

# CSCI1810: Computational Molecular Biology

## Course Information

- **Semester:** Fall 2015
- **Instructor:** Ben Raphael, CIT 447, braphael@cs.brown.edu
- **Time:** Tuesdays & Thursdays, 10:30-11:50 am
- **Location:** CIT 241 (SWIG Boardroom)
- **Website:** <http://cs.brown.edu/courses/csci1810/>
- **Piazza:** <https://piazza.com/brown/fall2015/csci1810/home>
- **TAs** (cs181tas@cs.brown.edu)
  - Younhun Kim (yk27) (HTA: cs181headtas@cs.brown.edu)
  - Samier Merchant (samercha)
- **Description**
  - High-throughput experimental approaches now allow molecular biologists to make large-scale measurements of the three fundamental molecules in the cell: DNA, RNA, and protein. These datasets are too large for manual analysis and demand computational techniques. This course introduces algorithms for analyzing DNA, RNA, and protein sequences, including: sequence comparison and alignment; molecular evolution and comparative genomics; DNA sequencing and assembly; recognition of genes and regulatory elements; and RNA structure and protein interaction networks. This course demonstrates how to model biological problems in terms of computer science. Students will also implement some of the algorithms and test on biological data.
- **Prerequisites**
  - One of: CS16, CS18 or CS19. Recommended: CS22, or some other course that introduces concepts from discrete math and probability.
  - Course overrides are available at the instructor's discretion.
- **Textbook**
  - **An Introduction to Bioinformatics Algorithms** by Jones and Pevzner (available in the bookstore, and online from the Brown library)
- **Piazza**
  - All students are required to enroll on Piazza. It will serve as a centralized location for all questions & discussion. Students are encouraged to ask questions! For any concerns regarding grading or administrative issues, please email the TA list instead.

- **Grading**

- Homeworks: 30%
- Programming projects: 30%
- Midterm (In class): 20%
- Final (Take home): 20%

- **Assignments**

- Assignments in this class will include 5 written problem sets and 3 programming projects in which you will explore basic bioinformatics algorithms covered in class. There will be an in-class midterm and a take-home final. All assignments will be posted on the course website and on Piazza.
- No late homework will be accepted without prior consent from course staff.
- Assignments will be handed in electronically, according to the instructions on the assignment handout.
- Students are strongly encouraged to use Julia for programming assignments. Java and Python are also permissible, but are not supported by the TAs. Support code and instructor assistance for programming assignments will be provided exclusively in Julia.
- Grading of written assignments will be based on both the accuracy of your answers and the work you show. Grading of programming assignments will primarily be based on correctness on test cases. Less emphasis will be based on coding style and performance, as long as they are within reasonable bounds. However, you are more likely to get partial credit for incorrect code if it is nicely formatted, thoroughly commented, and easily readable.

- **Collaboration Policy**

- There are three different policies regarding collaboration on written homework assignments, programming assignments, and exams. **Please read and note carefully the differences in these policies.**
- Problems on written homework assignments may be discussed with other students in the class. However, each student must independently write a solution to each problem. This written solution should reflect your own understanding of the problem. Written solutions from collaborating students should not be identical. **You are required to list any collaborators on each problem.**
- Programming projects are semi-collaborative. You may discuss the assignments with other students in the class and compare output on test cases. However, **you must write your own code and not examine code from other students.**
- The midterm and the final are **strictly non-collaborative**. This means that you must not discuss the problems on them with anyone besides the course staff. All solutions must be your own.

## Course Topics

- Pairwise Sequence Alignment
  - Global and Local Alignment, Affine Gap Penalties

- Space and time efficient alignment
  - Multiple sequence alignment dynamic programming
  - Star alignment and guide trees
- RNA Structure, Nussinov algorithm
- Phylogenetic Trees
  - Parsimony and Sankoff's Algorithm
  - Additive Distances
  - Neighbor Joining
  - Perfect Phylogeny
- DNA Sequence Assembly
  - Hamiltonian paths
  - Eulerian paths and de Bruijn Graphs
- RNA-Seq
- Pattern Matching
  - Exact pattern matching and suffix trees
  - Burrows-Wheeler transform
  - Approximate pattern matching and BLAST
  - Motif-finding
- Hidden Markov Models
- Networks