Genomics and Applied Bioinformatics Biol 4150/6150, Fall 2021

Instructor:	Jung H. Choi	jung.choi@biology.gatech.edu; Twitter @jung_gt
TA:	Nolan English	nenglish@gatech.edu

Required readings: Instead of a hard copy textbook, readings will be from the Biology Computes web pages and other on-line texts or articles posted on Canvas.

Recommended equipment: If you have not already done so, I recommend that your laptop be configured as a dual-boot linux/Windows configuration, or be able to run the 2nd OS as a virtual machine (e.g. using VirtualBox). There are many "flavors" or "distros" of linux to choose from; most are available free. If you have a Mac, you can install and run linux applications in OS X with some fiddling.

Learning objectives: You should be able to:

1. Locate and access a variety of public databases and software tools for bioinformatics.
2. Use/apply the common tools appropriately, with knowledge of their limits and pitfalls.
3. Evaluate the statistical significance of results of bioinformatics searches and analyses, where appropriate.
4. Evaluate the biological significance of, or formulate new questions based on the results of, bioinformatics searches and analyses.

Lectures: MW 2:00-2:50 MRDC 2407 or BlueJeans
For Fall 2021, students may opt to attend class in-person, or via BlueJeans. To accommodate students who may miss class for various reasons, BlueJeans recordings will be posted on Canvas.

Labs: F 12:30-3:15 Instructional Center 215
Attendance at labs is optional. Students should attempt the computer exercises before Friday, and come to the lab for any assistance required. The TA will also be available for on-line help – please contact Nolan by email to arrange an appointment.

Grading:

10%	Web posts on group project assignments
30%	Two midterm exams (take-home)
40%	Computational assignments
10%	Group project video presentations
10%	Participation in group work

Projects for graduate students will be more complex, involve writing scripts, and require a higher level of analysis.

Policy on collaboration and the Honor Code: Discussion with peers is encouraged, both during class group discussion & work periods and during computer exercises. However, all work turned in must be of your own original efforts**.** Using the work of your peers (copying and pasting, **even with alterations)** or from any other sources**, on-line or in print, without attribution** is **plagiarism. This includes photos and figures from websites. Plagiarism will result in zero credit for the assignment, and possible further penalties as an Honor Code violation.** All students are expected to abide by Georgia Tech's Honor Code.

Learning Accommodations: As needed, we will happily accommodate students with disabilities. These accommodations must be arranged in advance and in accordance with the Office of Disability Services (disabilityservices.gatech.edu).

Disclaimer: The course schedule and grading policies are subject to revision, depending on how the semester proceeds. Any changes will be announced in class and posted on Canvas.

Tentative Schedule of Topics: The table below is the first iteration of the schedule of topics. As the course progresses, this schedule may be adjusted to fit the pace of learning.

Week	***Topic***	***Computer Lab/Homework due Monday before class***
1:	Intro DNA sequencing, file formats, assembly and annotation overview Public databases and genome browsers	Genbank tutorial
2.	Genome assembly and annotation – intro to HMMs	No lab
3:	Pairwise sequence alignment – local and global alignments	Needleman-Wunch, Smith-Waterman
4:	BLAST & Advanced BLAST	BLAST, PSI-BLAST
5:	Multiple sequence alignment & molecular phylogeny	Take-home exam 1 includes phylogenetic trees
6:	High-throughput approaches to gene expression	Microarray data normalization
7:	Expression data analysis – normalization, statistical analysis	Gene expression data analysis project
8:	RNA seq	
9:	Gene expression data analysis – clustering, pathways	
10:	Gene expression data analysis	Take-home exam 2
11:	Gene expression group project presentations	group presentations
12:	Protein structures - classification and prediction	Homology modeling project
13:	Protein structures – homology modeling	
14:	Human exome/genome sequencing and annotation	Final group project on human exome variant calling and analysis
15:	Variant calling & analysis	
16:	What's trending in genomics	