CSE 527: Computational Biology, Autumn 2006

An introduction to the use of computational methods for the understanding of biological systems at the molecular level. Intended for graduate students in biological sciences interested in learning about algorithms and computational methods, and for graduate students in computer science, mathematics or statistics interested in applications of those fields to molecular biology.

Time: MW 12:00-1:20

Place: EEB 026

Instructor: Larry Ruzzo (CSE 554, ruzzo@cs.washington.edu)

Course web pages: `http://www.cs.washington.edu/527`

Course mailing list: `cse527a_au06@u.washington.edu`

Catalog Description (somewhat out of date): CSE 527 Computational Biology (3) Introduces computational methods for understanding biological systems at the molecular level. Problem areas such as mapping and sequencing, sequence analysis, structure prediction, phylogenic inference, regulatory analysis. Techniques such as dynamic programming, Markov models, expectation-maximization, local search. Prerequisite: graduate standing in biological, computer, mathematical or statistical science, or permission of instructor.

Workload: Notes, problem sets, project. We encourage projects in which a biologist and a mathematical scientist collaborate to model/solve a biological problem.

Desired Prerequisites: Ideally, students will have a considerable knowledge of one of computer science, biology, or probability/statistics, plus introductory knowledge of the other two. We'll try to supplement as needed (via lecture, outside reading, project teams, etc.) so that everyone has enough background in the immediately relevant areas to fruitfully proceed.

Rough Course Outline

I am unlikely to have time to cover all of this. If you have particular interests, let me know and I'll try to prioritize the in-demand topics.

Essential Background from Molecular Biology

Sequence Analysis Measures of sequence similarity, dynamic programming, pairwise and multiple alignment. Statistical modeling of families of DNA or protein sequences: profiles, motif discovery, hidden Markov Models, Expectation - Maximization algorithm Gibbs sampling. Gene finding.

Molecular Structure Prediction RNA secondary structure prediction, SCFGs and covariance models; the protein folding problem; protein threading.

Microarray Analysis Clustering, classification, feature selection for analysis of large scale gene expression data sets generated by microarrays and similar technologies.