CSCI 1820
Algorithmic Foundations of Computational Biology
Spring 2020

Professor Sorin Istrail
Department of Computer Science
Brown University

January 16, 2020

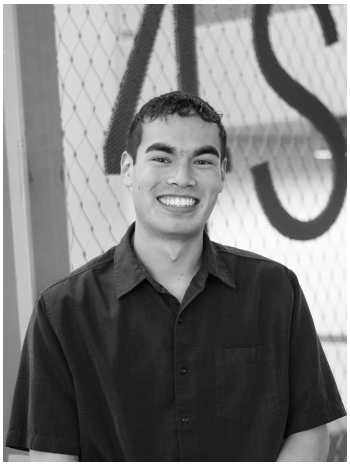
Course meeting: T/Th 2:30–3:50 PM, CIT 241 (SWIG)

Course website: `http://cs.brown.edu/courses/csci1820/`

TA email lists: `cs1820tas@lists.brown.edu`; `cs1820headtas@lists.brown.edu`

Staff:

Professor Istrail
`sorin_istrail@brown.edu`



Elliot Youth (HTA)
`elliot_youth@brown.edu`

Hersh Gupta (UTA)
`hersh_gupta1@brown.edu`

1 Course Description

Welcome to CS 182: Algorithmic Foundations of Computational Biology! The aim of this course is to provide mathematical and computer science foundations, as well as biological insights, for numerous seminal algorithms in the field of computational biology.

Course Topics

The course is organized into the following chapters:

Chapter 1: The BLAST Algorithm and Karlin-Altschul Statistics

Chapter 2: Genome Assembly Algorithms and Lander-Waterman Statistics

Chapter 3: Hidden Markov Model (HMM) Algorithms: The Learning Problem

Chapter 4: Spectral Graph Theory Algorithms: Clustering and Spectral Bounds

Chapter 5: Coalescent Theory and Ancestral Recombination Graph (ARG) Algorithms

Chapter 6: Gene Set Enrichment Analysis (GSEA) Algorithms

Chapter 7: Linear Time Suffix Tree Algorithms

Each chapter is devoted to a class of fundamental computational genomics problems, involving analysis of DNA, RNA and protein sequences as well as functional molecular biology. Our journey through each chapter is driven by a set of "beautiful" algorithms which will be presented with their theoretical algorithmic foundations in comprehensive analytical detail. "Beautiful" algorithms are rigorous, practical and elegant, yet also intuitive enough to be successfully understood and implemented. The algorithms covered in this course draw upon both seminal historical insights and state-of-the-art theory to solve the computational problems presented in each chapter. We will trace the origin, methodology and theory of each such algorithm, from biological impetus and statistical bases to underlying data structures and mathematical optimization, and develop a robust understanding of and appreciation for their value and purpose. Coursework will guide you on a journey through both historical and modern genomics, and provide you with advanced insights into the algorithmic tools of today's bioinformaticians and computational biologists.

Prerequisites

Required: any intro CS sequence, and CS 181 (Computational Molecular Biology).

Recommended: CS 22 (Introduction to Discrete Structures and Probability), or another course which introduces concepts from discrete math and probability theory.

Course overrides are available at the professor's discretion.

2 Course Format

CS 182 lectures will be held on Tuesdays and Thursdays 2:30–3:50 PM in SWIG (CIT 241). You are expected to attend all classes. Contributing to class-wide lecture notes throughout the semester is required and will constitute a portion of your class participation grade.

Assignments

Homeworks and projects will alternate throughout the semester. Each homework (HW) will focus on the algorithmic theory and mathematical/biological basis of the current chapter, and you will generally be given 1 week to complete these assignments. Each project (PR) will involve implementing algorithms discussed in lecture in the programming language of your choice, and you will generally be given 2 weeks to complete these assignments. There will be one in-class midterm exam and a take-home final exam.

Grading

Grades will be determined as follows:

Category	Percentage
Class participation	5%
Homeworks	30%
Projects	30%
Midterm exam (in-class)	15%
Final exam (take-home)	20%

Grades will be determined by your overall performance according to these metrics. At the end of the class, a *Pastiche Pie* award will be given to the student(s) with the overall most impressive performance in the class as judged by the TAs and the professor. All final grades will be determined by the professor.

Literature

There is no textbook for this course. However, suggested readings will be provided on the course website to complement and enhance the lecture content of the class.

Graduate Credit

In addition to all assignments listed above, graduate students must complete a final project selected in consultation with the professor to receive graduate credit. Undergraduates may choose to complete an optional final project for extra credit. Details regarding the final project assignment will be made available midway through the semester.

3 Course Policies

Collaboration Policy

In addition to Brown's Academic Code, CS 182 follows the collaboration policy below:

- You may discuss HW problems with other students in the class; however, all solutions must be written up independently and reflect your own understanding of the material.
- You may discuss PR assignments and compare output on test cases with other students in the class; however, all code must be written up independently. You may **not** examine code written by other students.
- You may **not** collaborate with anyone on the in-class midterm exam or the take-home final exam. You may only discuss the content of the exams with members of the course staff. All solutions must be entirely your own.

You will be required to accept this collaboration policy electronically at the beginning of the semester as a prerequisite for receiving grades for all subsequent assignments.

The course staff takes violations of the collaboration policy seriously and will prosecute with the standing committee on the academic code as necessary.

Late Handin Policy

You will receive 4 late days for use throughout the course. As all handins will be electronic, you may use these late days at your discretion, with two caveats:

- You may use a maximum of 2 late days per individual assignment
- You may **not** use late days on exams (only on HWs and PRs)

Extra late days will be penalized 15% each. Additional extensions on HWs and PRs will only be granted by the professor under extenuating circumstances and at his discretion. TAs cannot grant extensions.

Coursework Hours

The overall workload for this course (attending lecture, completing assignments and studying for exams) is estimated at 180 hours over the course of the semester.

Diversity, Inclusion, Accessibility & Accommodations

Brown is committed to the full inclusion of all students, and CS 182 strives to be a welcoming and inclusive place for the diverse student body. Please reach out to the professor if you have any concerns regarding inclusivity, accessibility, or SEAS accommodations.