Programming for Bioinformatics - BIOL7200

3 Credits
Tues, Thurs 6:30-7:45 pm

Course Summary

Bioinformatics, generally, is the use of computers to analyze biological data or otherwise address biological questions. The biological questions that can be addressed using bioinformatics are diverse. However, there are core bioinformatics skills which are broadly useful: the ability to work in a Linux environment and use scripting languages like Python and R. This intensive course covers some of the core programming concepts used in bioinformatics and provides opportunities to apply the concepts learned to bioinformatic software development through practical exercises.

Pre-Requisites

- This course assumes some prior knowledge of biology. Descriptions are provided for ay biological concepts required for assignments.
- This course does not require any prior programming experience. However, while the material begins with the basics, students with no programming experience should expect to commit significant time and effort to the course (see Course effort level section below).
- This course requires access to a Linux terminal. This means you must have access to a Mac, Windows, or Linux computer on which you have administrator privileges and the ability to install software.

Course Goals and Learning Objectives

Upon successful completion of this course, you should be able to:

- Navigate a filesystem and manipulate files in a Unix command line environment using a terminal and Bash scripting
- Use Git to create and version control software development projects
- Use Conda/Mamba to create and manage environments and install bioinformatics software
- Write Python scripts and execute them in a command line environment
- Write your own Python packages
- Design and develop simple bioinformatic software using Bash and Python
- Publish and distribute your software using Conda/Mamba

Course Assessments and Grading

Students will be assessed using take-home coding assignments which will be assigned most weeks. All assignments carry equal weight, although some include additional extra credit problems. Assignments will be set on Tuesdays and are due the following Monday by 11:59PM and must be submitted to Canvas before that time to avoid a late submission penalty. Any deviations from this assignment schedule will be communicated via the course canvas page and announced in class. Solutions to assignments will be demonstrated and discussed during Thursday sessions.

Final grades will be assigned according to the percentage of points awarded across all assignments. The score percentage corresponding to each grade is as follows (scores falling exactly on a breakpoint will round up - e.g., 90% is an A):

Grade	Percent Score
A	90-100
B	80-90
C	70-80
D	60-70
F	<60

Additional materials/Resources

This course does not use a textbook or other required texts. Optional readings are distributed through the class canvas page throughout the course. In addition, the following textbooks may be helpful to students who wish to read more about (or beyond) the topics covered in this course.

- The Linux Command Line by William Shotts. Free PDF from author available at link
- Practical Computing for Biologists by Steven Haddock and Casey Dunn. Published 2010, so uses Python2, but Bash sections and much of Python material is relevant. Includes freely distributed "cheat sheets"

Learning to code is like learning to speak a new language. Practice using the languages is an essential part of learning. The assignments set during this course are designed to reinforce the topics covered in class. In addition, the following resources provide problems which can be solved to practice coding in Bash and Python.

- Advent of code is an annual Christmas-themed advent calendar of coding problems/puzzles. This link takes you to an index of the years for which puzzles were released. Click on a year to go to the puzzles from that year.
- Rosalind.info is a repository of bioinformatics-related coding challenges with difficulties ranging from very easy to more difficult.

Academic Integrity

Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on Georgia Tech's Academic Honor Code, please visit here or here.

Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404)894-2563 or here, as soon as possible, to make an appointment to discuss your special needs

and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs.

Student-Faculty Expectations Agreement

At Georgia Tech we believe that it is important to strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. See here for an articulation of some basic expectations that you can have of me and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech while in this class.

Attendance

Attendance is strongly encouraged, but is not tracked and is not required. Classes will be streamed over zoom using a link provided on the course Canvas page and recordings made available through canvas after the class. Importantly, the Zoom chat is not monitored so questions asked by remote attendees will not be addressed. Additionally, the auditorium is not set up to record students so remote attendance and recordings of discussion sessions will be far less valuable than attending in person.

Collaboration and group work

Collaboration and discussion of assignments is acceptable as long as each student submits their own work (see the section on academic integrity and plagiarism). The primary goal of this course is for enrolled students to learn to use Bash and Python to write bioinformatics software and perform bioinformatic analyses. Any collaborative efforts that improve your understanding of these things is encouraged. Collaborative efforts that undermine that goal (e.g., copying without understanding) are not permitted.

Chat-GPT and other AI systems

The use of AI systems such as Chat-GPT and GitHub Copilot during this course is discouraged. Learning to read documentation and search forums like stackoverflow are valuable skills to develop when taking intro programming classes like this one and the use of AI may undermine progress towards those learning objectives.

If you do choose to use AI, you should be mindful to do so in a way that supports your learning and does not violate the Georgia Tech policies on plagiarism. The policy for what constitutes acceptable use of AI is similar to that for working with your peers. Discussion of course material or how to approach a problem with classmates is acceptable and so it is also acceptable to discuss these things with an AI chatbot. However, working together through an assignment or copying and pasting code is not permissible either with a human peer or with AI. You should never include anything in an assignment that was not written by you.

Two rules you can follow to ensure that your use of AI does not constitute academic misconduct:

1. Never copy anything from your conversations with AI assistants. You can copy code into the prompts for the AI assistant, but never copy the responses of the AI assistant.

2. Do not have an AI assistant and your assignment submission open at the same time. You can use a conversation with AI to clarify your understanding of a topic, but by closing the conversation before

working on your assignment, you ensure that the assignment reinforces and reflects your understanding.

Extensions and late submissions

Extensions will be granted at the discretion of the course instructor on a case by case basis. To increase the chance that you are granted an extension, please contact me as far in advance as possible.

Late submissions will result in a 20% penalty if submitted within 24 hours of the original deadline. Beyond 24 hours, a 50% penalty will be applied. Submissions received after the assignment has been discussed in class (i.e., after 6:30 PM on Thursday, when class starts) will not be accepted and will result in a score of 0 for that assignment. Late penalties may be waived on a case by case basis if mitigating circumstances exist. However, it is much easier to get an extension than to get a late penalty waived so please contact me in advance if you are unsure if you will be able to meet a deadline.

Food and drink in the classroom

Classes will take place around dinner time. However, please plan to eat before you arrive at class or wait until afterwards. Eating in class (both the act and any associated sounds/smells) is disruptive to other students.

Course effort level

This course is designed to provide students with the basic coding skills they need to succeed in a career in bioinformatics. It is intended to be accessible to individuals with no prior experience working in a terminal (using a command line). However, this course covers a lot of material. If you have limited coding experience or have never used a terminal before, then you may need to spend 20 hours or more a week on this course.

Both I and the course TAs are available to support your learning. However, learning to code unavoidably involves a large amount of self-directed learning and practice.

Posting code publicly

Through assignments over the course of the semester, we will write a functional Python package that can be used to perform bioinformatic analyses. In the past, students have expressed a desire to publicly host the code they have written on their GitHub account in order to showcase the skills they have developed. The public hosting of code written over the course of the semester is absolutely fine as long as you do not publicly post assignment solutions until after the class in which we go over those solutions.