

# Comp 555 - Bioalgorithms: Spring 2021

## COVID-19 statement:

This 2021 spring semester offering of Comp555 will begin as an online course with lectures and office hours offered using Zoom. Zoom links will be provided on the course website for lectures, office hours, and exams. If any student is unable to attend due to severe symptoms, exceptions to the late problem set policy and alternate exam times will be considered on a case-by-case basis, after notifying the instructor at least 24 hrs in advance. Please, be prepared to provide medical documentation. The path to a successful semester relies on every person in our community taking responsible actions that include mask-wearing, social spacing, frequent hand washing, and getting vaccinations.

## Bulletin Description

Bioinformatics algorithms. Topics include DNA restriction mapping, finding regulatory motifs, genome rearrangements, sequence alignments, gene prediction, graph algorithms, DNA sequencing, protein sequencing, combinatorial pattern matching, approximate pattern matching, clustering and evolution, tree construction, Hidden Markov Models, randomized algorithms.

## General Course Info

Term:	SPRING 2021
Department:	COMP
Course Number:	555
Section Number:	001
Time:	T Th, 9:30 - 10:45
Location:	SN 014
Website:	<a href="http://www.csbio.unc.edu/mcmillan/?run=Comp555S21">http://www.csbio.unc.edu/mcmillan/?run=Comp555S21</a>

## Instructor Info

Name:	Prof. Leonard McMillan
Office:	SN 316
Email:	<a href="mailto:mcmillan@cs.unc.edu">mcmillan@cs.unc.edu</a>
Phone:	919-590-6078
Web:	<a href="http://www.cs.unc.edu/~mcmillan">http://www.cs.unc.edu/~mcmillan</a>
Office Hours:	W 2pm-4pm

At the time of this writing, there are no assigned TAs or LAs for Comp555. Should this change, their names and supplemental office hours will be provided on the course website.

## Textbooks and Resources

There will be no textbook semester. In the past, I have used the following, which you might find useful as supplement reading material:

Bioinformatics Algorithms: An Active Learning Approach  
by Phillip Compeau and Pavel Pevzner  
Active Learning Publishers © 2014, ISBN: 978-0990374602.

An Introduction to Bioinformatics Algorithms  
by Neil Jones and Pavel Pevzner  
The MIT Press © 2004, ISBN: 978-0262101066

## Course Description

Computational methods are fueling a revolution in the biological sciences. Computers are already nearly as indispensable as microscopes for analyzing and interpreting biological data. As a result, two new multidisciplinary fields, bioinformatics and computational biology, have emerged. This course will explore the computational methods and algorithmic principles driving this revolution. It will cover basic topics in molecular biology, genetics, and proteomics. The course also addresses basic computational theory and algorithms including asymptotic notation, recursion, divide-and-conquer approaches, graph algorithms, dynamic programming, and greedy algorithms. These fundamental concepts from computer science will be taught within the context of motivating problems drawn from contemporary biology. Example biological topics include sequence alignment, motif finding, gene rearrangement, DNA sequencing, protein peptide sequencing, phylogeny, and gene expression analysis.

This course is suitable for both computer science and biology students at both undergraduate and graduate levels. Students taking this course should have some programming experience in a modern language.

## Target Audience

This course is intended for advanced undergraduate computer science majors and graduate students who are interested in exploring the challenges of bioinformatics and computational biology. It is also well suited for graduate students from Biostatistics and the Curriculum in Computational Biology who would like to better understand the algorithms underlying the common bioinformatics analysis tools.

## Prerequisites

COMP 410 and one of COMP 283 or MATH 381 are course prerequisites. However, I would like to offer the course to a multidisciplinary audience. Thus, I allow students, with permission, to take COMP 555 without the standard prerequisites. However, every student is expected to have taken a first and second course in programming equivalent to COMP 110/116 and COMP 401 offered in our department.

## Goals and Key Learning Objectives

Comp 555 emphasizes the design, performance, and application of algorithms. All algorithms are motivated by problems from modern biology. A major objective of Comp 555 is to develop a student's intuition for what is the appropriate algorithm to use in a given setting, what is a correct algorithm, and how to analyze the performance and scalability of an algorithm. Comp 555 also exposes the methods underlying many of the common tools used in modern quantitative biology.

## Course Requirements

Students will be assigned problem sets with associated programming assignments to be completed on their own computers using a programming language designated by the instructor (Python in this offering). There will be six homework assignments, a midterm, and a final exam.

## Key Dates

Midterm: March 18 (during class)  
Final Exam: May 11 (Tuesday, 8am-11am)

## Grading Criteria

The final grade will be based on the follow weighting factors:

6 – Problem Sets w/Programming Exercises	40% (lowest is dropped)
1 – Midterm Exam	25%
1 – Final Exam	25%
Many unannounced in-class exercises/quizzes	10% (lowest 2 dropped)

## Course Policies

This section should address the following:

- Attendance is expected, but no roll will be taken
- No late problem sets will be accepted. Submissions of problem sets are on-line and are “due” just before midnight on their due date.

The course final is given in compliance with UNC final exam regulations and according to the UNC Final Exam calendar.

## Cheating and the UNC Honor Code

Collaboration on assignments is encouraged. However, what you hand in must be your own work. Good scholarship requires that all collaboration must be acknowledged. Thus, if you collaborate on the solution of a problem set, I expect that you list your collaborators prior to your answer for each problem. Turning in copied code (any function or fragment) written by some other person or found on the Internet is not collaboration. Discussing algorithmic design approaches and sharing test cases are good examples of collaboration.

Collaboration on tests (midterms, final) in any form is, of course, a violation of the Honor Code. This includes discussion of questions on a midterm, or final with students that have not yet taken the test.

Using any unauthorized information source, such as social media (Facebook, WeChat, Slack, Shared Google Docs, etc.), texting, or any other interactive means of contacting another person during an exam is a violation of the honor code.

### Course Schedule

A course schedule with handouts and materials from each lecture will be posted on the course website.

### Diversity and Inclusion Statement:

Comp 555 will attempt to provide an objective learning experience that respects all demographic groups and LGBTQ+ identifications. The instructor acknowledges that many aspects of computer science are tinged by a history that over represents a small subset of privileged voices. Furthermore, it is possible that materials could be presented in ways that expose both overt and implicit biases of both the field and the instructor. The instructor is committed to integrating a more diverse point of view in this course, and always encourages feedback on how to improve the course materials to make them more inclusive.

### Disclaimer

“The professor reserves the right to make changes to the syllabus, including due dates and test dates. These changes will be announced as early as possible.”