparative database (VirGen, CORG, Gramene)
- CO4** Describe GWAS, molecular genetics, transcriptomics, metabolomics.

**Note: Nine questions are to be set in all and the candidates are required to attempt five questions including compulsory question. Question no. I is compulsory covering the whole syllabus. Out of remaining eight questions, two questions are to be set from each unit. Candidate is required to attempt four questions, selecting one question from each unit.**

## Unit I

### Introduction

Principles of Comparative genomics, Methods for genome assembly and annotation. Genomic approaches for the study of structural change, whole genome duplication, gene family evolution, gene networks, gene regulation and epigenetics.

## Unit II

### Genomics Algorithm

Objective and Overview of Genome Comparisons, Genome Alignments: BLAST2, MUMmer, PipMaker, VISTA.

## Unit III

### Comparative Databases

Comparison of Gene Order: Gene Order, Comparative Genomics: Viruses, Microbes, Pathogens, Eukaryotes, Comparative Genomics Databases: VirGen, CORG, HOBACGEN, Homophila, XREFdb, Gramene.

## Unit IV

### Genome Analysis

Single Nucleotide Polymorphism, db SNP and other SNP-related databases, Experimental Techniques and Bioinformatics Approaches, DNA Microarray: Techniques and analysis. Genome-wide association studies (GWAS). Association analyses, After GWAS: Next-gen sequencing, Defining function: Molecular genetics; Transcriptomics; Metabolomics. Regulatory issues in pharmacogenomics.

**Suggested Readings:**

1. Foundations of Comparative Genomics By Arcady R. Mushegian
2. Comparative genomics, Volume 1 Nicholas H. Bergman
3. Comparative Genomics edited by Melody Clark
4. Comparative Genomics: Empirical and Analytical Approaches to Gene Order Dynamics edited by D. Sankoff, J.H. Nadeau
5. Principles of Gene Manipulation and Genomics by S.B. Primrose and R.M. Twyman
6. Genomes by T.A. Brown
7. Related research articles.

# POST GRADUATE (P.G.) CERTIFICATE COURSE IN PHYLOGENOMICS

(Syllabus for the session 2014-15 onwards)

**CPG 103: Structural Phylogenomics**

**MM: Th 80 + IA:20**

**Time: 3 Hours**

**Students completing this course will be able to:**

- CO1** Explain the concept of molecular evolution and phylogenetic tree and comparisons.
- CO2** Discuss protein function in post genomics era, CASP, PHYRE server and structural Genomics.
- CO3** Describe transmembrane prediction, cellular localization webserver and use PYMOL to display functional residue.
- CO4** Explain phylogenetics trees, gene duplication, horizontal, transfer, domain evolution and viral evolution.

**Note: Nine questions are to be set in all and the candidates are required to attempt five questions including compulsory question. Question no. I is compulsory covering the whole syllabus. Out of remaining eight questions, two questions are to be set from each unit. Candidate is required to attempt four questions, selecting one question from each unit.**

## Unit I

### Basic Concept

Concepts in Molecular Evolution, Nature of data used in Taxonomy and Phylogeny: Morphological and molecular character data, Phylogenetic trees and their comparison: Definition and description, various types of trees; Consensus (strict, semi-strict, Adams, majority rule, Nelson). Data partitioning and combination. Tree to tree distances, similarity.

## Unit II

### Structural phylogenetics

Genome annotation protocols; intro to protein structure and function; intro to protein-protein interaction prediction methods including phylogenetic profiles, Rosetta Stone and other methods. Protein function in the post-genomic era. Phylogenomic inference of protein molecular function: advances and challenges. CASP; secondary and tertiary structure prediction; PHYRE server. Protein Structure Prediction and Structural Genomics.

## Unit III

### Cellular localization prediction

Transmembrane prediction; prediction of other cellular localizations. Cellular localization webserver (TargetP, TMHMM and other tools). Consensus approaches over a family and using multiple prediction methods.

Biology of active site residues, and methods to predict them. Predicting functional sites using manual analysis of an MSA and structure. Using PyMOL to display known and predicted functional residues.



## Unit IV

### Applications phylogenetic analysis

Comparison of Phylogenetic Trees obtained using DNA seq. Vs. protein seq. Vs. Full genomes. Need for addition of other properties towards total phylogenetic analysis, Comparative methods for detection of species / organism relationships, Gene duplication, Horizontal transfer, Domain evolution. Study of co-evolution: Plant-insect interactions. Host-parasite interactions. Viral evolution, tree of life.

### Suggested Readings:

1. Phylogenetic Networks: Concepts, Algorithms and Applications By Daniel H. Huson, Regula Rupp, Celine Scornavacca
2. BIOINFORMATICS: METHODS AND APPLICATIONS: (Genomics, Proteomics and Drug Discovery By S. C. RASTOGI, NAMITA MENDIRATTA, PARAG RASTOGI
3. From Genes to Genomes: Concepts and Applications of DNA Technology By Jeremy W. Dale, Malcolm von Schantz, Nicholas Plant, Nick Plant
4. Understanding Bioinformatics By Marketa J. Zvelebil, Jeremy O. Baum
5. Phylogenetics: Theory and Practice of Phylogenetic Systematics By E. O. Wiley, Bruce S. Lieberman

# POST GRADUATE (P.G.) CERTIFICATE COURSE IN PHYLOGENOMICS

(Syllabus for the session 2014-15 onwards)

**CPG 104: Tools and Techniques for Phylogenetics**

**MM: Th 80 + IA:20**

**Time: 3 Hours**

Students completing this course will be able to:

- CO1 Describe Dotplot analysis, scoring matrices, Needleman Wunsch algorithm and BLAST and FASTA programs.
- CO2 Explain CLUSTALW, hidden Markov models of a global, local multiple sequence alignment.
- CO3 Explain phylogenetic analysis algorithm, distance based and reliability of trees.
- CO4 Discuss database(orthoDB, immunoDB, OMA) and phylogenetics software and tools like MEGA, PHYLIP, PHYLODENDRON.

**Note: Nine questions are to be set in all and the candidates are required to attempt five questions including compulsory question. Question no. I is compulsory covering the whole syllabus. Out of remaining eight questions, two questions are to be set from each unit. Candidate is required to attempt four questions, selecting one question from each unit.**

## Unit I

### Sequence Comparison and Alignment

Introduction, sequence alignment reveals function, structure and evolutionary information, principal methods of pairwise sequence alignment, scoring matrices and gap penalties in sequence alignment. Dotplot analysis, measures of sequence similarity: Scoring schemes, Dynamic programming: Needleman-Wunsch algorithm, Significance of alignment, BLAST and FASTA programs, Variants of BLAST and FASTA.

## Unit II

### Multiple Sequence alignment and phylogenetic Analysis

Introduction, multiple sequence alignments as starting points for phylogenetic analysis, global multiple sequence alignment-CLUSTALW, hidden Markov models of a global, local multiple sequence alignment.

## Unit III

### Tree Construction method

Probabilistic models and associated algorithms: Probabilistic models of evolution and Maximum likelihood algorithm; Phylogenetic analysis algorithms: Maximum Parsimony, Distance-based: UPGMA, Transformed Distance, Neighbors-Relation, Neighbor-Joining. Reliability of trees. Bootstrap, jackknife, decay, randomization tests.

## Unit IV

### Phylogenetic Databases and tools

Databases: orthoDB, immunoDB, OMA, PANDITplus, Selectome. Tools: OMA, BLAST, TREEGEN, Supertree, BayeScan. Phylogenetics software and tools like MEGA, PHYLIP, PHYLODENDRON.

### Suggested Readings:

1. Essential Bioinformatics by Jin Xiong.
2. Bioinformatics : Sequence and Genome Analysis by David W. Mount.
3. Bioinformatics: Concepts, Skills and applications by R. C. Rastogi.
4. Introduction to Bioinformatics by Singhal and Singhal.
5. Bioinformatics: A Practical Guide to analysis of Genes and Genomes by Andreas D. Baxevanis and b. F. Francis Ouellette.
6. Molecular Evolution and Phylogenetics By Masatoshi Nei, Sudhir Kumar
7. Phylogenetics: Theory and Practice of Phylogenetic Systematics By E. O. Wiley, Bruce S. Lieberman