COURSE ANNOUNCEMENT FOR WINTER 2005

CSE 206B: Algorithms in Computational Molecular Biology (4 units)

Instructor: Vineet Bafna and Pavel Pevzner (CSE) phone: VB-24365, PP-24365, e.mail: vbafna*cs.ucsd.edu*, *ppevzner@cs.ucsd.edu*
$http://www-cse.ucsd.edu/users/vbafna/ http://www-cse.ucsd.edu/users/ppevzner/$
$Time : TuTh 12:30-1:50pm$
$Place : in York Hall 4060B$
$Office hours : by appointment VB : APM 3832 PP : APM 4802$

The course is a graduate seminar devoted to recent research in bioinformatics. The course will be rather self-contained but it implicitly assumes some graduate level algorithmic and mathematical culture. The goal of the course is to introduce students to recent advances in bioinformatics and to provide some research experience by guiding them through an open-ended short-term research project. Students interested in taking this course may contact the instructor.

The course will include 2 types of meetings: (i) lectures given by the instructor and guest speakers and (ii) presentations of research projects given by students.

The course will concentrate on computer science aspects of computational molecular biology and will emphasize the recent advances and open problems in the area. The computational techniques will include algorithms, graph theory, combinatorics, machine learning, etc. The students will be guided through various stages of bioinformatics research: formulating the problem, designing the research plan, studying relevant literature, responding to the criticism of the reviewers and the instructor, reviewing other research projects, preparing the presentation and the paper, etc. The projects in this class may evolve into research papers, for example a number of CSE 202B projects in the recent years evolved into some published and submitted bioinformatics papers. Ideally, we would like every student in the class to submit a research paper inspired by the 206B class project.

Textbooks: There is no required textbook for this class. Students may find the following books useful:

D. Gusfield. Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press. 1997

P.A. Pevzner Computational Molecular Biology: An Algorithmic Approach. The MIT Press, 2000

N.C, Jones and P.A. Pevzner. Introduction to Bioinformatics Algorithms. The MIT Press, 2004.

Grading: The goal of the course is to emulate work on a short-term bioinformatics research project. Grading in this course will consist of 5 components: (i) research project (50 %), (ii) presentation of research (10%), (iii) ability to formulate a research problem (10%), (iv) ability to evaluate other research projects (10%), and (v) take-home final exam based on questions motivated by research projects (20%).

The list of suggested class projects will be distributed in the first week of classes. You may suggest your own well-defined topic and it may be accepted as an extra class project if approved by the instructor. The research projects will be complemented by additional reading/presentation assignments. Reading/presentation assignments will require reading some recent papers (related to you class project) and will result in a class presentation that exposes the participants to the area covered by your research project. To grade your ability to evaluate other research projects, every student in the class will be assigned as a "reviewer" for another class project with the goal to provide critical comments and suggestions.

The class projects have to be posted on the web (see the schedule below) and the students are expected to make a presentation based on the class project. It is important to select the topic as soon as possible and to file the progress reports reflecting your work on the project according to the following schedule. It is important to complete the projects on time (to allow time for presentations) and the schedule below ensures the timely completion. Deviations from this schedule will negatively affect your grade in the class.

- Thursday, Jan, 13. Send an E.mail specifying your choice of the class project. If you prefer to conduct a class project that is not on the list you need to send a one page description of your project. The project should be at the frontier of current research and ideally should be inspired by recent bioinformatics papers. Your selection will either be accepted or rejected (the latter may happen if someone else selects the same paper or if your project proposal is not appropriate for this class). The papers are assigned on a a first come, first reserved basis. While selecting a project, please keep in mind that it should allow you to accomplish the tasks described below. Try to avoid very difficult, unclear and cumbersome problems. Keep in mind that you have limited time, and other classes, so try to choose a manageable topic that is just this side of easy. You are expected to spend about ten hours per week on the project.
- Tuesday, Jan, 18. Meet the instructor to discuss the project (from xx to xx pm in APM xxxx). You may want to sign up for a 20-30 minute meeting on April 14 to avoid waiting in line. Prior to the meeting please prepare (i) a list of questions to be answered during the project, (ii) a list of three-four recent papers related to your project, and (iii) a list of a few milestones, with a deadline for achieving each milestone.
- Thursday, Jan,20. Final assignments of the projects and related papers for class presentations. Start work immediately on your project. You also will be assigned a recent paper for presentations that will start in class in late April. Every project will be assigned a reviewer who is responsible for critical assessment of the project.
- Thursday, Jan, 27. Send a 1-page E.mail to the instructor and the reviewer outlining your research plan, algorithmic challenges, and software implementation efforts. Most suggested bioinformatics projects do not contain an explicit formulation of the algorithmic problem they address. Writing an explicit self-contained algorithmic formulation of the problem is a major component of the class project. The formulation (self-contained!) should be shorter than a page to be appropriate for the project. Try to choose projects that allow elegant and relatively short formulations. The key here is to abstract just enough complexity away to see a clean algorithmic problem without getting lost in biological detail.

- Monday, Jan. 31. Reviewers send their critical comments back to both authors and the instructor.
- Friday, February, 4. Send a 3-page E.mail to the instructor and the reviewer outlining the first algorithmic results and the state of software implementation efforts. This E.mail should contain the response to the reviewer's concerns. Sign up for a 15 min meeting on Thursday, April, 29 (from 2 to 6 pm) to discuss progress.
- Monday, February, 7. Reviewers send their critical comments back to both authors and the instructor.
- Tuesday, February, 15. The deadline for a 5-page intermediate report to be distributed in the class. You are expected to make a short presentation in the class describing your project, the challenges you are facing, and the preliminary results. Reviewers serve (and being evaluated) as moderators for these presentations.
- Tuesday, February, 22. The deadline for the final report and the powerpoint presentation. Sign up for a 15 min meeting on Tuesday, May, 11 (from 2 to 6 pm) to discuss your progress.
- Thursday, February, 24. The presentations of the class projects start (in class) according to the schedule provided by the instructor. Reviewers serve as moderators for these presentations.
- Tuesday, March 8. Design a new graduate-level homework problem that relates to your class project and write up its solution (this assignment evaluates your ability to formulate new problems). The formulation of the problem should be elegant and self-contained. Send the formulation/solution to the instructor. Please do not discuss your problem with students in the class and do not send the problem to the reviewer. Sign up for a 15 min meeting on Tuesday, May 18 (from 2 to 6 pm) to discuss your homework problem.
- Thursday, March, 10. Send the revised formulation/solution of the homework problem to the instructor.
- The best homework problems combined with problems motivated by the instructor's presentations will form the basis of the take-home final exam (so if your problem is selected you have one less problem to solve!). Since the final exam for this course will be mainly based on the presentations given in the class please be an active participant of all class presentation and make sure that you understand the presented material. Your class participation contributes to component (iv) of your grade (ability to evaluate other research projects).

Your class project and presentation will be graded according to the following 10 criteria:

- ability to formulate a computational problem.
- ability to review the previous research in the area.
- ability to write a self-contained and concise abstract and introduction.
- ability to propose efficient algorithmic solutions
- sensible implementation decisions
- sensible benchmarking design
- clear description of results in progress report
- clear presentation of results in the class, appropriate organization, and presentation style
- insightful discussion of major experimental results and further directions
- complete bibliographic review