

# **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**

**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

# **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**

**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.

# **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**

**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 webservers (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.

# **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**

**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.