George Mason University
Department of Bioinformatics and Computational Biology
Fall 2009
BINF630/BIOL580. Bioinformatics Methods

Thursday, 4:30-7:10 pm
OB Room 304B
3 credits

Instructor: Josif Vaisman
Office: OB, Room 312
Office Hours: By appointment
Phone: 703-993-8431

Course description: The course covers theoretical approaches, techniques and computational tools for DNA and protein sequence and structure analysis. The topics also include biological databases and internet-based bioinformatics resources.

Grading: grades will be based on homework assignments (60%), exam (30%), and class participation (10%).

Textbook:

Marketa J Zvelebil, Jeremy O Baum
Understanding bioinformatics

Syllabus

Lecture 1, 9/3/2009
Lecture 2, 9/10/2009
Lecture 3, 9/17/2009
Lecture 4, 9/24/2009
Lecture 5, 10/1/2009
Lecture 6, 10/8/2009
Lecture 7, 10/15/2009
Lecture 8, 10/22/2009

Reading Materials

Homework assignment 1

Exercises
BINF630: Bioinformatics Methods

Information Theory
Bioinformatics and Computer Networks
Databases and Database Management
Data Mining and Knowledge Discovery in Databases
Machine Learning and Artificial Intelligence
Pairwise Sequence Alignment and Database Searching
Multiple Sequence Alignments
Gene Identification and Prediction, Genome Annotation
Protein Secondary Structure Prediction
Protein Structure Classification
Protein Modeling and 3D Structure Prediction
Protein Structure-Function Relationships

Homework Assignment 1
Given: March 6. Due: March 27.

Homework Assignment 2

Final Exam: May 8, 4:30-6:30 pm

The report should be submitted by email as a Word or PDF file with the filename "b630_f09_hwl_Your_Name.doc or .pdf". The string "b630_f09_hwl" should be also included in the message subject line.

1. Write a regular expression that describes the alignment in the box.

2. Find all known protein sequences that contain the pattern described by the regular expression from Q1. List the IDs of found proteins.

3. Find all known protein structures that contain the pattern described by the regular expression from Q1. List the IDs of found protein structures.

4. Build a multiple sequence alignment for all protein sequences from Q3.

5. Change a single character in the regular expression from Q1 in such a way, that a search of protein sequence database for the pattern described by the modified regular expression would produce a larger number of hits than with the original regular expression.