BIOL526H Fall 2010

Biology 526H: Computational Genetics

Fall 2010

Description:

Modern biological research relies heavily on computers to manage and make sense of ever-growing volumes of data of ever-increasing complexity. Computational tools are needed to help make sense of everything from DNA sequences to global biodiversity hotspots. This interdisciplinary seminar pulls together threads from computer science and statistics used to study computational approaches to problems in one of the most information-rich disciplines of biology, genetics and genomics.

Seminar-style classes will be complemented by computer labs that will provide hands-on experience in genomic data analysis. The course is also intended to give students the opportunity to hone their scientific writing and presentation skills and get a taste of independent research. It is aimed at life science students who have an affinity for mathematical puzzles and programming.

4.0 credit hours.

Schedule

Seminars: Tues-Thurs 11:00-12:15, Stone Center Rm. 201
Computer Lab: Thurs 1:00-1:50, Stone Center, Rm. 201

Instructor

Todd Vision, Associate Professor of Biology
phone: 843.4507
email: tjv@bio.unc.edu
office hours: Tuesday 1-3pm or by appt, 102 Coker Hall

Teaching Assistant

Thomas (Toby) Clarke (TA)
phone: 843-4508
email: thclarke@email.unc.edu
office hours: Wed 1-3pm or by appt, 122 Coker Hall.

Enrollment

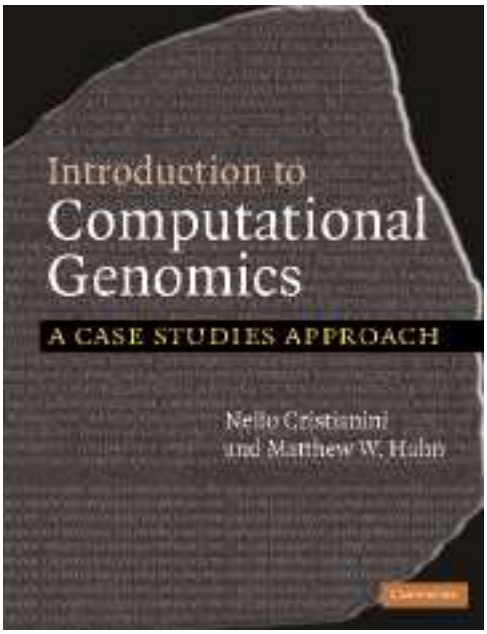
This course is intended for advanced undergraduates and beginning graduate students in the life sciences. Students should have taken the following UNC courses or their equivalents, or receive permission from the instructor. These may be satisfied as prerequisites or corequisites:

- BIOL 202: Genetics and Molecular Biology (or equivalent)
- COMP 101: Computers: Power Tools for the Mind (or equivalent). No specific programming language is required as a prerequisite.
- STOR 155 - Introduction to Statistics (or equivalent)

Enrollment is limited. Honors and graduate students are given priority. A grade point average of 3.0 is required for all undergraduates.

Matlab will be used in the course, but no prior experience is assumed. Familiarity with UNIX command line will be helpful.

Reading

	**Introduction to Computational Genomics: A Case Studies Approach** Nello Cristianini and Matthew W. Hahn Cambridge University Press (ISBN-13 978-0-521-67191-0).

Supplementary readings will be made available via class handouts, Blackboard, and library reserves.

Computer Labs

In the first two-thirds of the semester, the lab sessions will be an opportunity for guided work on programming projects in Matlab as well as pencil-and-paper problems. In the final third of the semester, the lab sessions will provide an opportunity to work on independent projects.

Course Information

Bio562H public course website
General information about the course. It is in a public Google Doc so that anyone can access it.

Schedule
Aug 24 Introduction to computational genetics
Aug 26 Basic sequence analysis (Ch 1), Lab 1
Aug 31 Gene finding (Ch 2)
Sep 2 Gene finding, Lab 2
Sep 7 Sequence alignment (Ch 3)
Sep 9 Sequence alignment, Assignment 1 due, Lab 3
Sep 14 Hidden Markov models (Ch 4)
Sep 16 Hidden Markov models, Lab 4
Sep 21 Population genetics (Ch 5)
Sep 23 Molecular evolution (Ch 6), Assignment 2 due, Lab 5
Sep 28 Sequencing of the Neandertal genome
Sep 30 Phylogenetics (Ch 7)
Oct 5 Phylogenetics
Oct 7 TBD, Assignment 3 due, Lab 7
Oct 12 HOLIDAY: University Day
Oct 14 Genome evolution (Ch 8), Lab 8
Oct 19 Genome evolution
Oct 21 HOLIDAY: Fall Break, No lab
Oct 26 Gene expression (Ch 9), Assignment 4 due
Oct 28 Gene expression, Lab 9
Nov 2 Finding regulatory motifs (Ch 10)
Nov 4 Finding regulatory motifs, Lab 10
Nov 9 Review session, Assignment 5 due
Nov 11 Journal club (Kellis et al., 2003, see Documents/Readings) and review session during lab
Nov 16 Exam
Nov 18 Exams returned, discussion of guidelines for paper presentations and independent projects
Nov 23 Paper presentations - note this class will run from 11am-1pm and be held in Coker 215
Nov 25 HOLIDAY: Thanksgiving Break
Nov 30 Discussion of independent project proposals
Dec 2 Independent project proposal presentations (during both seminar and lab)
Dec 7 Special topic: Ontologies and computing over qualitative data
Dec 16 Final paper due by 12pm. (Compile as one PDF document and submit through Dropbox)

Grading
In this instructor's opinion, grading is at best a necessary evil. Since this is an honors seminar, and the principal purpose is to give you the opportunity to think in ways that traditional classes do not, grading will in part be based on a subjective evaluation of how

engaged you (the student) are with the course. Make it your own and take from it what you think most useful. For the rest of your life, you will be responsible for your own education!

Assignments are intended largely to give you experience with the concepts and ideas in the course, and not as tests. That said, assignments will be graded, with weights as follows:

- Problem sets: five assignments worth 5% each, with the lowest problem set grade automatically dropped, thus cumulatively worth 20% of the final grade.
- A midterm exam, format TBA, worth 15% of the final grade.
- A class presentation on a primary-research paper: 15% of the final grade.
- An independent project, composed of both an oral presentation and a final paper (in lieu of a final exam): 30% of the final grade.
- Class participation: 20% of the final grade.

The distribution of final letter grades will not be curved. Letter grade divisions will be drawn at the instructors discretion, and will differ for graduate students and undergraduates.

Points obtained on assignments are halved if they are late, unless permission is granted at least the day before.

Students may have two unexcused absences without instructor permission. Beyond that, please inform the instructor if you need to be absent, or had to miss class unexpectedly. Unexcused absences will count against class participation.

Students are expected to be aware of and observe the UNC honor code (http://honor.unc.edu).

Background knowledge
Recommended places to read up for some of the background knowledge we will use in this class:

- Molecular biology and genetics: Ch 1 of Deonier et al (2005) or Gibson and Muse (2004)
- Probability and statistics: Chs 2 and 3 of Deonier et al (2005)
- Algorithms: Ch 2 of Jones and Pevzner

These books are available on reserve at the House Library.