

BIO337 Systems Biology/Bioinformatics (course # 50524)
Spring 2014 **Tues/Thurs** **11 – 12:30 PM** **BUR 212**

Instructor: Prof. Edward Marcotte marcotte@icmb.utexas.edu
Office hours: Wed 11 AM – 12 Noon MBB 3. 148BA

TA: Dakota Derryberry dakotaz@utexas.edu
Office hours: Mon 4-5 PM/Fri 9 – 10 AM MBB 3.304 TA Phone: 512-232-2459
Course web page: http://www.marcottelab.org/index.php/BIO337_2014

Open to upper division undergraduates in natural sciences and engineering.

Prerequisites: Biochem I or equivalent (e.g. CH339J or CH339K/L), basic familiarity with molecular biology.

An introduction to systems biology and bioinformatics, emphasizing quantitative analysis of high-throughput biological data, and covering typical data, data analysis, and computer algorithms. Topics will include introductory probability and statistics, basics of Python programming, protein and nucleic acid sequence analysis, genome sequencing and assembly, proteomics, synthetic biology, analysis of large-scale gene expression data, data clustering, biological pattern recognition, and gene and protein networks.

** Note that this is not a course on practical sequence analysis or using web-based tools (although we'll use a few), but rather on algorithms, exploratory data analyses and their applications in high-throughput biology. **

Most of the lectures will be from research articles and slides. For sequence analysis, there will be an **Optional text:** *Biological sequence analysis*, Durbin, Eddy, Krogh, Mitchison, Cambridge Univ. Press (available from Amazon, used from \$21.36)

For biologists rusty on their stats, *The Cartoon Guide to Statistics* (Gonick/Smith) is very good (really!).

We will also be learning some Python programming. The following is highly recommended:

Python programming for beginners: <http://www.codecademy.com/tracks/python>

Online homework will be assigned and evaluated using the free bioinformatics web resource Rosalind (<http://rosalind.info/faq/>). **Enroll here:** <http://rosalind.info/classes/enroll/32e1671726/>

No exams will be given. Grades will be based on online homework (counting 30% of the grade), **3 problem sets** (given every 2-3 weeks and counting 15% each towards the final grade) **and a course project** (25% of final grade), which will be collaborative (~4 students/project). The course project will consist of a research project on a bioinformatics topic chosen by the students (with approval by the instructor) containing an element of independent computational biology research (e.g. calculation, programming, database analysis, etc.). This will be turned in as a link to a web page. **The final project is due by midnight, April 28, 2014.**

All projects and homework will be turned in electronically and time-stamped. No makeup work will be given. Instead, all students have 5 days of free “late time” (for the entire semester, NOT per project, and counting weekends/holidays). For projects turned in late, days will be deducted from the 5 day total (or what remains of it) by the number of days late (in 1 day increments, rounding up, e.g. 10 minutes late = 1 day deducted). Once the full 5 days have been used up, assignments will be penalized 10 percentage points per day late (rounding up), e.g., a 50 point assignment turned in 1.5 days late would be penalized 20%, or 10 points.

Students are welcome to discuss ideas and problems with each other, but **all programs, Rosalind homework, and written solutions should be performed independently** (except the final collaborative project).

We'll cover the following topics, roughly in this order:

BASICS OF PROGRAMMING

A Python programming primer for non-programmers

Guest lecture (28-Jan): AppSoma: a collaborative cloud computing environment for UT students

Zack Simpson (CSSB; lead software architect, AppSoma)

BIOLOGICAL SEQUENCE ANALYSIS

Substitution matrices (BLOSSUM, PAM) & sequence alignment

Protein and nucleic acid sequence alignments, dynamic programming

Sequence profiles

BLAST! (the algorithm)

Biological databases

Markov processes and Hidden Markov Models

Gene finding algorithms

GENOMES, PROTEOMES, & "BIG BIOLOGY"

Guest lecture (18-Feb): Next- (& next-next-) generation DNA and RNA sequencing

Prof. Vishwanath Iyer (CSSB & Dept. of Molecular Biosciences)

Guest lecture (20-Feb): Genome engineering

Dr. Christopher Yellman (CSSB)

Genome assembly & how the human genome was sequenced

An introduction to large gene expression data sets

Clustering algorithms, hierarchical, k-means, self-organizing maps, force-directed maps

Classifiers, k-nearest neighbors, Mahalanobis distance

Promoter and motif finding, Gibbs sampling

Principal component analysis and data transformations

NETWORK & SYNTHETIC BIOLOGY

Guest lecture (1-Apr): Overview of mass spectrometry shotgun proteomics

Dr. Daniel Boutz (CSSB)

Guest lecture (3-Apr): Mapping universally shared animal protein complexes

Blake Borgeson (CSSB & Inst. of Cell & Molecular Biology)

Biological networks: metabolic, signaling, graphs, regulatory

Properties of biological networks

Network alignment and comparisons, network organization

Designing, simulating, and building gene circuits

***** THE FINAL PROJECT IS DUE by midnight, April 28, 2014 *****

The final two class days will be devoted to presenting your projects to the rest of the class.

Several lectures (indicated above) will be given by researchers from the University of Texas Center for Systems and Synthetic Biology (CSSB) on topics at the cutting edge of systems biology and bioinformatics.

Note that there is NO CLASS over spring break (March 11 & March 13).