MAT-5932, MAT-4930
Computational Biology
CRN 13372, Section # 007, 3 CR
College of Arts and Sciences,
Department of Mathematics and Statistics

COURSE SYLLABUS

Semester: Spring 2021
Class Meeting Days: TR
Class Meeting Time: 2:00 – 3:15 pm
Class Meeting Location: Online
Instructor: Dr. Abdulmelik Mohammed
Office Location: NA
Office Hours: NA, video chat by appointment
Phone Number: +8139936411
Email: abdulmelik@usf.edu

I. Introduction

Sequencing technologies are generating massive amounts of biological data, as exemplified with the seminal achievement of the sequence of the human genome in 2001. In order to efficiently analyze the big data obtained from the sequencing projects, researchers need to employ algorithmic design principles that systematically exploit the search space of mathematically formulated computational biology problems. In this course, students will learn about algorithmic techniques that allow them to analyze the massive the information content in the biomolecules that are at the core of life's driving mechanism. They will learn to model questions in molecular biology mathematically and how to solve them algorithmically. Equipped with the algorithmic foundations and especially with their an experience gained on the application of algorithmic principles in molecular biology, students will be well prepared for applied research as bioinformatics analyst or computational biologist in research labs, industry players such as Illumina or governmental institutes such as the NIH's NHGRI.

II. Course Prerequisites

A basic course in Linear Algebra (e.g. MAS 3105) and a basic course in Discrete Mathematics (e.g. MAD-2104). Consult with instructor if unsure.

III. Organization

Lectures are held online via Microsoft Teams. We use the textbook "*An introduction to bioinformatics algorithms*" (see below for full bibliographic information) as a guide for the lectures. We cover about one algorithmic design principle per week, apart from the more general techniques which take two weeks. See the course schedule section for more information.

Algorithms are best learnt through exercise and practice. Students must work on problems to best learn the algorithm design principles. Thus, homework exercises are given for each chapter. Homework problems include both theoretical problems and computer exercises. Students must work on the problems and submit their solutions as the homework problems constitute half the grade for the course.

Some additional items students should note in the online learning environment that we have at the course:

1. Check Canvas and emails daily for announcements.
2. Follow the lecture. In the online learning environment, students should take care to minimize their digital distraction. Thus, they are encouraged to close other programs and apps before class begins.
3. When possible, plug in to a wired internet connection, rather than rely on Wi-Fi.

IV. Student Learning Outcomes

By the end of this course, students are:

a. expected to have learned about different algorithmic principles such as branch-and-bound, divide-and-conquer, dynamic programming, graph algorithms, clustering algorithms, hidden Markov models and randomized algorithms.
b. expected to have learned to model problems in molecular biology in algorithmic terms such as in motif search, gene prediction, sequence alignment and sequence assembly.
c. expected to have learned how to apply algorithmic techniques to solve problems in computational biology.
d. expected to have learned how to use the software tools commonly used in the analysis of biological sequences.

V. Textbook and Course Materials

- **Textbook**: Jones, Neil C. and Pevzner, Pavel A. *An introduction to bioinformatics algorithms*. MIT press, 2004. ISBN: 9780262101066, 0262101068.
- The textbook is available as an electronic book in the USF library.

- Lecture slides based on the textbook will be made available. All course materials will be made available through Canvas.

VI. Supplementary (Optional) Texts and Materials

- Clote, Peter G., and Rolf Backhofen. *Computational molecular biology: an introduction*. J. Wiley, 2000. Accessible through USF library online. Focuses more on structure prediction, but also algorithmically oriented.
- Haubold, Bernhard and Wiehe, Thomas. *Introduction to Computational Biology: An Evolutionary Approach*. Springer Science & Business Media 2006. Accessible through USF library online.
- Pevsner, Jonathan. Bioinformatics and Functional Genomics, Second Edition. John Wiley & Sons 2009. Accessible through USF library online, biology oriented.

VII. Grading Scale

We use the grading scale shown in the table below.

Grading Scale (%)	
90-100	A
80 - 89	B
70 - 79	C
60 - 69	D
0 - 59	F

VIII. Grade Categories and Weights

The grades are distributed for undergraduates as follows:

Assessment	Percent of Final Grade
Homework	30%
Project presentations	30%
Final exam	40%

For graduate students, the project presentations constitute greater portion of the grades, as shown in the table below:

Assessment	Percent of Final Grade
Homework	30%
Project presentations	40%
Final exam	30%

IX. Project Assignments

Students will work in small groups on projects. Students will choose a research paper among a list of high impact research papers in computational biology. They will read their chosen papers and critically analyze them. They will analyze whether the algorithms are correct, give optimal results, analyze their guarantees, and whether the space and run-time of the algorithms can be improved. They will then report their results and present it to the class. They will also have an option to work on their own research problem, if it is a computational biology problem, and present their findings.

X. Grade Dissemination

Grades will be disseminated through Canvas.

XI. Course Schedule

We cover about one algorithmic principle per week. Given their generality, dynamic programming, graph algorithms and hidden Markov models are given two weeks each. The course schedule is shown in the table below. The chapters refer to chapters in the textbook. Homework is assigned by the final day of the week the week the chapter has covered. For instance, HW1, based on the material from Chapter 3 is assigned by January 17. The homework due dates will be posted when the the homework are posted, but are generally by Friday the week after they have been posted.

Date	Mathematics topic	Applications	Reading	HW due
1/12	Administrative matters, motivation, course overview, molecular biology primer			
1/14	Algorithms and complexity: run-time and space requirement of algorithms, Big-O notation, polynomial-time algorithms, hardness of some problems		Chapter 3	
1/19	Exhaustive search: brute force vs branch and bound algorithms	Restriction mapping	Chapter 4	
1/21	Exhaustive search: search trees, motif-finding problem, median string problem	Regulatory motifs	Chapter 4	HW1
1/26	Greedy algorithms: sorting by reversal	Genome rearrangement	Chapter 5	
1/28	Greedy algorithms: approximation algorithms	Genome rearrangement	Chapter 5	HW2
2/2	Dynamic programming: edit distance, longest common subsequence	Sequence alignment	Chapter 6	
2/4	Dynamic programming: global and local sequence alignment	Sequence alignment	Chapter 6	HW3
2/9	Dynamic programming: exon chaining problem, spliced alignment problem	Gene prediction	Chapter 6	
2/11	Dynamic programming: free energy calculations	RNA folding		
2/16	Divide and conquer algorithms	Space-efficient alignment	Chapter 7	
2/18	Divide and conquer algorithms	Subquadratic alignment	Chapter 7	HW4
2/23	Graph algorithms: shortest superstring problem, overlap graphs	Sequence assembly	Chapter 8	
2/25	Graph algorithms: De Bruijn graphs	Sequence assembly	Chapter 8	HW5
3/2	Graph algorithms: spectrum graphs	Peptide sequencing	Chapter 8	
3/4	Graph algorithms: arc diagrams, matchings and trees	RNA structure modeling		
3/9	Combinatorial pattern matching: hash tables, suffix trees	Repeat finding	Chapter 9	
3/11	Combinatorial pattern matching: approximate matching	Database search	Chapter 9	HW6

3/16	Clustering and trees: hierarchical, k-means clustering	Gene expression analysis	Chapter 10	
3/18	Clustering and trees: trees	Phylogenetics	Chapter 10	HW7
3/23	Hidden Markov models: Markov chains		Chapter 11	
3/25	Hidden Markov models: structure	CG-islands	Chapter 11	HW8
3/30	Hidden Markov models: decoding algorithm		Chapter 11	
4/1	Hidden Markov models: parameter estimation	Profile alignment	Chapter 11	
4/6	Randomized algorithms: Gibbs sampling	Profile motif search	Chapter 12	
4/8	Randomized algorithms: random projections		Chapter 12	HW9
4/13 - 4/15	Spring break			
4/20	Project presentations			
4/22	Project presentations			HW10
4/27	Project presentations			
4/29	Project presentations			
5/4	**Exam (12:30 – 2:30)**			
5/7				**Project reports due**

* Note: The Schedule is subject to revision.

XII. Standard University Policies

Policies about disability access, religious observances, academic grievances, academic integrity and misconduct, academic continuity, food insecurity, and sexual harassment are governed by a central set of policies that apply to all classes at USF. These may be accessed at: https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx

XIII. Covid-19 Procedures

All students must comply with university policies and posted signs regarding COVID-19 mitigation measures, including wearing face coverings and maintaining social distancing during in-person classes. Failure to do so may result in dismissal from class, referral to the Office of Student Conduct and Ethical Development, and possible removal from campus.

Additional details are available on the University's Core Syllabus Policy Statements page: https://www.usf.edu/provost/faculty/core-syllabus-policy-statements.aspx.

XIV. Course Policies: Grades

Late Work Policy:

Late homework submissions will be penalized at 5 percent per late day.

Grades of "Incomplete":

The current university policy concerning incomplete grades will be followed in this course. For undergraduate courses: An "I" grade may be awarded to a student only when a small portion of the student's work is incomplete and only when the student is otherwise earning a passing grade. The time limit for removing the "I" is August 6th, 2021. For undergraduate students, this time limit may not exceed two academic semesters, whether or not the student is in residence, and/or graduation, whichever comes first. For graduate students, this time limit may not exceed one academic semester. "I" grades not removed by the end of the time limit will be changed to "IF" or "IU," whichever is appropriate.
For graduate courses: An Incomplete grade ("I") is exceptional and granted at the instructor's discretion only when students are unable to complete course requirements due to illness or other circumstances beyond their control. The course instructor and student must complete and sign the "I" Grade Contract Form that describes the work to be completed, the date it is due, and the grade the student would earn factoring in a zero for all incomplete assignments. The due date can be negotiated and extended by student/instructor as long as it does not exceed two semesters for undergraduate courses and one semester for graduate courses from the original date grades were due for that course. An "I" grade not cleared within the two semesters for undergraduate courses and one semester for graduate courses (including summer semester) will revert to the grade noted on the contract.

Group Work Policy:

Students can discuss solution ideas for homework problems with their peers, but solutions must be written up and submitted individually. Students are also strongly encouraged to think about the homework problems individually before discussing with their peers. The projects are done in assigned groups. Everyone must take part in a group project. All members of a group will receive the same score; that is, the project is assessed, and everyone receives this score. Once formed, groups cannot be altered or switched, except for reasons of extended hospitalization.

XV. Course Policies: Technology and Media

Email:

Course content is delivered through canvas. All other forms of communication are to be carried out via email.

Canvas:

This course will be offered via USF's learning management system (LMS) Canvas, available under my.usf.edu -> Learning and teaching tools -> Canvas. If you need help learning how to perform various tasks related to this course or other courses being offered in Canvas, please consult the Canvas help guides. You may also contact USF's IT department at (813) 974-1222 or help@usf.edu.

XVI. Course Policies: Student Expectations

Attendance Policy:

Students are encouraged to attend classes, but attendance is not mandatory. **In order to fulfill the first day attendance requirement, a class role will be taken the first day of class and it will be required from everyone to present a short introduction of oneself with camera on. This is the only time camera-on will be required. Anyone missing this activity will be automatically dropped from class.**

Professionalism Policy:

Students are expected to mute their audio in the class sessions in order to minimize noise from their environment. They can unmute anytime they wish to speak. Students who habitually disturb the class by talking, arriving late, etc., and have been warned may suffer a reduction in their final class grade.

XVII. Important Dates to Remember

All the dates and assignments are tentative and can be changed at the discretion of the instructor. Recall the following important dates, mostly from the registrar https://www.usf.edu/registrar/calendars/:

Spring classes begin:	Jan 11, 2021
Our class begins:	Jan 13, 2021
Drop/Add deadline:	Jan 15, 2021
Mid-term grading opens:	Feb 22, 2021

Mid-term grading closes:	Mar 5, 2021
Withdrawal deadline:	Mar 27, 2021
Last day of lecture:	Apr 29, 2021
Canvas closes for final grades:	May 11, 2021