

Course description:

The course provides fundamental background in bioinformatics, both theoretical (bioinformatics algorithms) and practical (databases and web-based tools used to study problems in biology), to students in computer science or in biological sciences. It is designed for graduate students. However, undergrad students can also enroll in this course. The requirements for the undergrad students are a strict subset of the requirements for graduate students. Introduction to the biological problems addressed in this course will be provided, as well as a formal definition of the computational problems and a deep exploration of the algorithms for solving these problems. Practical use of topics introduced in class is demonstrated by laboratory exercises and homework problems. Students are grouped for class projects such that each group contains at least one life scientist and one computer scientist. Practical applications of the algorithms and tools discussed in this class to ongoing biological research projects at KSU will be encouraged as topics for course projects.

The course includes the following topics:

- Course overview: bioinformatics
- Pairwise Sequence Alignments
 - Biological Motivation & Problem definition
 - Dot plots
 - Dynamic Programming
 - Substitution Matrices
 - Database Searching
 - BLAST and PSI-BLAST
 - Alignments Statistics and Evaluation
- Gene Finding
 - Sequencing Genomes
 - Gene Structure
 - Problem Definition
 - Methods for Gene Finding
 - Markov Chains
 - Hidden Markov Models
- Multiple Sequence Alignments (MSA)
 - Biological Motivation & Problem Definition
 - ClustalW
 - MSA Tools: Accuracy versus Efficiency Tradeoff
 - Profiles
 - Profile HMMs
- Genetic Variation
 - Single Nucleotide Polymorphism (SNP)
 - Haplotype
 - Hapmap Project
- Phylogenetic Trees
 - Biological Motivation
 - Basic Definitions, Types of Trees
 - Distance-Based Methods

- Parsimony-Based Methods
- Maximum Likelihood Methods
- Bayesian Inference
- Evaluation: Bootstrap
- Comparison of Methods

- Gene Expression Data Analysis
 - Microarray Technology
 - Gene Expression Data
 - Data Preprocessing: Normalization, Replicate Removal, Missing Values, etc.
 - Differentially Expressed Genes
 - Clustering Gene Expression Data
 - Classifying Gene Expression Data
 - Time Series Data Analysis

- Protein Function and Structure Prediction
 - GO Ontology
 - Types of Protein Structures (e.g., SCOP categories)
 - Prediction Methods

- Ethical Issues Related to Bioinformatics and Biotechnology
 - Genetically Modified Organisms
 - Cloning
 - Genetic Testing
 - Stem Cells

- Special Topics and Research Projects

Most of these topics have been taught in CIS 798 / Biol 697 (Spring 2007), so class materials are already available.