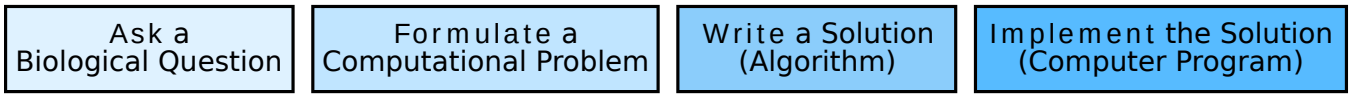
Introduction to Computational Biology

Professor: Anna Ritz, she/her/hers (aritz@reed.edu, Bio 200B)

Goals and Objectives

High-throughput technologies produce massive amounts of data, much too large to analyze by hand. The goal of this course is to learn how to analyze DNA, RNA, and protein sequences using computers. Through a combination of foundational examples and current research questions, this course aims to demystify computer science, molecular biology, and some of the ways they intersect.

1. You will know about the properties of DNA, RNA, and proteins, the relationships among these molecules, and some *biological questions* that have puzzled researchers.
2. You will know how to convert a biological question into a *computational problem* that can be solved using computers.
3. You will know how to read and understand solutions to computational problems, which will be formalized as a series of tasks (an *algorithm*). You will learn about general approaches for solving computational problems, and you will be able to apply these approaches to new problems you encounter.
4. You will know how implement the algorithms by writing *computer programs* in Python, which can be run and understood by others.



Using this general framework, you will learn how to to analyze DNA content, identify protein binding patterns, compare sequences, and discover variation within genomes. In the last two weeks, you will formulate your own sequence analysis problem, implement a solution, and present your findings to the class. By the end of this course, you will understand what types of biological questions can be investigated using computers, and what limitations computational methods impose on the understanding of biology.

---

1 Course Details

Time & Place: Lecture is MWF 10:05am-10:55am online. There are three lab sections:

Lab S1 W 1:25pm-3:15pm, ETC 211
Lab S2 Th 1:40pm-3:30pm, ETC 205
Lab S3 Th 3:44pm-5:35pm, ETC 205

Lab sections are in person, with an option to zoom in.

Textbook: Freely available online!
Bioinformatics Algorithms: An Active Learning Approach.
Edition 2, Volume 1. Phillip Compeau & Pavel Pevzner.
https://www.bioinformaticsalgorithms.org/read-the-book
You can also buy the book, and it is on two-hour reserve a the library. Different editions are fine, but page numbers may shift.

1.1 Collaboration & Support

We will use Slack for text-based communication and support (the tutors will also have access to the slack channel).
We will use repl.it for programming assignments (this will be taught).
There will be a Google Calendar with course content and deadlines that students can add to their own accounts. This will be updated regularly.
Students can nd additional challenge problems on Rosalind.

2.2 Schedule Overview

We will dive into ve main topics related to biological sequence analysis, and we will touch on a few more topics near the end of the semester. The course content is exible and the assignment timelines are approximate; refer to the Google Calendar on the course website and Moodle for up-to-date information.

Week	Dates	Topic	Assignment
Week 1	1/25{1/29	Intro to Intro & Central Dogma	HW1: Spreadsheet Warm-Up
Week 2	2/1-2/5	DNA & RNA	HW2: Python Practice
Week 3	2/8-2/12	Protein & Translation	HW3: Central Dogma
Week 4	2/15-2/19	Origin of Replication	
Week 5	2/22-2/26	DNA Replication	HW4: Frequent Words
Week 6	3/1{3/5	DNA Motifs	
Week 7	3/8{3/12	Formulating Motif Problems	
Week 8	3/15{3/19	*De novo* Assembly	HW5: Geedy Motif Finder
Week 9	3/22{3/26	*De novo* Assembly	
Week 10	3/29{4/2	Sequence Alignment	HW6: Sequence Alignment
Week 11	4/5{4/9	Sequence Alignment	

next morning. This should *not* be used to check whether your assignment is complete, but rather used when you are stuck on a bug.

I will provide an expected number of hours for each assignment. Keep track of the number of hours you spend. If it takes you signi cantly *less* time than expected, then attempt the challenge questions. If it takes you signi cantly *more* time, check in with me or attend a drop-in tutoring session.

Deadlines & Late Policy: Many programming assignment deadlines will depend on your lab section time, since they typically follow lab activities. Assignments will generally come out after your lab section and will generally be due before your lab section. You have two 24-hour extensions to use for any programming assignment (including using them back-to-back for one 48-hour extension). You must email me before the deadline to let me know you are using one or both of your extensions.

30% Quizzes: There will be a short (about 15{20 minute) take-home quiz at the end of each week that will assess knowledge about the biology topics and computational thinking. You complete these quizzes on your own over the weekend; they may be open or closed notes (which will be speci ed at the start of each quiz).

Deadlines & Late Policy: Quizzes will be due by Monday before lecture at 10:05am. If you get below 70% on your rst attempt, you may retake the quiz to get a maximum of 70% of the points. If you miss a quiz, you can complete it any time before the end of the semester to get a maximum of 70% of the points.

15% Participation & Engagement: Participation will be assessed in three ways:

Labs. Labs will not be graded, but will be used for participation points.

Code Conferences. There will be at least one Code Conference during the semester to discuss assignments to-date. No preparation beyond your assignment solutions are required.

Engagement with Material. Participation in the form of discussion, questions, and submitted videos will be considered.

5% Final Assignment: During Finals Week, you will make a video (e.g. by recording a Zoom session) where you explain how one of your programming assignments works. In this video, you will address the limitations of the algorithm in helping answer the underlying biological question.

---

4 Diversity and Inclusion

Bio131 is a combination of biology and computer science, and each eld has been claimed to be free of racism and prejudice. This is simply not true. Historically, biological discoveries and advances in computer science have been dominated by privileged voices, namely those of white men. Computational biology, while more recent of a eld, also lacks diversity along many important axes (including race, gender, nationality, class, sexuality, religion, ability, etc.). To foster an inclusive learning environment:

I acknowledge the bias in course materials that stem from systemic privilege, and I aim to make Bio131 content more inclusive each time I teach the course.

This course is designed for students with no programming experience, but you may come to Bio131 with varying levels of preparation in both biology and computer science. It is important to remember that this is an introductory course, and my priority is to teach from the ground up. Use the Collaboration Policy as a way to work with others and help them get comfortable with programming concepts.

We will be attending two seminars hosted by the Black Women in Computational Biology Network