General course information

Lectures: Tues/Thurs 10:10–11:25, Weill Hall 226
Discussion: Fri 12:20–1:10, Caldwell Hall 100
Credit hours: 4 (S/U or letter)
Instructor: Amy Williams, 102G Weill Hall
Office hours Tues 3:00-4:00 PM
TAs: Dilip Thiagarajan, Andrew Vaughn
Office hours TBA
Course webpage: http://williamslab.bscb.cornell.edu/btry4840/
Course forum: Piazza (linked from course webpage)

Prerequisites

The prerequisites for this course are listed as “BTRY 3010 and CS 2110 or equivalents.” In practice, those who are willing to work at understanding algorithms or statistics can do well. However, students with strong backgrounds in these disciplines will find the course easier than those who don’t. The following are general guidelines for background in the three major topic areas in this course:

- **Statistics.** A general statistical methods course (e.g., BTRY 3010 or 6010) would provide helpful background. Those who have taken only a probability course will need to spend additional time understanding the statistical inference approaches we will cover. However, such a student can succeed if they expend sufficient effort, including consulting outside textbooks where needed.

- **Computer science.** The ability to program is essential to completing this course, and we will accept problem set solutions written in Python or R. Additionally, because we will discuss algorithms and algorithm complexity, an intermediate or advanced programming/algorithms course such as CS 2110 or CS 3110 would be best. Students that have not taken these courses can succeed if they take time to learn algorithm complexity and design.

- **Molecular biology and genetics.** We will be discussing concepts in molecular biology and genetics. However, we will not delve deeply into these topics. We aim to discuss all biological and genetics concepts that students need in this course within the lectures and discussion sections. Occasional outside reading from available online resources on these topics is likely to be beneficial to gain greater clarity on these concepts, although detailed understanding is not needed.

Textbooks

The primary textbook for this course is:


This book covers much of the material we will discuss in this course and it will be on reserve at Mann library within two weeks. There are a number of errors in the book that are addressed in the online errata at [http://selab.janelia.org/cupbook_errata.html](http://selab.janelia.org/cupbook_errata.html). Most errors have been corrected in the latest version (10th reprinting, 2006), which you may wish to obtain. New copies cost ~$48 on Amazon, and used copies are also available from both Amazon and half.com.

Other books on bioinformatics that may be worth referencing are:


The Felsenstein book contains very detailed information about methods for phylogenetic inference. It also contains introductions to continuous-time Markov models of DNA substitution, maximum likelihood phylogeny reconstruction, and other topics we will cover. The Jones and Pevzner book focuses on algorithms, with a clear introduction to dynamic programming.

Students who wish to study more on probability and statistics, algorithms, or molecular biology can explore books on these topics. Some relevant books include:


**Grading**

Grades will be based on a combination of five problem sets, a final project, and participation, with each contributing to final grades as follows:

- Problem sets: 12% × 5 = 60%
- Final project and presentation: 30%
- Participation: 10%

Problem sets will be fairly challenging and take some time. Start early and make use of discussion section and office hours. These will be assigned roughly every other week and you will have two weeks to complete them. These will stop before the end of the term so you can devote time to your final project.

The project is an opportunity to apply the concepts and skills you have learned in this course to a real research question. It should be substantial. Start thinking about it early, even if it takes until late in the semester for you to decide what you will do. A list of possible project ideas will be posted on the class website soon after Fall Break. You will be asked to submit a brief project proposal by late October.

For graduate students (BTRY 6840), the project should involve original research, and except in rare cases (e.g., a challenging theoretical project) should involve some programming and some analysis of real biological data. Undergraduates (BTRY 4840/CS 4775) will not be expected to do original research, but should attempt a substantial programming project, a comprehensive literature review on a topic of interest, or something of similar scope. In all cases, a written project report and five-minute presentation to the class will be due and take place the last week of class. Some class projects may lead to (Master’s or Ph.D.) thesis projects, and/or to conference or journal publications.

Note the significant class participation component. You are expected to participate in classroom discussions, ask questions, and solve problems in class. Full participation credit will be awarded to students that attend 60% of discussion section, and attendance will be based on iClicker responses to questions.

**Late policy:** Late problem sets will have their score reduced by 10% per day late. Two free late days will be granted without any reduction in score. Final Projects must be in on time.

**Collaboration and academic integrity**

Collaboration is encouraged, but everyone must turn in his or her own work and must acknowledge all sources and collaborators. Be sure you understand the work you turn in. Final projects must be done individually. No group projects are allowed.

Each student in this course is expected to abide by the Cornell University Code of Academic Integrity. Any work submitted by a student in this course for academic credit will be the student’s own work. For this course, collaboration is allowed on the problem sets so long as students turn in their own work and list the individuals they collaborated with.