STA 4953 (Spring 2001)
Special Studies in Statistics: Introduction to Bioinformatics

Objectives: The purpose of this course is to introduce students to bioinformatics - 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 programs from the Wisconsin package of molecular sequence analysis software and other Bioinformatics resources on the Internet. Lectures and class discussion will emphasize on the statistical models and methods in Bioinformatics and biological interpretation of computational results.

Instructors: Dr. Ming-Ying Leung
Office: BB 4.03.32
Phone: 458-5535
Email: mleung@utsa.edu
http://www.math.utsa.edu/~leung/
Office Hours: TR 10:00 – 10:55 a.m. and 12:25 - 1:00 p.m., or by appointment

Dr. J. Aaron Cassill
Office: SB 3.02.27
Phone: 458-5491
Email: acassill@utsa.edu
Office Hours: TR 11:00 a.m. - 12:00 p.m., or by appointment

Ms. Kassandra K. Garnett
Email: kgarnett@texas.net
http://www.biostat.utsa.edu/~kgarnett
Consultation hours: TR 2:00 - 3:25 pm in the BB lab.

Syllabus: We shall cover topics 1 – 5 in the first half of the semester. Depending on available time and students' interest, topics from 6 – 9 will be picked for in-depth discussion.
1. **What is Bioinformatics?** Historical development, the role of mathematics and statistics, current challenges and future outlook.
2. **Introduction to Computational Resources for Bioinformatics.** The UNIX system, World Wide Web, sequence and structure databases, bioinformatics and statistics 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, statistical inference in a Bayesian framework, 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. **Genetic Engineering Analyses of Living Systems.** Sequencing techniques, large scale sequencing projects, physical genome maps and clone libraries.
7. **Prediction of Functional Units on Genomes.** Prediction of coding regions and regulatory sites
8. **Evolution.** Phylogenetic models, sequence alignment, tree building methods, and tree evaluation.
9. **DNA Sequence Patterns.** Repeating elements, over- and under-representation of short oligonucleotides, palindrome clusters, detecting periodicity in DNA.
10. **Protein Structure Prediction.** Molecular organization of protein molecules, from Chou and Fasman to neural networks, secondary structure prediction, tertiary structure prediction.

Reference Books:
No textbook will be required. 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).

Grading:			
	Exercises	30%	(Due in class almost every Thursday)
	Test I	20%	(Thursday, 2/22)
	Test II	20%	(Thursday, 4/5)
	Final project	30%	(In class presentations 4/24 & 4/26;
			Written report due by 1:15 p.m. on 5/11)

NO MAKEUP EXAM will be given except for emergency or medical reasons. In such cases, the student must inform the instructor with 24 hrs from the scheduled time of the test and then submit a written request accompanied by official documents to arrange for a makeup test. **Overdue assignments** will only be accepted for a good reason. However, the instructor reserves the right to discount part or all of the credit for any late homework.

LAST DAY TO DROP an individual course and receive a grade W is 3/23.