

BIOL 698: current trends in computational biology

January 18, 2005

Course goals:

- A breadth knowledge of current problems and techniques used in computational biology
- Depth in some particular topic

Prerequisites: Basic knowledge of data structures, algorithms, heuristic search methodologies, and some programming background (if doing a programming project)

Textbooks: The course will be loosely based on journal and conference publications, various texts, and Ron Shamir's lecture notes available online from <http://www.cs.tau.ac.il/~rshamir/algmb/01/algmb01.html>. The following textbooks may be used for further study:

- Algorithms on strings, trees, and sequences: computer science and computational biology, Dan Gusfield
- Introduction to computational molecular biology, Setubal and Medianis
- Introduction to bioinformatics, Arthus Lesk
- Introduction to Algorithms, Cormen, Leiserson, Rivest, and Stein
- Computational molecular biology, Pavel Pevzner

- Biological sequence analysis: probabilistic models of proteins and nucleic acids, Durbin et. al.
- Protein bioinformatics, Eidhammer et. al.
- Introduction to computational biology: maps, sequences and genomes, Michael Waterman

All the textbooks are available from the NJIT library and from Amazon as well.

Topics covered: (not necessarily in this order)

- Phylogenetics: applications, models, distance-based methods, maximum likelihood, maximum parsimony, exact algorithms and heuristics
- Multiple Sequence alignment: applications, pairwise alignment, sum of pairs optimization, tree alignments, exact algorithms and heuristics
- Gene expression analysis: hierarchical, k-means, and graph-based clustering approaches, applications
- Hidden Markov Models: algorithms for learning HMMs and applications
- Gene finding techniques
- Protein structure prediction: lattice and analytical models, heuristics under simple and complex models
- More topics if time...

For each topic we will discuss its importance, formalize the problems, and look at various algorithms and heuristics to solve them.

Course grade:

- **Quizzes:** About three quizzes will be held for a total of 60% of the class grade. They will be announced ahead of time and will cover the material covered to date since the previous quiz.

- **Project:** This will make up about 20% of the class grade. A project could be
 - A presentation and critique of a particular paper on some topic covered in class. I will make a list of papers available. If the paper is small then two presentations will be required of two separate papers. Please have the papers approved by me first. Your grade here will be determined by the quality of the presentation, your understanding of the paper, and questions asked by the audience.
 - A programming project where you will conduct an experimental performance study. Such a study could be comparing a set of tools for solving a particular problem such as protein structure prediction or multiple sequence alignment. The programming part, depending upon the project, could involve writing some Perl scripts to run and parse the output of tools or C/C++ code if coding up a new algorithm.
- **Final:** This will make up 20% of the total class grade and will be comprehensive.

There will absolutely be NO make-up exams except on production of a medical excuse with a letter from the doctor.

Office hours: To be announced later. However, feel free to email me at usman@cs.njit.edu anytime regarding questions related to the course, and you may come by my office anytime as well. I am normally in everyday till at least 5pm.