CS 5124: Algorithms in Bioinformatics
Syllabus
Spring, 2019

1 General Course Information

CRN	19518
Meeting Time	8:00 AM–9:15 AM; Tuesday/Thursday
Classroom	McBryde 238
Midterm Exam	In class, March 7, 2019
Final Exam	7:45 AM–9:45 AM, May 10, 2019

Instructor: Lenwood S. Heath

- **Office:** 2160J Torgersen Hall
- **Office Hours:** 1:00–3:00 Tuesday/Thursday
- **Email:** `heath@vt.edu`

Web Site: `http://courses.cs.vt.edu/~cs5124/Spring2019/index.php`

Piazza Signup: `http://piazza.com/vt/spring2019/cs5124`

Canvas: `https://canvas.vt.edu/`

CentOS Linux Server: `cs5045.cs.vt.edu`; if you would like an account, send me email

Prerequisites:

- CS 2604, Data Structures and File Management, or equivalent
- An undergraduate course in algorithms is not required, but is highly desirable.
- **Corequisite:** GBCB 5314 — Biological Paradigms for Bioinformatics — or equivalent coursework in genetics and molecular cell biology

Required Textbook: Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Dan Gusfield. Cambridge University Press, 1997. ISBN: 0521585198.

Books On Reserve: For current list, see class web site.

2 Course Description

This course emphasizes algorithms to solve problems found in biology, especially molecular biology. A variety of current problems in computational molecular biology will be introduced, investigated, analyzed for computational complexity, and solved with efficient algorithms, when feasible. A number of such problems will be shown to be NP-complete or other evidence of their difficulty will be presented.

3 Grading Policy

Grading for the course is on a 1000-point scale, with the points distributed as follows:

Homework assignments: 10 at 60 points each	600
Midterm exam: March 7, 2019	150
Final exam: May 10, 2019	250

A typical homework assignment consists of two or three problems, posted on the course web site approximately one week before the due date. During the first week of class, students will be organized into **discussion groups** of two (or perhaps three) students each. These discussion groups are meant to help students learn the subject by providing **partners** to discuss material with. Also, with his or her partner(s), each student may discuss the homework problems and devise solution strategies. However, each student must write up his or her solutions individually. It is acceptable for a student to proofread and offer feedback on a partner's solutions before they are submitted.

All homework must be prepared with LaTeX[1] or other word processing system, printed, and submitted in class on the due date[2]. Use of LaTeX is **strongly** recommended, though not absolutely required. **No late homework will be accepted.**

4 Readings

For most classes, there is a reading assignment (see Section 7) to be completed by class time. Each assignment consists of sections in the textbook.

5 Ethics

The Honor Code applies. All work submitted must be the student's own work. A student may solicit help with homework assignments only from the instructor and his or her partner(s). A student must complete the take-home midterm and final exams without any outside help of any kind.

[1] See LaTeX resources on the course web site.

[2] See due dates on the Calendar on the course web site.

6 Announcement

If any student needs special accommodations because of a disability, please contact the instructor during the first week of classes.

7 Course Schedule

Dates	Reading Assignment	Topics
	January	
1/22–1/25	Chapter 1	Exact matching: first algorithms; fundamental preprocessing
1/28–2/1	Sections 2.1–2.3	Exact matching: classical algorithms; Boyer-Moore and Knuth-Morris-Pratt
	February	
2/4–2/8	Section 3.4	Exact set matching; keyword trees
2/11–2/15	Chapter 5; Sections 6.1–6.2	Suffix trees; linear-time construction of suffix trees
2/18–2/22	Sections 7.2, 7.4–5, 7.11–12	Selected applications of suffix trees; exact set matching again, longest common substring, DNA contamination, and finding repeats
2/25–3/1	Chapters 10 and 11	Core string edits, alignments, and dynamic programming
	March	
3/4–3/8	Chapter 11	More dynamic programming
3/7	**Midterm**	**In class:** Topics through Chapter 11
3/11–3/15	Spring Break	
3/18-3/22	Chapter 14	Multiple string comparison and multiple sequence alignment
3/25–3/29	Chapter 15	Sequence databases and searching — BLAST and scoring matrices
	April	
4/1–4/5	Chapter 17	Reconstructing phylogenetic trees
4/8–4/12	Handouts	Models of molecular evolution
4/15–4/19	Handouts	Models of molecular evolution continued
4/22–4/26	Handouts	Some NP-completeness proofs
4/29–5/3	Handouts	Advanced topics in phylogenetic reconstruction
	May	
5/7	**Last Day of Class**	Review for final
5/10	**Final**	**7:45 AM–9:45 AM:** Comprehensive exam

End of Syllabus