

BI 6102 Introductory Bioinformatics (Y1S1) (39 hrs)

This subject teaches basic bioinformatics concepts, databases, tools and applications. It is meant for students to apply their computational knowledge into one of the most promising industries, life science. Upon the completion of this subject, the students should be able to

- Understand the central dogma: basic genetic information transformation from DNA through mRNA to proteins;
- Get familiar with the organisation of the main databases and the accessing tools used in bioinformatics;
- Understand the main algorithms and tools applied for analyzing the biological data:
 - Sequence analysis
 - Similarity search
 - Gene expression analysis
 - Structure prediction
- Be aware of the bioinformatics applications in biomedical industry

Case studies and hands-on sessions are conducted in the BioInformatics Lab to prepare the participants for their individual thesis projects.

1. Introduction to bioinformatics (2 hrs)

Basic concepts of molecular biology: the central dogma, replication of DNA, genes and genome, transcription of DNA into RNA, translation of mRNA into protein. Introduction to methods in bioinformatics: sequence alignment and searching, pairwise comparison and sequence analysis, DNA microarray analysis, protein analysis, molecular phylogeny. Access to sequence data and literature information.

2. Pairwise alignment: algorithms and matrices (5 hrs)

Introduction to pairwise sequence alignment: Dayhoff Model, PAM1 matrix, PAM250 and other PAM matrices. Global and local alignment algorithms: Needleman and Wunsch algorithm, Smith and Waterman algorithm.

3. BLAST search (5 hrs)

BLAST search steps and strategies.

4. Advanced database searching (4 hrs)

Organism-specific BLAST sites, specialized BLAST-related algorithms, BLAST-like tools for genomic DNA searches, PSI-BLAST, PHI-BLAST

5. Multiple sequence alignment (4 hrs)

Introduction to multiple sequence alignment, profile HMMs, databases of multiple sequence alignments, how to make your own multiple sequence alignments, how a multiple sequence alignment provides the basis for phylogenetic trees.

6. Gene expression (3 hrs)

Introduction to bioinformatics approaches to gene expression. cDNA libraries, SAGE, and microarrays.

7. Microarray data analysis (7 hrs)

Microarray data analysis: preprocessing, normalization, transformation. Significance analysis of microarrays, clustering analysis, association rule mining.

8. Protein families & proteomics (3 hrs)

Four perspectives to proteins. Bioinformatics tools for high-throughput protein analysis: 2-D gels electrophoresis, affinity chromatography and mass spectrometry, and yeast two hybrid systems.

9. Protein structure (3 hrs)

Overview of protein structure and structural genomics. Protein data bank. Computational biology approaches to structure.

9. Molecular phylogeny (3 hrs)

Introduction to molecular evolution and phylogeny. Four stages of phylogenetic analysis: phylogeny performed with DNA, RNA, or protein, multiple sequence alignment, tree-building methods, and evaluating trees using randomizing tests and bootstrapping.

Text books:

1. Bioinformatics and Functional Genomics, by Jonathan Pevsner, Publisher: Wiley-Liss (November 4, 2003), ISBN: 0471210048
2. Fundamental Concepts of Bioinformatics, by Dan E. Krane, Michael L. Raymer, Publisher: Benjamin Cummings; 1st edition (September 12, 2002) , ISBN: 0805346333
3. Bioinformatics: Sequence and Genome Analysis, by Mount, David W., Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory Press, [available in NTU e-book]