INTRODUCTION TO BIOINFORMATICS (Fall 2008)

For Undergrads: BIOL 478 / CS 478 / STAT 490B (3 credits)
For Grads: BIOL 595B (4 credits)

Graduates and undergraduates will attend the same lectures and take the same exams. Graduates will also participate in a one hour per week seminar with Dr. Gribskov (time to be arranged)

Course Description: Bioinformatics is broadly defined as the computational study of biological information, targeting particularly the enormous volume of information in DNA and protein sequences, gene expression data, and protein structure and function. Topics in this course will include understanding the sequences of DNA and proteins, the evolution of genes and genomes, the structure and function of proteins, and the dynamics of gene expression in biological processes (transcriptomics). Inherently, bioinformatics is interdisciplinary, melding biology with applications of computer science and statistics. This course introduces analytical methods from biology, statistics and computer science that are necessary for bioinformatics investigations, and demonstrates some application of these methods to biological problems.

The undergrad course is intended for juniors and seniors from various science backgrounds. Our objective is to develop the skills of both tool users and tool designers in this important new field of research.

The grad course has the same goals as the undergrad course with the additional goal of learning to read and critically analyze bioinformatics research literature.

Course Coordinator: Michael Gribskov (Gribskov@purdue.edu)
Lilly G-233, 494-6933; office hours by appointment

Co-Instructors: Daisuke Kihara (DK; Biol. Sci. & CS; dkihara@purdue.edu)

Grad TA: Yifeng (David) Yang (Biol. Sciences; yang41@purdue.edu)

Time & Place: MWF, 12:30 - 1:20 pm, Lilly G-432

Course Material: Biol478/595 provides a survey of the major areas of bioinformatics at the macromolecular level. There are many interesting areas of bioinformatics that we won't be able to cover in a semester, such as cellular modeling, organismal interactions, and ecology..

The course material is divided into four modules:

1. Genomics (DNA and protein sequence analysis)
2. Evolution and Phylogenetics
3. Systems Biology
4. Protein structure

Prerequisites: Background in both biology and computer science is helpful in this course. Background in molecular biology as represented by either BIOL 295E or both of BIOL 231 and 241, and background in computer science as represented by CS 158, CS 177, CS 180 or an equivalent first-year introductory programming course. Students who have not achieved grades of C or better in these courses should consult with the instructor. Admission for students who have not taken these courses is also possible by consent of the instructor.

Course Activities

Regular lectures: Unless indicated on the schedule below all classes will be regular lectures. Readings for the next lecture will be announced at the beginning of each lecture. See course materials below for a description of the text and background references.

Tutorials: Practical demonstrations of bioinformatic analyses using available programs.

Homework: Homework assignments, generally weekly. Generally homework assignments will be published on Monday or Wednesday, and will be due the following Friday, e.g., published on blackbopard 27 Aug, due 5 Sep.

Mini-projects: There will be three long homeworks or mini-projects during the semester. Each mini-project will involve a multidisciplinary approach, and contain both biological and computational analyses. The due dates for each mini-project will be listed on the schedule. The assignment for each will be published on blackboard 2-4 weeks before it is due.

Evening exams: A midterm is tentatively scheduled for 7:00-9:00 PM, Monday, 29 September (location TBA). The comp time will be provided by having no required lecture during class time on that day, instead class time on 29 September will be used for an optional review session. The exam is open book, open notes, but no calculators and no computers. It will cover material through Friday, 26 September.

Quizzes: There will be two quizzes. Each quiz will last 30 minutes. The time before the quiz will be used as catch-up time for the lectures and time for a question and answer review of the material. You will be able to make a single sheet of notes (one side) for reference, but otherwise the quizzes will be closed book, closed notes and no calculators or computers.

Assessment and Grading: Grades in the course will be based on one mid-term and one final exam, together worth a total of 40% of the course grade. The balance of the course grade (60%) will be based on homework assignments, mini-projects and quizzes. The breakdown of points follows (note that this may be tweaked slightly):

Activity	**Points**
Midterm	100
Final	300 (See note below)
Homework	200 (20 points each)
Quizzes	100 (50 points each)
Mini-projects	300 (100 points each)
Total	1000

Note: If the grade on the Final is better than the grade on the Midterm, the grade on the final can replace the Midterm, making the Final worth 400 points.

Semester grades will be awarded based on the following minima of performance: 85% = A, 75% = B, 65% = C and 50% = D; < 50% = F. These thresholds may be adjusted downwards at the instructors' discretion so that it may be easier to get each letter grade, but the thresholds will not be raised. There will be no strict curve (everyone can earn an A, and we hope they do).

Course Policies

Academic behavior: Academic dishonesty of any kind (cheating, plagiarism, fabrication of data, improper collaboration, etc.) is not tolerated and is grounds for failing the course (grade F) and notification of University administration for further disciplinary action. All assignments will be explicitly labeled for individual versus group effort; groups will be instructed as to the rules for collaboration. All questions about course policy and administration should be directed to the Course Coordinator.

Important dates and policies to note if the course proves unsuitable:

Last day to drop without it appearing on your record	8 Sep
Last day to drop without a grade	22 Sep
Last day to drop with W or WF on your record	29 Oct

We'll always give W (and not WF) to encourage those who find the course is unsuitable to drop and not feel like they have to stay with it.

Course Materials

Course Web Page: http://blackboard.purdue.edu

Primary Text:
Mount, David W. 2004. Bioinformatics: sequence and genome analysis. Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor MY, USA. ISBN 0-87969-712-1 (paperback)

Additional readings will be provided for specific lectures; generally these will be posted on blackboard as PDF files.

Background Biology References:
NCBI website (http://www.ncbi.nlm.nih.gov). See Science Primers on Molecular Genetics, Bioinformatics, etc.

Biology and Biochemistry:
Berg, Jeremy M. 2006. Biochemistry. Sixth Edition. W. H. Freeman. ISBN 0-71678-724-5 (hardcover)

The Fifth Edition is also available on the web at:
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?call=bv.View..ShowTOC&rid=stryer.TOC&depth=2.

An excellent text to refresh your knowledge of the biochemistry underlying bioinformatics. Any edition newer than the third is equally valuable for background reading.

Syllabus

Readings in the Mount text are to be completed before the day when they are shown, generally a Monday of the week in which the topic will be covered.

#	Day	Inst.	Topic	Hwk	Reading
			August		
1	M 25	MG	Introduction		
2	W 27	MG	Sequences and Evolution		Handout
3	F 29	MG	Sequences and Evolution		
			September		
	M 1		Labor Day		
4	W 3	MG	Database Searching		Ch. 6
5	F 5	MG	Database Searching	Hw1	
6	M 8	MG	Scoring Matrices		Ch 3 and Ch 4
7	W 10	MG	Pairwise Alignment		
8	F 12	MG	Pairwise Alignment	Hw2	
9	M 15	MG	Pairwise Alignment		Ch 9 and Handout
10	W 17	MG	Genome Sequencing		
11	F 19	MG	Gene Finding/Annotation	Hw3	
12	M 22	MG	Sequence Motifs		Ch. 7
13	W 24	MG	Sequence Motifs		
14	F 26	MG	Evolution & Phylogeny	Hw4	
X	M 29	Both	Exam		
			October		
15	W 1	MG	Evolution & Phylogeny		Ch 5.
16	F 3	MG	Evolution & Phylogeny	(no hw)	
17	M 6	MG	Evolution & Phylogeny		Handout
18	W8	MG	Phylogeny Statistics		
19	F 10	MG	Phylogeny Statistics	Mp1	
	M 13		October Break		
20	W 15	DK	Comparative Genomics		Ch 11
21	F 17	DK	Comparative Genomics	Hw5	
22	M 20	DK	Comparative Genomics Statistics		Ch 13 and Handout
23	W 22	DK	Systems Biology Gene Expression		
24	F 24	DK	Systems Biology Gene Expression	Hw6	
25	M 27	DK	Systems Biology Gene Expression Statistics		TBA
26	W 29	DK	Systems Biology Gene PPI		
27	F 31	DK	Systems Biology Gene PPI	Mp2	
			November		
28	M 3	DK	Intro to protein structure		Ch 10
29	W 5	DK	Protein Families/Protein Evolution		
30	F 7	DK	Protein Families/Protein Evolution	Hw7	
31	M 10	DK	Protein Families/Protein Evolution		TBA
32	W 12	DK	HMM (protein structure/sequence analysis)		
33	F 14	DK	HMM (protein structure/sequence analysis)	Hw8	
34	M 17	DK	HMM (protein structure/sequence analysis)		TBA

35	W 19	DK	Protein Activity and binding		
36	F 21	DK	Protein Activity and binding	Hw9	
37	M 24	DK	Protein Structure Comparison		
	W 26		Thanksgiving		
	F 28		Thanksgiving		
December					
38	M 1	DK	Protein Structure Comparison		TBA
39	W 3	DK	Protein Dynamics		
40	F 5	DK	Protein Dynamics	Hw10	
41	M 8	DK	Homology Modeling		TBA
42	W 10	DK	Homology Modeling		
43	F 12	Both	Review	Mp3	