CSE3800/BME4800/CSE5800: Bioinformatics – Fall 2020

This course is an introduction to the fundamental mathematical models and algorithmic techniques used in bioinformatics. Emphasis will be placed on modeling computational problems arising in biology as graph-theoretic, statistical, or mathematical optimization problems, and on designing, analyzing, and implementing efficient algorithms for the latter. 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 covered will include motif finding, sequence assembly, pairwise sequence alignment, genome rearrangement analysis, gene expression analysis, and evolutionary tree reconstruction.

Prerequisites: BIOL 1107, CSE 1100 or 1010 or 1729, and either STAT 3025Q or STAT 3345Q.

Instructor

Ion Măndoiu
ion@engr.uconn.edu
Office: ITE 261
Office Hours:
    Tu/Wed/Th 12pm-1pm
    (in person or online at http://snp.engr.uconn.edu/b/ion-3bm-yp0)

Course Logistics

Meeting times: TuTh 3:30pm-4:45pm

In person participation: AUST 108

Remote participation: Links for two-way real-time interactive communication will be posted for each class meeting on the course website. Non-interactive, 90-second delayed video streaming will be available at https://classrooms.uconn.edu/classroom/aust-108/ as a backup.

Course website

We will use a course website hosted using Moodle at https://edx.engr.uconn.edu/. *Note that this is a local Moodle installation and you will not be able to login using your netid credentials.* Please check the course site regularly to access assignments, grades, and course materials including videos, handouts, etc. The Moodle site also includes discussion forums and chat rooms to ask class-related questions and communicate with the instructor and your peers. Please observe basic etiquette by keeping your postings polite, concise, and on-topic. Appropriate questions are 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.
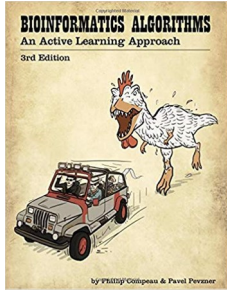
Textbook

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

All textbook chapters (but not charging stations) are available for free on the book website. Additional materials will be distributed on Moodle.

Course objectives

After completion, you will be able to:

- Translate biological questions into computational problems using mathematical models

- Employ advanced techniques to design computer algorithms for answering relevant biological questions

- Use high-level programming languages to implement computer programs meeting given requirements

- Evaluate efficiency and accuracy of existing bioinformatics software tools

Course work

The course will be delivered using a blended modality, consisting of both asynchronous and synchronous activities and several types of assignments. Asynchronous course content (short videos, book readings) will be linked on Moodle prior to each class meeting and will include **interactive exercises and quizzes**. It is essential that you review the asynchronous content and complete associated exercises prior to each class meeting to ensure you are able to participate in class discussions. Participation in synchronous class meetings can be either in person or via two-way real-time communication via a local installation of the BigBlueButton conferencing system. Non-interactive, 90-second delayed video streaming of class meetings will be available as a backup and recorded for post-meeting review.

In addition to interactive exercises and quizzes, grading will be based on theoretical homeworks and programming assignments reinforcing the covered material, as well as a final group project. Solutions to **theoretical homeworks** will be submitted in electronic format via Moodle. **Programming assignments** will be submitted electronically via the Rosalind site at rosalind.info. Rosalind is a repository of intellectually stimulating problems of varying difficulty that are extracted from real challenges of molecular biology. Solutions can be prepared using any high-level programming language. You will be asked to process a dataset generated by Rosalind on your own computer and then upload or copy-paste the solution to Rosalind along with your source code. Each submitted solution is automatically checked for correctness, allowing you to fix potential problems before the due date. The **final 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 group presentations. Suitable final project topics include surveys of bioinformatics topics not covered in the 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 exercises & quizzes	20%
Theoretical homeworks	20%
Programming assignments	30%
Final project	30%

Late policy

Interactive exercises and quizzes are due before the beginning of class. All other assignments are due by midnight on the specified due date. Late submissions are allowed for up to three days. Assignments that are more than three days late and make-up quizzes will only be allowed in extenuating circumstances, however, to accommodate unforeseen circumstances that may prevent timely submission, the lowest homework and programming assignment scores will be dropped from the overall grade calculation.

Academic integrity

You are expected to adhere to the highest standards of academic integrity. All submitted solutions must be your own work. For homework assignments and programming projects you may discuss ideas and concepts with others, but must not share written solutions or code. Use of published materials (including web resources) is allowed, but all sources should be explicitly acknowledged in your solutions. Violations will be reviewed and sanctioned according to university policies.

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 better ensure that such accommodations are implemented in a timely fashion.

Tentative Schedule

Meeting dates	Topic	Textbook chapter
Sept 1 & 3	Finding DNA replication origin	Ch. 1
Sept 8 & 10	Motif finding	Ch. 2
Sept 15 & 17	Genome assembly	Ch. 3
Sept 22 & 24	Sequence alignment	Ch. 5
Sept 29 & Oct 1	Sequence alignment	Ch. 5
Oct 6 & 8	Pattern matching	Ch. 9
Oct 13 & 15	Pattern matching	Ch. 9
Oct 20 & 22	Antibiotics sequencing	Ch. 4
Oct 27 & 29	Computational proteomics	Ch. 11
Nov 3 & 5	Genome rearrangements	Ch. 6
Nov 10 & 12	Evolutionary tree reconstruction	Ch. 7
Nov 17 & 19	Gene expression analysis	Ch. 8
Dec 1 & 3	Randomized algorithms	Ch. 10