

## **OBT552 BASICS OF BIOINFORMATICS**

### DETAILED SYLLABUS

#### **UNIT I BIOLOGICAL DATA ACQUISITION**

The form of biological information. Retrieval methods for DNA sequence, protein sequence and protein structure information

#### **UNIT II DATABASES**

Format and Annotation: Conventions for database indexing and specification of search terms, Common sequence file formats. Annotated sequence databases - primary sequence databases, protein sequence and structure databases, Organism specific databases

#### **UNIT III DATA PROCESSING**

Data – Access, Retrieval and Submission: Standard search engines; Data retrieval tools – Entrez, DBGET and SRS; Submission of (new and revised) data; Sequence Similarity Searches: Local versus global. Distance metrics. Similarity and homology. Scoring matrices.

#### **UNIT IV METHODS OF ANALYSIS**

Dynamic programming algorithms, Needleman-wunsch and Smith-waterman. Heuristic Methods of sequence alignment, FASTA, and PSI BLAST. Multiple Sequence Alignment and software tools for pairwise and multiple sequence alignment

#### **UNIT V APPLICATIONS**

Genome Annotation and Gene Prediction; ORF finding; Phylogenetic Analysis: Comparative genomics, orthologs, paralogs. Genome analysis – Genome annotation

#### **TEXT BOOKS:**

1. Introduction to Bioinformatics by Arthur K. Lesk , Oxford University Press.
2. Algorithms on Strings, Trees and Sequences by Dan Gusfield, Cambridge University Press.
3. Biological Sequence Analysis Probabilistic Models of proteins and nucleic acids by Durbin, S.Eddy, A.Krogh, G.Mitchison.
4. Bioinformatics Sequence and Genome Analysis by David W. Mount, Cold Spring Harbor Laboratory Press.
5. Beginning Perl for Bioinformatics: An introduction to Perl for Biologists by James Tindall, O'Reilly Media.

#### **REFERENCE**

1. Bioinformatics "The Machine Learning Approach" by Pierre Baldi and Soren Brunak