CSE3800/BME4800/CSE5800: Bioinformatics – Fall 2023

TuTh 3:30pm-4:45pm, CHM T309

This course is an introduction to the fundamental mathematical models and algorithmic techniques used in bioinformatics. Emphasis will be placed on modeling biological applications as graph-theoretic or mathematical optimization problems, and on designing, analyzing, and implementing efficient algorithms for solving these formulations. Covered algorithmic techniques will include exhaustive search, greedy algorithms, dynamic programming, divide-and-conquer, graph algorithms, combinatorial pattern matching, clustering, and randomized algorithms. Biological applications will include motif finding, sequence assembly, pairwise and multiple sequence alignment, genome rearrangement analysis, gene expression analysis, evolutionary tree reconstruction, and single-cell sequencing data analysis.

Prerequisites: BIOL 1107; CSE 1729 or CSE2050; and one of STAT 3025, STAT3345, STAT3375, or MATH 3160.

Instructor

Ion Măndoiu
ion@engr.uconn.edu
Office: ITE 261
Office Hours:
M/Tu/W/Th 11:30am-12:30pm

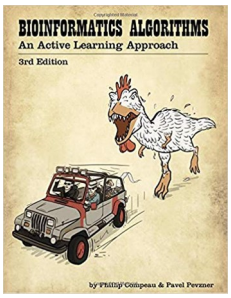
Teaching Assistant

Rye Howard-Stone
rye.howard-stone@uconn.edu
Office: ITE 219
Office Hours:
Fr 1-2pm

Course objectives

Upon completion of the course you will be able to:

- Translate biological questions into computational problems using mathematical models
- Design computer algorithms for answering relevant biological questions
- Use high-level programming languages to implement bioinformatics algorithms
- Evaluate efficiency and accuracy of existing bioinformatics software tools



Textbook

Phillip Compeau and Pavel Pevzner, *Bioinformatics Algorithms*, 3rd Edition.
Book website: https://www.bioinformaticsalgorithms.org

Additional materials will be distributed on Moodle.

Course website

We will use a course website hosted using Moodle at https://edx.engr.uconn.edu/. Please check this site regularly to access assignments, grades, and course materials including videos, handouts, etc. The Moodle site also integrates a discussion forum to ask class-related questions and communicate with the instructor and your peers. Please use this forum for general questions about the covered material and clarifications on the assignments. For questions that are specific to your own work you should contact the instructor directly.

Course work

The course will be employ a blended modality, consisting of both asynchronous and synchronous activities and several types of assignments. Asynchronous course content will be linked on Moodle prior to each class meeting and will include *interactive videos and quizzes*. It is essential that you review the asynchronous content and complete associated quizzes prior to each class meeting to ensure that you are prepared to actively participate in class discussions.
In addition to the interactive videos and quizzes, grading will be based on bi-weekly *theoretical homeworks* and *programming assignments* reinforcing the material covered in lectures along with a *final group project*. Solutions to both theoretical homeworks and programming assignments must be submitted in electronic format via Moodle. The recommended language for solving programming assignments is Python, but common programming language including Java, C, C++, and R can also be used. Solutions to the programming assignments will be automatically checked for correctness on a set of standard test cases, allowing you to receive immediate feedback and fix potential problems before the due date. The final group project is designed to give you the opportunity to study a bioinformatics application in more depth. Project requirements will include submitting several written reports and short presentations. Suitable final project topics include surveys of bioinformatics topics not covered in lectures, design and implementation of novel algorithms, and empirical comparisons of existing bioinformatics tools. Full final project details including a list of potential topics will be provided towards the middle of the semester.

Grade breakdown

Interactive videos & quizzes	10%
Theoretical homeworks	20%
Programming assignments	30%
Final project	40%

Late policy

Unless otherwise specified, interactive videos and quizzes are due before the beginning of class meetings. Theoretical homeworks and programming assignments are due by midnight on the specified due date. Late submissions are allowed for up to three days with a 10% penalty for each late day. To accommodate unforeseen circumstances that may prevent timely submission, the lowest theoretical homework and programming assignment scores will be dropped from the final grade calculation.

Academic integrity

You are expected to adhere to the highest standards of academic integrity. For homework assignments and programming projects you may discuss ideas and concepts with others, but must not share written solutions or code. All submitted solutions must be your own work. *Submitting solutions from various web sources as your own is considered academic misconduct and will be sanctioned according to the University's Academic Integrity Policy.*

Students with disabilities

If you have a documented disability for which you are or may be requesting an accommodation, you are encouraged to contact the instructor and the Center for Students with Disabilities or the University Program for College Students with Learning Disabilities as soon as possible to ensure that such accommodations are implemented in a timely fashion.

Tentative Schedule

Meeting dates	Topic	Textbook chapter
Aug 29 & 31	Course intro; Finding DNA replication origins	Ch. 1
Sept 5 & 7	Motif finding	Ch. 2
Sept 12 & 14	Genome assembly	Ch. 3
Sept 19 & 21	Sequence alignment	Ch. 5
Sept 26 & 28	Sequence alignment	Ch. 5
Oct 3 & 5	Pattern matching	Ch. 9
Oct 10 & 12	Pattern matching	Ch. 9
Oct 17 & 19	Antibiotics sequencing	Ch. 4
Oct 24 & 26	Computational proteomics	Ch. 11
Oct 31 & Nov 2	Genome rearrangements	Ch. 6
Nov 7 & 9	Evolutionary tree reconstruction	Ch. 7
Nov 14 & 16	Gene expression analysis	Ch. 8
Nov 28 & 30	Randomized Algorithms	Ch. 10
Dec 5 & 7	Hidden Markov Models	Ch. 10