Math/Stat 547, Winter 2004.

Probabilistic Modeling in Bioinformatics: Biological Sequence Analysis

Instructor: D. Burns

Office: 5834 East Hall

Phone/E-mail: 763-0152/dburns@umich.edu

Office Hours: MW, 10-11, and by appointment beforehand.

Webpage: `http://www.math.lsa.umich.edu/~dburns/547/547syll.html`

The webpage is under revision from last year. It is a good place to see ahead in some detail (caution, there are gaps) what was done last year, and will be done this year. There will be significant differences as the we get further into the term, however. There are also links there to the later part of the lab component (Math/Stat 548).

I have not required a specific textbook this year. A standard reference which will be helpful for some parts of the course is:

Text: "Biological Sequence Analysis", by R. Durbin, S. Eddy, A. Krogh and G. Mitchison, Wiley & Co., 1998.

A fuller bibliography will be provided. The later part of the course is based on papers from the literature. Check the webpage for these details. The papers used for last year are still on the webpage.

Course Requirements: There will be five or six problem sets. Some will be done in groups. In addition, there will be a final project, which can be done in a small group or by yourself. There will be student presentations to the class at the end of the term. There will be no examinations.

It is required that you have an active e-mail account, and check the mail at least once a week. Students in the past have also found it a good way to drop questions to me outside of class or office hours. It is not, however, a substitute for office hours. I will try to get to all your messages the day you send them, but I cannot guarantee this.

Grading: The problem sets will each count towards the final grade with equal weight. The final project/presentation will be about 30% of the total grade. In addition, I may reserve some small percentage of the final grade for general participation.

Syllabus: As of now, I intend to cover many of the topics in the book by Durbin, et al. Of special interest besides these will be

> genefinding and parsing
> protein structure, especially transmembrane proteins
> DNA mechanics and duplex destabilization
> *de novo* peptide identification from mass spectrometry data

There have been many books in the last two or three years on the subject. It is convenient that there are now more books on the mathematical/statistical aspects of the subject, most notably the books by Ewens and Grant, and by Clote and Backofen. Time permit-

ting, we could look at some background on artificial neural networks from Brian Ripley's "Pattern Recognition and Neural Networks" (Oxford Univ. Press, 1995). Michael Waterman's "Introduction to Computational Biology" (Chapman & Hall, 1995) is a standard reference for some of the more mathematical aspects of what we will discuss. I will begin with a review of some elementary facts about probability. You should read Chapter 1 of Durbin and start Chapter 11, the chapter on probability background. I will distribute a list of those books that will be put on reserve in the science library on the main campus, and in the Taubman Library in the Medical School.

I would like to emphasize gene finding or "parsing" and structural issues for proteins and RNA (secondary structure and location of RNA genes).

Computer Laboratory: The laboratory course Math/Stat 548 is scheduled for Tuesday, 9-10, room TBA. It will be scheduled either in 5631 Medical Sciences II (Bioinformatics Computer Lab) or in the basement of East Hall. Either will require special accounts valid for the semester only. The lab component of the course is not required for the course, but is recommended.

Bioinformatics Seminars: We will be having guest speakers from the U of M and Pfizer addressing us on areas of their special expertise. Some of these talks will be incorporated as regular lectures in the course.

There are two seminar series which I call to your attention, though most of you probably are already aware of them. These are the regular Bioinformatics Seminar, (approximately) every other Thursday at 4PM in the West Lecture Hall, room 5330 in Medical Sciences I building (usually! check schedule), but so far not scheduled to start this term until January 29. The schedule can be viewed at

$$http://www.bioinformatics.med.umich.edu/bioinfoseminar.html.$$

Another seminar of interest is the Bioinformatics Journal Club. Again, many of you are already familiar with this. This seminar meets weekly on Mondays from 12 PM to 1PM in Room 5631, Medical Sciences II building (Bioinformatics Computer Lab). The schedule can be viewed at

$$http://www.bioinformatics.med.umich.edu/journalclub.html.$$

Students can enroll in Bioinformatics Journal Club (BI 602) for one credit.
Finally, in a slightly different direction is the Mathematical Biology Research Group seminar. Their emphasis is more on deterministic models, but the seminar series is good. The schedule is found at

$$http://sitemaker.umich.edu/mathbio/upcoming_seminars$$