BIOL 495S / CS 490B / MATH 490B / STAT 490B (3 cr.; Fall, 2005)

INTRODUCTION TO BIOINFORMATICS

Course Description: Bioinformatics is broadly defined as the study of molecular biological information, targeting particularly the enormous volume of DNA sequence and functional complexity embedded in entire genomes. Topics will include understanding the evolutionary organization of genes (genomics), the structure and function of gene products (proteomics), and the dynamics of gene expression in biological processes (transcriptomics). Inherently, bioinformatics is interdisciplinary, melding various applications of computational science with biology. This jointly taught course introduces analytical methods from biology, statistics and computer science that are necessary for bioinformatics investigations. The course is intended for junior and senior undergraduates 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.

Course Policies: 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. Each mini-project will involve an multidisciplinary approach, i.e., containing both biological and computational assessments. Mini-projects will be focused on various aspects of genomics, transcriptomics or proteomics. Semester grades will be awarded based on the following minima of performance: 90% = A, 80% = B, 70% = C and 60% = D; < 60% = F. These thresholds will not be raised but may be lowered at the instructors' discretion.

Attendance is mandatory. 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.

Course Coordinator: Morris Levy (ML1; Biological Sciences, levym@bilbo.bio.purdue.edu; G-420 LS, #48134; office hours by appointment)

Co-Instructors: Alan Friedman (AF; Biological Sciences: afried@purdue.edu)
Daisuke Kihara (DK;Biol. Sci. & CS; dkihara@purdue.edu)
Mark Levinthal (ML2; Biol. Sciences: marklev@bilbo.bio.purdue.edu)
Katy Simonsen (KS; Statistics: simonsen@stat.purdue.edu)

Grad TA: Tim Williamson (Biological Sciences; williate@purdue.edu)
Chetak Sirsat (Computer Sciences; sirsat@purdue.edu)

Time & Place: MWF, 12:30 - 1:20 pm, **3-102 LS (Lilly Hall)**

Course Web Page: (Register via WebCT-Vista) www.itap.purdue.edu/tlt/ecourses/

Primary Text: (additional readings will be announced)

Mount, David W. 2004. Bioinformatics: sequence and genome analysis. Second edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor MA, USA. ISBN USA. ISBN 0-87969-712-1 (paperback) $75

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

BIOL 495S / CS 490B / MATH 490B / STAT 490B
Fall, 2005: Provisional Schedule (Readings/Assignments-to be announced)

Date	Lec#	Instr.	Topic
Aug. 22	1	ML1	Course introduction: Managing and mining biomolecular information Student biographies
Aug. 24	2	ML2	Information Flow in Biological Systems
Aug. 26	3	ML1	Tutorial on NCBI data bases and BLAST **Homework #1:** DNA-RNA sequence exercise (ML1)
Aug. 29	4	KS	Fundamentals of Data Analysis
Aug. 31	5	KS	Fundamentals of Data Analysis, cont. **Homework#2:** Probability & Statistics (KS)
Sept. 2	6	DK	Overview of Algorithms and Data Structures:Sequences
Sept. 5			Labor Day Holiday – class cancelled
Sept. 7	7	DK	Pairwise Sequence Alignment
Sept. 9	8	DK	Pairwise Sequence Alignment, cont. **Homework #3** (DK)
Sept. 12	9	DK	Sequence searching: FASTA and BLAST
Sept. 14	10	DK	Sequence searching, cont.
		AF	Protein structure: introduction
Sept. 16	11	AF	Protein structure: representations, scale, shape and information
Sept. 19	12	ML1/KS	**Mini-project #1**/ Scoring Matrices (posted only)
Sept. 21	13	AF	Protein sequence conservation, mutation and diversity
Sept. 23	14	DK	Algorithms for Multiple Sequence Alignments
Sept. 26	15	ALL	**Quiz #1**
Sept. 28	16	DK	Hidden Markov Model Algorithms: Dynamic Programming
Sept. 30	17	DK	Hidden Markov Model Algorithms: Profile Alignments
Oct. 3	18	DK	HMMs, cont. **Homework #4?**
Oct. 5	19	ALL	Midterm exam review **Midterm Exam, 7-9 pm, Open notes/no laptops**
Oct. 7	20	ML1	Tutorial on multiple sequence alignment
Oct. 10			October Break – class cancelled
Oct. 12	21	ML2	Why We Sequence Genomes
Oct. 14	22	ML2	Comparing Genomes for Phylogeny
Oct. 17	23	ML1	Phylogenetics I. Why Phylogeny Can be Reconstructed
Oct. 19	24	ML1	Phylogenetics II. Reconstruction Methods
Oct. 21	25	KS	PHYLIP demonstration
Oct. 24	26	KS	Statistical analysis of phylogenetic trees

Oct. 26	27	ML1	Introduction to **Mini-project #2**
Oct. 28	28	ML2	Functional Genomics
Oct. 31	29	DK	Comparative genomics #1
Nov. 2	30	DK	Comparative Genomics #2
Nov. 4	31	KS	Statistical methods for comparative genomics
Nov. 7	32	KS	The statistics of human diseases
Nov. 9	33	ML2	Gene expression and the transcriptome
Nov. 11	34	ML2	Construction and use of microarrays, **Mini-project #3**
Nov. 14	35	KS	Statistical analysis of microarray data #1
Nov. 16	36		**Quiz#2**
Nov. 18	37	KS	Statistical analysis of microarray data #2
Nov. 21	38	AF	Proteins: molecular associations **Homework #5**
Nov. 23-25			Thanksgiving Vacation – classes cancelled
Nov. 28	39	AF	Protein Informatics: Energetics and Dynamics
Nov. 30	40	AF	Binding, Catalysis and Macromolecular Associations
Dec. 2	41	DK	Protein structure comparisons **Homework #6**
Dec. 5	42	DK	Homology Modeling and Threading
Dec. 7	43	DK	*Ab initio* prediction
Dec. 9	44	All	Course evaluation
Dec. 16		All	**Final exam, 10:20 am -12:20 pm, 3-102**