Biol 536: CIS 535, Spring 2005
Introduction to Computational Biology
MW 3:30-5:00 Leidy 10

Instructor:
Junhyong Kim
Junhyong@sas.upenn.edu
Goddard Labs 206
215-746-5187
Office Hours: Any time by appointment

Teaching Assistants:
Lucia Peixoto
luciap@sas.upenn.edu
Office Hours: TBA

Prerequisites:
(You may talk to me if you would like to take the course but cannot meet the prereqs). College level introductory biology required; undergraduate or graduate level statistics taken previously or concurrently required; molecular biology and/or genetics encouraged; familiarity with computers encouraged.

Course Description:
Biology 536 is an introductory computational biology course designed for both biology students and computer science, engineering students. The course will cover fundamentals of algorithms, statistics, and mathematics as applied to biological problems. In particular, emphasis will be given to biological problem modeling and theoretical perspectives. Students will be expected to learn basic algorithm principles, basic mathematical and statistical proofs, and molecular biology. As mentioned, emphasis is on general principles of biological information analyses, this course will not cover the usage of standard bioinformatics tools.

Course Web page:
This course will utilize Penn's "Electronic Blackboard" web software at: https://courseweb.upenn.edu/. All updates, materials, discussion, and grades will be posted here. Visit the site often. This is a secure site requiring your PennKey. If you have trouble with access, please contact the support staff at bb-support@ccat.sas.upenn.edu.

Grading and other important stuff:

In-course exams (3 total): 75%
Final Project: 25%
Exercises: Bonus points

There will be three short exams given during the semester. These exams will be modularized to the material covered in the previous periods. Each student will be expected to also submit a final project. The final project will be specific individualized assignments handed out after Spring break. You will have a choice of three areas: Programming, Literature Survey, or Data Analysis.

Periodically, exercise problems will be posted to help with concrete understanding of the lecture material. These problems will not be graded per se, but if you hand them in BONUS points will be given (see below for how bonus points will be used). Other impromptu opportunities to collect BONUS points will be presented during lectures.

Make-up exams:
Make up exams will be given only if you submit a request for permission prior to the exam. No makeup will be given if requested after the exams unless you have documented emergency.

Grades:
Numerical grades will be converted to letter grades based loosely on distribution and absolute standards. If your numerical grade is close to a letter grade cutoff, I will use your BONUS points to raise your letter grade.

Grade disputes:
Grade disputes must be submitted in writing. You must write me a letter stating in scientific manner why your answer is correct. I will not accept arguments of the form "but it's all there." No grade dispute letter will be accepted after the Reading Period.

Recommended Reading:
There are no required textbooks for this course. The following is a list of reference books you may use. Lecture notes will be posted weekly.

Bioinformatics: Sequence and genome analysis
David W. Mount, Cold Spring Harbor Lab Press

A Primer of Genome Science
Greg Gibson and Spencer Muse, North Carolina State University

Molecular Evolution: A Phylogenetic Approach
Roderic Page and E. C. Homes

Other books of interest:

Algorithms on strings, trees, and sequences: computer science and computational biology, Dan Gusfield, Cambridge Univ Press. (A very comprehensive algorithms book with easy to read style)

Computational Molecular Biology: An Algorithmic Approach

Pavel A Pevzner (Somewhat hard core computational biology)

Genes VII, Ben Lewin, Oxford Press. (Standard textbook for molecular biology course. Rather lengthy)

Molecular Biology of the Cell, Alberts et al. (Again standard textbook—long)

Recombinant DNA, Watson et al., W.H. Freeman. (Quite outdated but still a great introduction to molecular biology)

Statistical methods in bioinformatics: An introduction, Grant and Ewens, Springer-Verlag. (This is one of the few statistically focused books on bioinformatics. And, you can directly talk to the author!)

Course Schedule:

1/10 M	Introduction	Course organization and goals, Data-Information-Knowledge, survey of computational biology, student survey	JK
1/12 W	Introduction to Web-based Resources	NCBI, EMBL, BLAST	WB
1/19 W	Structure of Biological Information	Basic Biology Concepts	JK
1/24 M	First algorithmic characterization of the genome	Repeat finding problem: Brute force algorithm Z-Box algorithm	JK
1/26 W	More sequence analysis algorithms	Suffix Trees, FSA	JK
1/31 M	No Lecture	No Lecture	JK
2/2 W	Probability and Stochastic modeling	Probability and counting Introduction to stochastic processes Rates and distances	JK
2/7 M	Hidden Markov Models I	Introduction to HMM	JK
2/9 W	Hidden Markov Models II	Computational algorithms, additional models	
2/14 M	Test I		JK
2/16 W	Pattern of biological information	Molecular Evolution, Homology, Micro/Macro evolution	JK

2/21 M	Sequence alignment I	Pairwise alignment/global alignment/local alignment	JK
2/23 W	Alignment Continued	Multiple alignment	JK
2/28 M	Homologous sequence analysis	BLAST and other sequence searches/Distance computation	JK
3/2 W	Statistics Primer	Data organization Computing a statistic Model specification Parameter estimation	JK
3/14 M	Estimating genealogical relationships I	Tree of Life, gene trees, and species trees Terminology Tree graph structures	LW
3/16 W	Estimating genealogical relationships II	The four-taxon problem Estimation principles Distance, Likelihood, and Parsimony	LW
3/21 M	Test II		
3/23 W	Estimating genealogical relationships III	Stochastic models of sequence evolution over a tree Likelihood estimate	JK
3/28 M	Pattern detection: Introduction to statistical learning I	Generative Models of Information/Classification Models of Information General Problem Structure General classification functions Fisher discriminant analysis	JK
3/30 W	Statistical Learning II	Cluster analysis SVM Neural Networks	JK
4/4 M	Gene Prediction	TBA	SH
4/6 W	Promotor Prediction	TBA	SH
4/11 M	Motif Prediction	TBA	SJ
4/13 W	Global gene expression I	Genome-wide gene expression Data quantification	JK
4/18 M	Global Gene expression II	Computing distance measures on gene expression Classification of expression patterns	JK
4/20	Test III		

W			
4/22 F	Special Lecture	Future of computational biology and all things asked. Room to be announced	