# 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.

Ask a Biological Question	**Formulate** a Computational Problem	**Write** a Solution (Algorithm)	**Implement** the Solution (Computer Program)
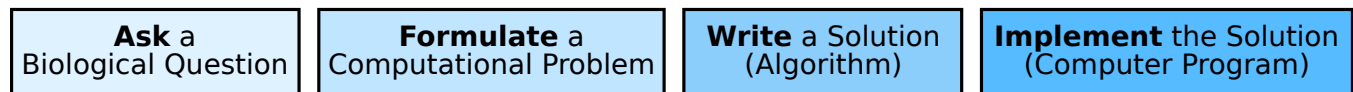
Using this general framework, you will learn how to to analyze DNA content, identify protein binding patterns, compare sequences, and discover variation within genomes. In the last two weeks, you will formulate your own sequence analysis problem, implement a solution, and present your findings to the 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

Office Hours: Wed 3:00pm-4:00pm & Th 10:00am-11:00am **online**. Drop in during these times to talk about anything (class related or otherwise), no appointment needed. Email Anna to set up an appointment if you cannot make office hour times.

Tutors: There are two Bio131 tutors with posted DoJo drop-in tutoring hours. They are also available for one-on-one tutoring. You can find their hours on the Moodle page as well.

- Iki Edreva, any pronouns – interdisciplinary biology/art junior
- Tayla Isensee, she/her/hers – biology senior

Collaboration Policy: You can work with other students on any programming assignment, provided you note that you worked together. This policy is notably different from other courses (e.g., CSCI121). More details are available in the Collaboration Policy.

Academic Accommodations: Please discuss any documented accommodations with me, especially those that include extensions on assignments. In the world of hybrid and remote-learning, please let me know if there are useful supplements to course content (e.g. closed captioning). Disability Accommodation Notification Letters can be obtained from Disability & Accessibility Resources. All discussions will remain confidential.

Obligated Reporting: I am happy to talk with you about anything that concerns you, but note that I (and many faculty) are *obligated reporters*, who must report possible violations of the Title IX and/or DHSM policies, which govern discrimination and harassment on the basis of sex, gender, and gender identity. If you need additional support, please reach out to Assistant Dean for Student Support Britt Hoover at hooverb@reed.edu or careteam@reed.edu.

2 Course Structure

Lectures. Bio131 will have synchronous lectures on Monday, Wednesdays, and Fridays (10:05–10:55) via Zoom. Lectures will be a mix of biology concepts and programming concepts.

Labs. Bio331 labs are ∼2-hours long, and are in person. Labs are designed to reinforce programming concepts, which will help you with the programming assignments. Students will work in small groups during the lab time unless there are unavoidable time conflicts. There is an option for attending the labs remotely, which is designed for students who cannot come in person. Talk with me if you would prefer to attend the labs remotely, rather than in person.

2.1 Communication and Technology

The official communications about all course materials and assignments will be done through the Moodle page. The course content may be adjusted (for example, if we move completely online) - you should check Moodle and your email every day to stay updated. Additionally,

- We will use Zoom for video conferencing. The zoom link is on the Moodle page.

- 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 find additional challenge problems on Rosalind.

2.2 Schedule Overview

We will dive into five 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 flexible 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	
Spring Break			
Week 12	4/19–4/23	Genome Rearrangements	HW7: Dealer's Choice
Week 13	4/26–4/30	Sequencing Antibiotics	
Finals Week			Final Assignment

* We will attend two #BlackInCompBio seminars during class.

** Wed lecture and all labs are canceled (mid-term break).

3 Assessment

There are three general types of student assessment, which allows you to demonstrate acquired knowledge in multiple ways.

50% **Programming Assignments:** There will be approximately seven programming assignments of varying lengths. Beginning with an assignment using spreadsheets, the assignments are designed to gradually make you comfortable with programming in Python.

- You may work with others on the programming assignments, within the limits of the Collaboration Policy.

- There will be at least one **checkpoint** midway through each assignment. If you submit your code with a question by the evening of a checkpoint, I will give quick feedback the

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 significantly *less* time than expected, then attempt the challenge questions. If it takes you significantly *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 specified 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 first 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 field 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 field, 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 to give you a view of cutting-edge research in computational biology.
- Many students will have personal circumstances that may affect their performance in the course. I will work with you to make adjustments to the course schedule as needed, and students should be encouraged to seek guidance on any anticipated or realized issues.
- Mental and physical health is more important than attending all parts of class. If you miss multiple days of class, I will reach out to make sure you are okay.

Bio131 is a better course when students have a diversity of lived experience and previous knowledge.

4.1   Feedback for the Instructor

Please give me feedback on anything in the syllabus or course, especially with respect to making Bio331 an inclusive learning environment. Let me know if anything makes you uncomfortable in class, if you would like more instruction on a topic, or if you are experiencing a hardship outside of class. Anonymous feedback is available on Moodle, though with a small class your comments may be identifiable.

Keys for Success

Start the assignments early. The time you spend on each assignment might vary drastically, depending on how much you debug your programs.

- Have a plan for the work you need to do for each component of the assignment.
- Work through problems on paper if you are confused with a concept.
- It is better to have something working that is not quite complete than something that does not work but "has all the parts."

This course is cumulative – topics build upon each other throughout the semester, and attendance is critical. Let Anna know if you miss a lecture or a lab to get caught up. While there is no penalty for missing lecture, lab attendance is expected unless explicitly stated or accommodations have been arranged.

When in doubt, ask. Asking for help is sometimes unintuitive, but a little clarification goes a long way. If you have no idea how to start a problem (which happens!), stop by office hours or the DoJo. Also take advantage of the one-on-one tutoring available.

Working collaboratively is fun! If possible, try to work with others on the assignments within the limits outlined in the Collaboration Policy.

Start the assignments early. Really.

Even though this spring will have its academic and personal challenges, I am looking forward to working with you all this semester.