Course Number: 3460:445/545  
Course Name: Introduction to Bioinformatics  
Course Credits: 4 (including 1 credit for the 446/546 lab)  
Schedule: Spring  
Syllabus Date: October 2007  
Prepared by: Zhong-Hui Duan

Class web site: [http://www.cs.uakron.edu/~zduan/class/bioinformatics/](http://www.cs.uakron.edu/~zduan/class/bioinformatics/)

Prerequisite

Computer science majors
- Data Structure and Algorithms II (3460:316)

Biology majors
- Cell & Molecular Biology (3100:311) or Principles of Systematics (3100:406)

Open to other majors with consent of the instructors

Textbook

Goals
Bioinformatics is an interdisciplinary field that studies biological systems and biological data (such as DNA/protein sequences, macromolecular structures and functional genomics data) using analytic theory and practical tools of computer science, mathematics and statistics. The goal of this course is to provide students with a broad scope of the new field, introduce the fundamental methods and techniques, and discuss major topics and state-of-art research in bioinformatics. The topics include basic concepts of molecular genetics, biological databases, database searching, sequence alignments, phylogenetic trees, structure prediction, and microarray data analysis.

The goals of this course are to provide students with a broad scope of the field of bioinformatics; to communicate the fundamental methods and techniques in bioinformatics; and to discuss major topics and state-of-art research in bioinformatics.

Bibliography
- Draghici, Data Analysis Tools for DNA Microarrays, Chapman & Hall/CRC, 2003

**Grading Policies**
Your grade will be based on 1 exam (15%), a final exam (20%), 4-5 projects (25%), labs, quizzes, and homework (30%), class participation (10%). Exams are closed book and closed notes. The final is comprehensive. Each project is worth roughly the same amount. Late homework and late labs will not be accepted. Late project will be penalized 10% per day late. Quizzes (if any) will be unannounced.

**Grading scale (**+/- grades may be assigned at instructors' discretion)**
A 90-100; B 80- 89; C 70- 79; D 60- 69; F 0- 59

**Ethics**
"Plagiarism is the intentional or unintentional use of the words or ideas of another without acknowledging their source." (University's Office of General Counsel) All the assignments you submit must be your own work. You must give detailed references to the sources of information if you include any ideas, statements or programs that are not yours. Plagiarism, cheating, undue collaboration, or other forms of academic dishonesty will be reported to the Student Disciplinary Office as a violation of the Student Honor Code.

**Four teaching and learning modules**

**Module 1 (2.5 weeks)**
- Overview of multiple genome projects and biological databases
- Introduction to molecular genetics (for computer science students)
- Introduction to programming environment and basic data structures (for biology students)
- (lab) DNA isolation, PCR amplification, gel electrophoresis, computing environments

**Module 2 (5 weeks)**
- Sequence alignment and searching
  - Pairwise alignment, multiple sequence alignment, dynamic programming, heuristic methods, Bayesian analysis, genetic algorithms, etc.
- Phylogeny construction
  - maximum parsimony, maximum likelihood, and distance methods, etc.
- (lab) DNA sequencing, database searching, software (BLAST, PAUP, BioEdit)
- Two hands-on group projects on sequence alignment and phylogenetic tree construction
- Class presentation of the project reports

**Midterm examination**
Module 3 (2.5 weeks)

- Overview of protein structures and terminology
- RNA secondary structure prediction
- Protein motif analysis, clustering of orthologous groups, protein classification and structure prediction, distance matrix analysis, double dynamic programming, etc.
- (lab) protein database searching, protein structure prediction and visualization tools such as RasMol, Chime
- One hands-on group project on protein secondary structure prediction
- Class presentation of the project reports

Module 4 (4 weeks)

- Overview of the microarray technology and gene ontology
- Clustering methods for microarray data analysis such as hierarchical, K-means, nearest neighbors, and singular value decomposition
- Analysis of clustering results using gene ontology
- (lab) microarray database searching, software such as TreeView, Gene Cluster
- Two hands-on group projects on microarray data acquisition and analysis
- Poster presentation of the project reports

Final examination