MANGALORE UNIVERSITY

DEPARTMENT OF BIOSCIENCES

MSc BIOTECHNOLOGY

BTS 504   BIOINFORMATICS AND BIOSTATISTICS

(SOFT CORE COURSE)

Hours: 40

Course outcome

Students get knowledge and hands on experience of:

- Primary, secondary, genome and structural databases.
- Various sequence alignment programs- BLAST and FASTA along with algorithms and applications.
- Constructing phylogenetic tree to understand evolutionary relationship among the organisms of interest.
- Carrying out protein structure analysis and also protein prediction tools.
- Basics of Computer Aided Drug Design (CADD) and applications in designing new drugs.
- Basics of statistics with measures of dispersion, Normal, binominal and Poisson distribution, student t-test, ANOVA, chi-square etc

UNIT I (13 hrs)

Introduction to Bioinformatics. Basics of UNIX OS and PERL Programming. Biological databases: Nucleotide and protein sequence and structure (primary and secondary) databases, File formats, Molecular visualization softwares. Sequence analysis. Sequence Alignment: Gap penalties, scoring matrices, Alignment algorithms - Global and Local alignments, Dynamic programming and Heuristic methods (BLAST, FASTA). Multiple Sequence Alignment: Tree alignment, Star alignment, Progressive alignment methods and tools. Stand alone packages for sequence alignment: GCG Wisconsin and EMBOSS package.

UNIT II (13 hrs)

Phylogenetics. Representation of phylogeny. Methods of phylogeny: Maximum Parsimony, Maximum Likelihood, Distance method, UPGMA. Softwares for phylogenetic analysis: PHYLIP, CLUSTAL, Tree viewing and editing softwares. Nucleotide sequence and structure prediction methods and tools:

Promoter Scan, Gen Scan, CENSOR, Repeat Masker. Whole genome analysis. Genome sequencing strategies, Restriction mapping, Primer designing. Gene Expression analysis - microarray techniques. Protein sequence and structure prediction, Molecular modeling softwares and servers, Protein folding, Threading. Computer-aided Drug Designing: Molecular Docking. Distributed computing approach: Genome@home, Folding@home.

UNIT III (14hrs)

Statistics – Definition, Application of statistics in Biosciences, Classification and tabulation, Graphical representation of data, Histogram, frequency polygon, frequency curve. Measures of central tendency, Measures of dispersion. Normal distribution, Binomial, Poisson, Probability, non-parametric statistics, Correlation and regression; Sign test, Rank sum test, Rank correlation. Testing of hypothesis: Significance of t-test and ANOVA, Multiple range test, Chi-square test. Experimental designs. Diversity measures and evenness (e.g. Simpson and Shannon). Statistical packages.

References

1. Beginning Perl for Bioinformatics. Tisdall, J.D., San Val Pub., 2001
2. Bioinformatics: Sequence and Genome Analysis. Mount, D.W., CSHL Press, 2004
3. Bioinformatics: Methods and protocols. Misener, S., & Krawetz, S. A., Humana Press, 2000
4. Fundamental Concepts of Bioinformatics. Krane, D.E.& Raymer, M.L., Pearson Ed., 2002
5. Introduction to Protein Structure. Branden, C.-I. &Tooze, J., Garland Pub., 1999
6. Introduction to Bioinformatics. Attwood, T.& Parry-Smith, D., Prentice Hall Pub., 1999
7. Introductory Statistics for Biology. Parker, R.E., Hodder Arnold Pub., 1979
8. Statistics for Biological Sciences. Schefler,W. C., Addison Wesley Pub., 1979
9. Biostatistical Analysis. Zar, J. H. Prentice Hall, 2010
10. Biostatistics. Lewis,A. E.. Prentice Hall, 2010