Subject: INFO
Catalog Level: 390C
Course Name: Introduction to computational biology and bioinformatics

Meeting Days: M/W
Times: 75 min@2 days/wk
Start/End Times: 11:30-12:45
Capacity: ~50
Room: CS 142
Credits: 3

Instructor:
Anna Green
annagreen@umass.edu
Office: CS 348
Pronouns: she/her

TA:
Bryn Reimer
breimer@umass.edu
Office: n/a
Pronouns: she/her

Office Hours:

Anna Green	Wednesday 10:30– 11:30 AM	Office: CS 348
Bryn Reimer	Tuesdays 1 – 2 PM	Zoom: link

Course Description:
This course is designed to provide Informatics students with a broad, practical introduction to the field of computational biology and bioinformatics. The course will discuss at a conceptual level the models and algorithms used to analyze biological sequence data, as well as practical applications and data analysis. Background in biology is not assumed. The primary focus of the course will be analysis of genomic data, including sequence alignment, genome assembly, genome annotation, phylogeny construction, mutation effect prediction, population genetics, and genotype-phenotype association studies. Throughout the course, we will emphasize the unique challenges to working with biological data. Through lectures and hands-on programming problem sets, students will develop the necessary skills to tackle computational challenges in biology.
Course URL: https://people.cs.umass.edu/~annagreen/courses/info390c/

Learning Objectives:

- Gain an understanding of fundamental biological concepts and their relevance to computational biology.
- Understand a broad survey of current computational biology problems and their respective solutions, including the statistical and computational foundations of these solutions
- Acquire practical skills in programming and data analysis relevant to computational biology.
- Apply computational techniques in ways that demonstrate an understanding of bioinformatic data analysis

•

Prerequisites: For Informatics majors, a C or better in INFO248 is required. For CS majors, a C or better in COMPSCI240 and CS210 is required. Other majors are required to have earned at least a C in STAT 240 (or equivalent) and one of COMPSCI119, CICS108, or CICS110.

Textbook: None (Required readings will be free electronic materials provided by the instructor, including journal and conference articles and free textbooks)

Eligibility: This course is open to Junior and Senior-level students in the Informatics and Computer Science majors. Students from other majors with suitable background (see prerequisites) are eligible for a waiver.

Course Access:
We will be using Canvas to post course resources and lecture slides. We will be using Gradescope for coursework submissions and Piazza for communication with the course staff. Note that you may post anonymously to other students on Piazza, but instructors are able to see your identity.

The Gradescope entry code for this class is: (see Canvas)
The Piazza link is: (link posted on Canvas)

Grading Criteria:
60% problem sets. The problem set with the lowest grade will not be counted.
20% Midterm exam
20% Final exam
2% BONUS participation – either via attendance, Canvas, or office hours

Late work policy: Problem sets are due at 11:59pm on the day specified. You have four late days which you may use for any problem set or combination of problem sets, **with a maximum of two late days per problem set**. **Work that is more than 2 days (48 hours) late will not be graded.** You do not need to request to use late days, they are automatically tracked by Gradescope. However, **you are responsible** for knowing how many late days you have used. Remember that your lowest problem set grade will be dropped (*i.e.*, not counted towards your final average), even if that lowest grade is a 0.

If you require an exception to the policies stated above, you must either:

1. Have an accommodation from the Disability Services office which provides you with extra time on assignments, and have discussed the accommodation with me at the beginning of the semester
2. OR Contact your advisor (link for CS and INFO students) to discuss your need for an exception to course policies due to emergency circumstances, and have them contact me
3. OR, Contact the UMass Dean of Students Office, to discuss your need for an exception to course policies due to emergency circumstances, and have them contact me

Grading Scale: Grading is on a letter scale, listed below. Grades will be rounded to the nearest integer. The lowest problem set grade will be dropped (*i.e.,* not counted towards your final average problem set grade), even if that lowest grade is a 0 (exception: academic honesty violations).

Undergraduate students:

- A: 93-100%
- A-: 90-92%
- B+: 87-89%
- B: 84-86%
- B-: 80-83%
- C+: 77-79%
- C: 74-76%
- C-: 70-73%
- D+: 67-69%
- D: 60-66%
- F: below 60

Attendance policy: Regular attendance and participation is critical to developing an understanding of the material and achieving success in the course. There is no formal attendance policy.

Collaboration policy: In corporate and academic settings, it is encouraged that you collaborate with your colleagues and use available resources to complete work. In this spirit, I encourage you to discuss course material with your classmates and use online resources to extend your understanding. However, in order to fairly evaluate your understanding in a classroom setting, **I expect all versions of all assignments to be produced independently by you, in your own words (or code), and to reflect your own understanding of the problem.** Copying any component of an assignment from your fellow students or any other resource (including chatGPT and similar technology) is not permitted. Each assignment will include an option to describe any resources you used or fellow students you discussed with answers with, please answer these honestly.

Accommodation Statement: The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For further information, please visit Disability Services (https://www.umass.edu/disability/)

Academic Honesty Statement: Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst. Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent (http://www.umass.edu/dean_students/codeofconduct/acadhonesty/).

Title IX Statement: In accordance with Title IX of the Education Amendments of 1972 that prohibits gender-based discrimination in educational settings that receive federal funds, the University of Massachusetts Amherst is committed to providing a safe learning environment for all students, free from all forms of discrimination, including sexual assault, sexual harassment, domestic violence, dating violence, stalking, and retaliation. This includes interactions in person or online through digital platforms and social media. Title IX also protects against discrimination on the basis of pregnancy, childbirth, false pregnancy, miscarriage, abortion, or related conditions, including recovery. There are resources here on campus to support you. A summary of the available Title IX resources (confidential and non-confidential) can be found at the following link: https://www.umass.edu/titleix/resources. You do not need to make a formal report to access them. If you need immediate support, you are not alone. Free and confidential support is available 24 hours a day / 7 days a week / 365 days a year at the SASA Hotline 413-545-0800.

Course Inclusiveness Statement: It is important to me that this course be a welcoming environment to people of all backgrounds. My goal as an instructor is to help you learn the subject material in a way that is useful and empowering, and I believe that the best learning happens on a foundation of mutual respect.

This course will discuss subject matter related to human genetics and evolution. While in an ideal world science would be objective, the reality is that false beliefs about genetic differences between humans have been and continue to be used to justify racism and oppression. I will strive to teach accurate information about the data and techniques used in human genomics while also acknowledging their potential for problematic misinterpretations.

I anticipate that students in this interdisciplinary course may come from different intellectual backgrounds, and thus there may be substantial differences in terms of familiarity with the concepts. I expect you to be patient with your fellow students, and hope that you will help one another in learning the material. I also hope that you will ask questions when something in the course is confusing or unfamiliar! There are no stupid questions.

Please know that my door is open to you if you wish to bring any issues to my attention. I especially encourage you to come to me if something that I say or do as an instructor makes you feel unwelcome in the course.

Pronouns Policy Statement: Everyone has the right to be addressed by the name and pronouns that they use for themselves. You can indicate your preferred/chosen first name and pronouns on SPIRE, or indicate them to me directly. I will do my best to ensure that I address you with your chosen name and pronouns. Please remember: A student’s chosen name and pronouns are to be respected at all times in the classroom.

Modules:

Module 1: Introduction to computational biology and relevant concepts in biology
Overview of computational biology and its interdisciplinary nature
Introduction to basic biological concepts, including genes, genomes, and genetic variation

Module 2: Sequence search and alignment
Concepts underlying sequence alignment
Probabilistic foundations of alignment
Sequence scoring matrices
Needleman-Wunsch algorithm and Smith-Waterman algorithms
Fast sequence database search using BLAST
Models and Algorithms: Needleman-Wunsch algorithm

Module 3: Classifying variation in DNA
Introduction to DNA sequencing technologies and read formats
De novo DNA sequence assembly algorithms (for short-read sequencing)
Analysis of long-read sequencing data
Reference-based assembly algorithms and tools
Models and Algorithms: Building a classifier – logistic regression

Module 4: Annotating genome sequences with functional information
Introduction to genome annotation and its importance
Gene prediction algorithms: *ab initio* and comparative methods
PFAM domains
Models and Algorithms: Markov chains

Module 5: Sequence evolution, population genetics, and tests for selection
Basic concepts in population genetics: allele frequencies, genetic drift, natural selection
Detecting signatures of selection (eg, dN/dS, linkage disequilibrium)
Analysis of genomic variation in populations
Ethical considerations of human genome sequencing and population genetics
Models and algorithms: computing dN/dS

Module 6: Association studies
Introduction to GWAS and its role in identifying genetic factors associated with traits and diseases
Encoding DNA for use in association studies
Statistical methods for identifying significant genetic variants
Models and algorithms: linear regression and significance testing

Module 7: Population structure analyses
Understanding the origins of population structure
Computing principal components analysis for DNA
Models and algorithms: PCA

Module 8: Mutation effect prediction
Introduction to genetic mutations and their impact on proteins and non-coding regions
Functional consequences of genetic variants: missense, nonsense, frameshift, etc.
Prediction of protein structure and function changes
Prediction of functional changes to non-coding DNA
Models and algorithms: Building a classifier, part 2

Module 9: Protein structure prediction
Basics of protein structures and protein folding
Why predicting protein structure is hard
Prior solutions for protein structure prediction
Current solutions: AlphaFold
Note: there is no associated problem set for this module