MATH 4370 or CS 4390

Topics in Bioinformatics

SPRING 2004, Tues. & Thurs. 3:00 - 4:20 pm

Objectives: This course introduces an interdisciplinary area of study that combines techniques and knowledge in mathematical, statistical, computational, and life sciences to understand the biological significance of genetic sequence data. With hands-on projects, students will learn to use molecular sequence analysis software and other bioinformatics resources on the Internet for database search, gene prediction, phylogenetic studies, etc. Lectures and class discussions will emphasize probabilistic models and statistical methods in Bioinformatics as well as biological interpretation of computational results.

Instructors: Dr. Ming-Ying Leung, Professor of Mathematics
Dr. Steve Aley, Associate Professor of Biology
Dr. Elizabeth Walsh, Associate Professor of Biology

Content: Because of the interdisciplinary nature of Bioinformatics, we are expecting a class of students of highly mixed background. Everyone will have to apply the knowledge from their own major field to contribute to active cooperative learning through hands-on projects. We intend to cover the fundamentals of topics 1-5 in the first half of the semester and then students will choose a group project among the remaining topics for in-depth study.

1. **What is Bioinformatics?** Historical development, current challenges and future outlook.
2. **Introduction to Computational Resources for Bioinformatics.** World Wide Web, sequence and structure databases, bioinformatics software packages.
3. **A Primer in Molecular Genetics.** Definition of a gene, gene expression, protein structure and function, conserved sequence motifs, introduction to genome projects.
4. **Probability and Statistics.** Probability and statistical inference, random sequence models, Markov chains and hidden Markov models.
5. **Sequence Alignment and Database Search**. Alignment methods for two sequences, database similarity search using BLAST and FASTA, multiple sequence alignment.
6. **Prediction of Functional Units on Genomes.** Prediction of coding regions and regulatory sites
7. **Evolution.** Phylogenetic models, sequence alignment, tree building methods, and tree evaluation.
8. **DNA Sequence Patterns.** Repeating elements, over- and under-representation of short oligonucleotides, palindrome clusters, detecting periodicity in DNA.
9. **Protein Structure Prediction.** Molecular organization of protein molecules, from Chou and Fasman to neural networks, secondary structure prediction, tertiary structure prediction.
10. **Microarrays and Gene Expression Data.** Design of array experiments and statistical analysis of gene expression data.

Reference Books: No textbook will be required for the course. We shall rely a lot on web resources but the following references may be useful.

1. Introduction to Computational Biology – Maps, Sequences and Genomes, by Michael Waterman (1995)
2. Bioinformatics: A practical guide to the analysis of genes and proteins, edited by A.D. Baxevanis and B.F. Francis Quellette (1998)
3. Bioinformatics: The machine learning approach, by Pierre Baldi and Soren Brunak (1998).
4. Biological Sequence Analysis - Probabilistic models of proteins and nucleic acids by Richard Durbin *et al.* (1999).
5. Bioinformatics: Sequence and Genome Analysis by David W. Mount (2001)
6. Statistical Analysis of Gene Expression Microarray Data by T.P. Speed, Editor, (2003)