PROGRAMMING FOR BIOINFORMATICS – BIOL 7200 – Syllabus – Fall 2023

Instructor: **Alan Collins** (biol7200.gt@gmail.com)
Teaching Assistants: **Haojun Song** (hsong343@gatech.edu)
Daniel Groves (Dgroves6@gatech.edu)

Course summary

The fields of Bioinformatics and Computational Biology occupy the intersection of the life sciences and information technology. As sequencing technologies have improved and become cheaper, the amount of data generated by Biologists working in academic research, biotech, and public health has grown enormously. Computers are needed to handle and analyze this massive amount of information. In addition, highly specialized knowledge is required to perform manipulation and analysis of the data, leading to a growing need for bioinformatic expertise throughout the Biological sciences. A fundamental element of bioinformatic expertise is the ability to use programming languages to work in a Unix environment and perform analyses.

This project-based/lab course will introduce programming for bioinformatics. We will begin by introducing you to the command line environment in the Unix / Linux operating system – where bioinformatics is done. This will include a broad overview of Unix / Linux utilities as well as shell scripting. Next, we will introduce Python, a programming language used commonly by bioinformaticians. We will cover the basics of writing simple scripts and then progress to more complex analyses of real biological data. The goal of this course is to provide a foundation in programming which can then be applied to any of the many areas of bioinformatics.

All required and recommended readings, lectures and exercises will be made available on the course Canvas site (https://canvas.gatech.edu/). Students will be required to post exercise answers and codes to the course Canvas site for evaluation. Assignments are due by 11:59 PM on the Monday after the exercise is assigned and must be posted to Canvas by that time. Late assignments will not be accepted and will result in a score of 0. Thursday follow-up programmatic discussion sessions will serve as code demonstration and lab sessions where students will get the opportunity to discuss class problems with the TAs. Office hours can be requested by email.

Required equipment

In order to participate in this class, you will need access to a Unix terminal (Linux or Mac). If you are using a Mac computer, you can use the built-in terminal application. If you are using a Windows computer, you will need to install software to access a Linux environment. See the following links for guides to set up a Linux environment on Windows:

- Installing Windows Subsystem for Linux (easiest, recommended):
  https://learn.microsoft.com/en-us/windows/wsl/install
- Installing a virtual machine (requires a computer with >= 8 CPU threads. 16GB RAM, 50GB free hard drive space):
  https://ubuntu.com/tutorials/how-to-run-ubuntu-desktop-on-a-virtual-machine-using-virtualbox
- Running Ubuntu in a docker container (advanced):
  https://ubuntu.com/tutorials/windows-ubuntu-hyperv-containers

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 http://www.catalog.gatech.edu/policies/honor-code/ or http://www.catalog.gatech.edu/rules/18/.

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 http://disabilityservices.gatech.edu/, 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 http://www.catalog.gatech.edu/rules/22/ for an articulation of some basic expectation 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.

Course Evaluation:

Demos	25%
Code/exercise submission & evaluation	75%
• Code/assignment submission	*10%*
• Code evaluation	*65%*

Schedule of lecture / lab sessions

Classes are 6:30-7:45 Tuesdays and Thursdays

Date	Topic	Room
8/22/2022	Introduction to Unix and Bash	Instructional Center 215
8/24/2021	Glob and working with text files	Instructional Center 215
8/29/2022	Environments and installations	Instructional Center 215
8/31/2021	Permissions, writing and using Bash scripts	Instructional Center 215
9/5/2022	Symlinks, Regex	Instructional Center 215
9/7/2021	Version control with git	Instructional Center 215
9/12/2022	Loops and conditionals	Instructional Center 215
9/14/2021	Parameter expansion and process substitution	Instructional Center 215
9/19/2022	Introduction to version control and Bash scripting	Instructional Center 215
9/21/2021	Flex session	Instructional Center 215
9/26/2022	Introduction to Python	Instructional Center 215
9/28/2021	Reading and writing files	Instructional Center 215
10/3/2022	Loops and control flow	Instructional Center 215
10/5/2021	Functions and docstrings	Instructional Center 215
10/10/2022	*Fall Break*	
10/12/2021	Commandline arguments and modules	Instructional Center 215
10/17/2022	Plotting data	Instructional Center 215
10/19/2021	Statistics in python	Instructional Center 215
10/24/2022	Subprocess	Instructional Center 215
10/26/2021	Try-catch and exceptions	Instructional Center 215
10/31/2022	Classes and custom modules	Instructional Center 215
11/2/2021	Flex session	Instructional Center 215
11/7/2022	Decorators	Instructional Center 215
11/9/2021	Writing a fasta reader	Instructional Center 215
11/14/2022	Implementing algorithms	Instructional Center 215
11/16/2021	Implementing algorithms	Instructional Center 215
11/21/2022	Flex session	Instructional Center 215
11/23/2022	*Thanksgiving break*	
11/28/2021	Review of concepts covered in the class	Instructional Center 215
11/30/2021	Review of concepts covered in the class	Instructional Center 215
12/5/2022	Final instructional Day	Instructional Center 215

Note that the syllabus is subject to change depending on the speed at which the class progresses and the needs of the students.