

# CSE3800/BME4800/CSE5800: Bioinformatics – Fall 2021

**Instructor:** Ion Măndoiu, [ion@enr.uconn.edu](mailto:ion@enr.uconn.edu)

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 1729, and one of STAT 3025/3345/3375 or MATH 3160.

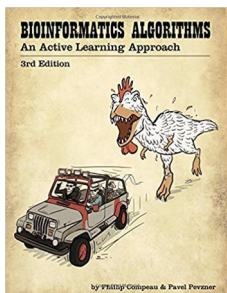
## Course Logistics

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

*Room:* TBD

## Course website

We will use a course website hosted using Moodle at <https://edx.enr.uconn.edu/>. Please check the course website 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

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

## 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 actively participate in class discussions. 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.

## Grade breakdown

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

## 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 with a 10% penalty for each late day. 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.

## Tentative Schedule

Meeting dates	Topic	Textbook chapter
Aug 31 & Sept 1	Course intro; Finding DNA replication origin	Ch. 1
Sept 7 & 9	Motif finding	Ch. 2
Sept 14 & 16	Genome assembly	Ch. 3
Sept 21 & 23	Sequence alignment	Ch. 5
Sept 28 & 30	Sequence alignment	Ch. 5
Oct 5 & 7	Pattern matching	Ch. 9
Oct 12 & 14	Pattern matching	Ch. 9
Oct 19 & 21	Antibiotics sequencing	Ch. 4
Oct 26 & 28	Computational proteomics	Ch. 11
Nov 2 & 4	Genome rearrangements	Ch. 6
Nov 9 & 11	Evolutionary tree reconstruction	Ch. 7
Nov 16 & 18	Gene expression analysis	Ch. 8
Nov 30 & Dec 2	Hidden Markov Models	Ch. 10
Dec 7 & Dec 9	Hidden Markov Models	Ch. 10