

## PROGRAMMING FOR BIOINFORMATICS – BIOL 7200 – Fall 2018

Instructors: *Lavanya Rishishwar*

*School of Biology, Engineered Biosystems Building, 2200*

[lavanya.rishishwar@gatech.edu](mailto:lavanya.rishishwar@gatech.edu)

**Course summary:** The fields of Bioinformatics and Computational Biology occupy the intersection of the life sciences and information technology. Over the last decade, there has been an explosion of data in the life sciences and the proliferation of raw information promises to continue at an even more rapid pace. Computers are needed to handle and assimilate this massive amount of information. More importantly, the role of bioinformatics is to convert information, in the form of data, into biological knowledge. In order to do this, bioinformaticists and/or computational biologists must be adept at the use of computers, *i.e.* **YOU MUST KNOW HOW TO CODE.**

This project-based/lab course will provide an introduction to programming for bioinformatics. We will begin by introducing you to the command line environment in the Unix / Linux operating system – this is where real scientific computing gets done. This will include a fairly broad coverage of Unix / Linux utilities as well as shell scripting. The course will then go on to use the Python programming language to illustrate the fundamentals of bioinformatics programming.

All required and recommended readings, lectures and exercises will be made available on the course T-square site. This is an exclusively practical and active learning class. Students will complete exercises in order to learn how to code and how to do bioinformatics. The only way to learn the course material is by doing. Accordingly, attendance and participation are mandatory and critical. Students who show up late or miss lecture sessions will be penalized. Participation in lecture sessions will be judged by the degree to which each student participates in class discussions and exercise sessions. Students will also have the opportunity to demonstrate and explain their code to the class. Students will be required to post their code and exercise answers to the course T-square site for evaluation. Assignments are due by 12:00 am on Mondays and must be posted to T-square by that time. Late assignments will not be accepted and will result in a score of 0.

Please see [www.honor.gatech.edu](http://www.honor.gatech.edu) for Georgia Tech's Academic Honor Code, which you are required to uphold.

### Course Evaluation:

Class participation (attendance)	20 %
Code / exercise evaluation	80 %

### Schedule of lecture / lab sessions

Date	Topic	Room
TBD	Introduction to *nix environment and permissions	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Installation, streams, and pipelines	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Regex, file handling/text processing (sed, grep, awk)	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Introduction to Programming	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Data structures and file handling	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Subroutines, references, and system calls	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Syntactical sugars and basic modules	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Common Bioinformatics Formats	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Bioinformatics Pipelines	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Introduction to Bioinformatics Tool Development	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Dynamic Programming	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Threading and parallelization	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	Bioinformatics Simulations	TBD
TBD	Follow-up Programmatic Discussion	TBD
TBD	High-Throughput Analysis & Review	TBD
TBD	Follow-up Programmatic Discussion	TBD

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