Introduction to Bioinformatics – AS 250.265
Course Syllabus

Nicholas Fitzkee [nickfitzkee (@) jhu (.) edu]
Jenkins Hall 204

Course Description

Bioinformatics is a rapidly growing field that integrates molecular biology, biophysics, statistics, and computer science. Fundamentally it is a field focused on comparison: how similar are two given proteins? What are the differences between various DNA sequences? How is the data from one microarray assay different from another? Furthermore, bioinformatics is concerned with quantifying the significance of these differences. In any of the examples above, once a metric for similarity is obtained, it must also be statistically characterized to determine the likelihood that such a relationship could occur by chance. In this course, you will learn many of the popular tools for performing bioinformatics analysis and you will be introduced to the thinking that drives the algorithms.

Meeting Times / Attendance

The course will meet Thursdays and Fridays from 1:00 to 2:30 PM in Jenkins 122. The general format of the course will be a lecture on Thursday and a brief lecture/lab on Friday. A specific schedule of meeting times is given below. Because of the interactive nature of both lectures and labs, attendance is mandatory for the course. Please contact the instructor if you must miss a class meeting.

Reading / Textbook

The textbook for the class is *Bioinformatics and Functional Genomics* by Jonathan Pevsner (2003 John Wiley and Sons, Inc., ISBN 0-471-21004-8). The book is available in the bookstore, but it may be cheaper for you to purchase it online (try www.addall.com). If you have a choice, try to obtain the third printing of the book, with chapter headings in red. In addition, there will be supplemental readings throughout the semester of papers and additional handouts. It is expected that you come to class having read the material.

Homework

Throughout the course, you will have weekly problem sets to test you on the material taught. These assignments are graded and constitute 20% of your final course grade. Homework assignments will often have a computational component to them, and they may partially consist of the lab from that week. You will submit these assignments using the course web server. You are free to submit the written portions of each assignment either electronically (Word document, postscript file, or PDF) or by hand at the beginning of the class period when the assignment is due. Late assignments will not be accepted unless prior arrangements with the instructor are made. In these cases, there must be extenuating circumstances (not simply a test in another course scheduled for the same day).

Grade Distribution

The grades for the course will be calculated according to the table below. Although attendance is not explicitly included, the instructor reserves the right to deduct 5 percentage points from your final course grade if you have more than three unexcused absences.

Course Component	Percentage
In-Class Exams (3)	60%
Homework (10)	20%
Final Project	20%

If the distribution of grades necessitates it, the final grades will be curved. However, if all students do well, they should not expect to receive poor grades simply because of a forced bell curve on the final distribution. Students are encouraged to concentrate on learning, which has lifelong benefits, rather than grades, which are useful to your mid-twenties at best. No student who has shown discipline in pursing educational excellence will fail this course.

Final Project

The ultimate goal of this course is application. To this end, there will be a final project in lieu of a final exam. This project will be due during the final examination period, on or before 12:01 am (the morning of) Monday, May 15, 2006. You will be able to choose from a choice of final projects: one project will involve exhaustively characterizing an unknown genetic sequence provided by the instructor. Another project possibility will involve identifying an as yet undiscovered gene and describing its possible function. Finally, you are welcome to design a web application implementing one of the algorithms discussed in the course. If you choose this last option, you are allowed to work with one partner. More information on the final project will be provided at a later date.

Office Hours

The instructor will be available for course help in his office on most Thursday mornings from 10:00 am to 12:00 pm. If you are unavailable then and wish to schedule an alternative time to meet, please send him an email. Drop-in appointments are welcome, but may need to be rescheduled depending on the instructor's schedule.

Course Web Page

The web page for this course is located at http://roselab.jhu.edu/bioinfo/. Please check this site frequently for course updates. Additionally, many of the homework assignments will be distributed through this site, and you will be able to submit your assignments through this web site as well.

Course Schedule and Topics Covered

Week 1			
Date	**Description**	**Reading**	**Assignment**
February 2	Lecture 1: Introduction, review of biochemistry and molecular biology		
February 3	Lecture 2: Computer tutorial, introductory programming	Supplement I	PS#1 out
Week 2			
Date	**Description**	**Reading**	**Assignment**
February 9	Lecture 3: Computers and biology, online resources for bioinformatics	Chapter 1, 2	
February 10	Lab 1: Simple pairwise alignment	Chapter 3	PS#1 due, PS#2 out
Week 3			
Date	**Description**	**Reading**	**Assignment**
February 16	Lecture 4, Part 1: Pairwise alignment		
February 17	Lecture 4, Part 2: Pairwise alignment		PS#2 due, PS#3 out
Week 4			
Date	**Description**	**Reading**	**Assignment**
February 23	Lecture 5: BLAST searching	Chapter 4	
February 24	Lecture 6: BLAST and PHI/PSI-BLAST	Chapter 5	PS#3 due, PS#4 out
Week 5			
Date	**Description**	**Reading**	**Assignment**
March 2	Lecture 7: Gene expression and microarrays	Chapter 6	
March 3	**Exam I:** Biochemistry, programming, and pairwise alignment		
Week 6			
Date	**Description**	**Reading**	**Assignment**
March 9	Lab 3: Microarray data analysis	Chapter 7	
March 10	Lecture 8: Introduction to protein structure and structural databases	Chapter 9	PS#4 due, PS#5 out

Week 7			
Date	**Description**	**Reading**	**Assignment**
March 16	Lab 4: Molecular visualization	Supplement II	PS#5 due, PS#6 out
March 17	No Class		
Week 8			
Date	**Description**	**Reading**	**Assignment**
March 23	No Class – Spring Break		
March 24	No Class – Spring Break		
Week 9			
Date	**Description**	**Reading**	**Assignment**
March 30	Lecture 9: Proteomics	Chapter 8	
March 31	Lab 5: Bioinformatics methods for studying proteins		PS#6 due, PS#7 out
Week 10			
Date	**Description**	**Reading**	**Assignment**
April 6	Lecture 10: Multiple sequence alignment	Chapter 10	
April 7	Lab 6: Programming and multiple sequence alignment algorithms		PS#7 due, PS#8 out
Week 11			
Date	**Description**	**Reading**	**Assignment**
April 13	**Guest Lecture:** Protein structure prediction (George Rose)	Supplement III	
April 14	**Exam II:** Microarrays and Proteomics		
Week 13			
Date	**Description**	**Reading**	**Assignment**
April 20	Lecture 11: Phylogeny, cladistics, and evolution	Chapter 11, 12	
April 21	Lab 7: Phylogenetic trees		PS#8 due, PS#9 out

Week 12			
Date	**Description**	**Reading**	**Assignment**
April 28	Lecture 12: The human genome project	Chapter 17	
April 29	Lecture 13: Bioinformatics and human disease	Chapter 18	PS#9 due, PS#10 out
Week 12			
Date	**Description**	**Reading**	**Assignment**
May 4	Make up day, review for exam		
May 5	**Exam III:** Phylogenetic analysis, evolution, genomes		PS#10 due

The supplemental reading list is below. Additional readings may be assigned throughout the course.

Supplement I – "Python Tutorial" by Guido van Rossum. In particular, we will emphasize sections 1-5 and 7. Available at http://docs.python.org/tut/tut.html.

Supplement II – Gong, H, and Rose G. D. (2005) "Does secondary structure determine tertiary structure in proteins?" *Proteins* **61** (2): 338-43.

Supplement III – Bradley, P., Misura, K. S. M., and Baker, D. (2005) "Toward High-Resolution de Novo Structure Prediction for Small Proteins." *Science* **309** (5742): 1868-1871.