

**BIOL 495S / CS 490B / MATH 490B / STAT 490B (3 cr.; Spring, 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; Biological Sciences:  
marklev@bilbo.bio.purdue.edu)

Katy Simonsen (KS; Statistics: simonsen@stat.purdue.edu)  
**Grad TA:** Adam Boyko (Biological Sciences; aboyko@bilbo.bio.purdue.edu)  
Chetak Sirsat (Computer Sciences; sirsat@purdue.edu)

**Time & Place:** MWF, 12:30 - 1:20 pm, G-432 LS (Lilly Hall)

**Course Web Page: (Register via WebCT) [www.itap.purdue.edu/tlt/ecourses/](http://www.itap.purdue.edu/tlt/ecourses/)**

**Primary Text:** (additional readings will be announced)

Pevsner, Jonathan. 2003. Bioinformatics and Functional Genomics. JohnWiley & Sons, Inc., Hoboken, NJ, USA. ISBN 0-471-21004-8 (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**  
**Spring, 2005: Provisional Schedule (Readings/Assignments-to be announced)**

Date	Lec#	Instr.	Topic
Jan. 10	1	ML1	Course introduction: Managing and mining biomolecular information Student biographies
Jan. 12	2	ML2	Information Flow in Biological Systems
Jan. 14	3	ML1	Tutorial on NCBI data bases and BLAST <b>Homework #1:</b> DNA-RNA sequence exercise (ML1)
Jan. 17			No classes: M. L. King Jr. Day
Jan. 19	4	KS	Fundamentals of Data Analysis
Jan. 21	5	KS	Fundamentals of Data Analysis, cont. <b>Homework#2:</b> Probability & Statistics (KS)
Jan. 24	6	DK	Overview of Algorithms and Data Structures:Sequences <b>Homework #3</b> (DK)
Jan. 26	7	DK	Pairwise Sequence Alignment
Jan. 28	8	DK	Pairwise Sequence Alignment, cont. <b>Homework #4</b> (DK)
Jan. 31	9	DK	Sequence searching: FASTA and BLAST
Feb. 2	10	AF	Protein structure: representations, scale, shape and information content
Feb. 4	11	AF	Protein structure: cont.
Feb. 7	12	ML1/KS	<b>Mini-project #1/</b> Scoring Matrices
Feb. 9	13	DK	Hidden Markov Model Algorithms: Dynamic Programming
Feb. 11	14	DK	Hidden Markov Model Algorithms: Profile Alignments
Feb. 14	15	ALL	<b>Quiz #1</b>
Feb. 16	16	DK	HMMs, cont.
Feb. 18	17	ML1	Tutorial on multiple sequence alignment
Feb. 21	18	DK	Algorithms for Multiple Sequence Alignments
Feb. 23	19	ML2	Why We Sequence Genomes
Feb. 25	20	ML2	Comparing Genomes for Phylogeny
Feb. 28	21	ML1	Phylogenetics I. Why Phylogeny Can be Reconstructed
Mar. 2	22	ALL	Midterm exam review <b>Midterm Exam, 7-9 pm,</b>
Mar. 4	23	ML1	Phylogenetics II. Reconstruction Methods
Mar. 7	24	KS	PHYLIP demonstration
Mar. 9	25	KS	Statistical analysis of phylogenetic trees
Mar. 11	26	ML1	Introduction to <b>Mini-project #2</b>
3/14- 18			Spring Break

Mar. 21	27	ML2	Functional Genomics & <b>Mini-project #2</b>
Mar. 23	28	DK	Comparative genomics #1
Mar. 25	29	DK	Comparative Genomics #2
Mar. 28	30	KS	Statistical methods for comparative genomics
Mar. 30	31	KS	The statistics of human diseases
Apr. 1	32	ML2	Gene expression and the transcriptome
Apr. 4	33	ML2	Construction and use of microarrays, <b>Mini-project #3</b>
Apr. 6	34	KS	Statistical analysis of microarray data: clustering methods
Apr. 8	35	KS	Statistical analysis of microarray data: clustering methods, cont.
Apr. 11	36		<b>Quiz#2</b>
Apr. 13	37	AF	Tutorial on protein sequence conservation, mutation and protein diversity
Apr. 15	38	AF	Protein flexibility and function
Apr. 18	39	AF	Proteins: molecular associations <b>Homework #5</b>
Apr. 20	40	DK	Motif profile search
Apr. 22	41	DK	Protein structure comparison (alignment)
Apr. 25	42	DK	Protein structure prediction (homology modeling)
Apr. 27	43	DK	<b>TBA</b>
Apr. 29	44	All	Course evaluation
May 6	Fri		<b>Final exam, 8-10 am, G-432</b>