

BIOL 698: current trends in computational biology

November 10, 2004

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 and some programming background (if doing a programming project)

Textbooks: The course will be based upon various papers and tutorials in the literature, along with various lecture notes. The following textbooks may be used for further study:

- Biological sequence analysis: probabilistic models of proteins and nucleic acids, Durbin et. al.
- 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
- Computational molecular biology, Pavel Pevzner
- Introduction to Algorithms, Cormen, Leiserson, Rivest, and Stein

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

Topics covered: (tentative and not necessarily in this order)

- Brief overview of molecular biology and basic computer science we will be needing (this, of course, comes first!)
- Sequence alignment
- Multiple sequence alignment
- Hidden Markov Models
- Gene finding techniques
- Phylogenetics (evolutionary tree reconstruction)
- Mapping of DNA
- Genome rearrangements
- DNA chips and gene networks
- Protein structure prediction
- 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:

- Short quizzes—just to make sure everyone is following the class material. The quizzes will make up a small portion of the class grade and will be cover only the class material. They will be announced ahead of time and will cover the material covered to date since the previous quiz.
- Project: this will make up a larger (about 50% or more) portion 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.

The goal would be to (preferably) have a useful result from the project. For example, your project may be to compare tools for clustering microarray data and report on the best tools and why they perform well. The project may continue into a Masters report or thesis and you may publish your results in a national conference proceedings or a journal.

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.

Finally, I hope you enjoy the course and don't be scared of this outline! I want everyone to do well and learn something from this course.