COMP 5970/6970-004
Computational Biology: Genomics and Transcriptomics

Haynes Heaton

Spring, 2022

E-mail: HaynesHeaton@auburn.edu
Course Discord: https://discord.gg/my9hTgndcH
Office Hours: T/Th 1:45-3pm Shelby 3101F (Additional hours before projects are due)
Class: T/Th 12:30-1:45pm Shelby 1120

Course Description

This course is a broad introduction to computational genomics and transcriptomics. We will cover algorithms for analyzing DNA and RNA sequencing data. This will include sequence analysis, comparison, search, genome assembly, and analysis of RNA data including bulk and single cell sequencing. This will include statistical modeling and inference of biological data. We will also cover practical skills, software tools, and pipelines used in the bioinformatics industry.

Course Materials

- Lecture notes will be made available on Canvas.
- Lecture videos will be available on Canvas.
- Slides will also be available on Canvas.

Prerequisites/Corequisites

Prerequisites: Introduction to Algorithms, Discrete Math, Linear Algebra, or Departmental Approval

Course Objectives

Successful students:

1. Understand and implement core alignment algorithms

2. Recognize problems that are applicable to hidden markov models, and solve those problems
3. Can describe the properties of problems that make dynamic programming a viable algorithmic technique
4. Have familiarity with a broad set of biological concepts in genomics and transcriptomics
5. Can describe and run standard genomics pipelines and analyze and visualize results
6. Understands RNA and single cell RNAseq basic analysis

Course Structure

Assignments

Assignments will include problem sets and larger projects. Problem sets should take no longer than 2 hours. Projects will involve coding a solution to a problem in python (with the last project being a practical bioinformatics project using bash and open source software to analyze a dataset). The assignment sheet will make it clear what part of the project should be coded without specialized packages. General portions of packages such as numpy will always be allowed.

Grading

Your letter grade will follow the standard numerical scale – A: [90, 100], B:[80, 90), C: [70, 80), D: [60, 70), F: [0, 60)

1. **Exams** (35%): Midterm: 15%, Final: 20%
2. **Homeworks** 15%
3. **Class participation** 10%
4. **Projects** 40% (3 projects – grad students will have additional requirements for these projects)

The final exam will be given on the last day of the course and not on the day designated in the final exam schedule.

Course Policies

0.1 Attendance

Class attendance is not mandatory except for the exam days. But because class participation is 10% of your grade, regular attendance is highly encouraged. Phones should be on silent and any audible ring or notification will result in public shaming.

0.2 Excused Absences

Excused Absences only apply to the cases of mid-term and final examination. Students are granted excused absences from class for the following reasons: Illness of the student or serious illness of a member of the student?s immediate family, death of a member of the student?s immediate family, trips for student organizations sponsored by an academic unit, trips for University classes, trips for participation in intercollegiate athletic events, subpoena for a court appearance and religious holidays. Students who wish to have an excused absence from this class for any other reason must contact the instructor in advance of the absence to request permission. The instructor will weigh the merits of the request and render a decision. When feasible, the student must notify the instructor prior to the occurrence of any excused absences, but in no case shall such notification occur more than one week after the absence. Appropriate documentation for all excused absences is required.

0.3 Make-Up Policy

Arrangements to make up missed major examination (e.g. mid-term exams) due to properly authorized excused absences. Except in unusual circumstances, such as continued absence of the student or the advent of University holidays, a make-up exam will take place within two weeks from the time the student initiates arrangements for it.

0.4 ADA Policy

Students who need accommodations are asked to electronically submit their approved accommodations through AU Access and to make an individual appointment with the instructor during the first week of classes — or as soon as possible if accommodations are needed immediately. If you have not established accommodations through the Office of Accessibility, but need accommodations, make an appointment with the Office of Accessibility, 1228 Haley Center, 844-2096 (V/TT).

0.5 Academic Honesty

All portions of the Auburn University Student Academic Honesty code (Title XII) found in the Student Policy eHandbook at http://www.auburn.edu/student_info/student_policies/ will apply to this class. All academic honesty violations or alleged violations of the SGA Code of Laws will be reported to the Office of the Provost, which will then refer the case to the Academic Honesty Committee.

Schedule and weekly learning goals

The schedule is tentative and subject to change. The learning goals below should be viewed as the key concepts you should grasp after each week, and also as a study guide before each exam, and at the end of the semester. Each exam will test on the material that was taught up until 1 week prior to the exam.

Week 01, 01/10 - 01/14: Alignment Introduction

- Introduction to Dynamic Programming
- Sequence alignment algorithm(s)

Week 02, 01/17 - 01/21: Probability and Statistics

- Bayes Theorem
- Likelihoods vs probabilities and maximum likelihood estimation
- Distributions and the physical processes that produce them

Week 03, 01/24 - 01/28: Statistical modeling and inference

- How to model complex processes with distributions
- Statistical Inference via EM or numerical optimization

Week 04, 01/31 - 02/04: Biology Background

- Nucleic acid structure
- Double Helix
- Central Dogma of Molecular Biology
- Structure of the genome and genes

Week 05, 02/07 - 02/11: Further Biology Background

- RNA, tRNA, rRNA, amino acids, proteins
- DNA sequencing technologies

Week 06, 02/14 - 02/18: Alignment revisited

- Global, semi-global, local alignment
- Backtrace/Viterbi algorithm revisited
- Scoring model
- Protein sequence alignment and scoring

Week 07, 02/21 - 02/25: Hidden Markov Models (HMM)

- Observed and Hidden states
- Optimal state sequence via dynamic programming and Viterbi algorithm
- Forward-Backward algorithm for probabilistic underlying states

Week 08, 02/28 - 03/04: Alignment revisited again and **Exam 1**

- Sequence alignment as pair-HMM
- Forward-Backward algorithm for probabilistic alignments

Week 09, 03/07 - 03/11: Spring Break

Week 10, 03/14 - 03/18: Multiple sequence alignment

- Dynamic programming - Optimal but intractable
- Partial order alignment
- Guide trees and other popular MSA methods

Week 11, 03/21 - 03/25: Kmers

- Exact matching
- Counting
- Bloom filters
- De brujin graphs and genome assembly

Week 12, 03/28 - 04/01: Genome Indexes

- Prefix/suffix trees
- Suffix array
- Burrows wheeler transform and the FM-index (mostly for historical context)
- Minimizers

Week 13, 04/04 - 04/08: Bioinformatics practical

- File Formats
- Resequencing workflow
- Software tools

Week 14, 04/11 - 04/15: RNA sequencing analysis

- Gene counts and differential expression
- Single cell sequencing
- Cell barcode detection

Week 15, 04/18 - 04/22: RNA sequencing analysis continued

- Doublet detection
- Cell type clustering, annotation

Week 16, 04/25 - 04/29: Spillover week or special topics / research plug

1 Statements

1.1 Diversity & Inclusion

It is my intent that students from all diverse backgrounds and perspectives be well served by this course, that students? learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, religion, sexuality, disability, age, socioeconomic status, veteran status, ethnicity, race, and culture. All students in this course are expected to respect their fellow classmates and actively participate in fostering an inclusive learning environment. If you experience anything in this class that makes you feel uncomfortable, please bring it to my attention and we will formulate a response. If you would prefer to remain anonymous you may complete a Bias Incident Report which will maintain your confidentiality at: http://studentaffairs.auburn.edu/bert/submit-a-report-of-bias/

Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups.