COMPUTATIONAL GENOMICS – BIOL 7210 A – Spring 2020

Professor: ***I. King Jordan*** *and TA:* ***Shashwat Deepali Nagar***
School of Biology, Engineered Biosciences Building (EBB), Office: 2109 Lab: 2200
king.jordan@biology.gatech.edu   *shashwat@gatech.edu*

Course summary: The science of genomics involves the intersection of experimentation and computation. Computers are quite obviously required to handle the massive amount of data produced by genome sequencing projects. More importantly however, genome sequencing efforts yield ‘information’ alone, which can only be converted into ‘knowledge’ through the use of computers. In this class, the students will convert raw genomic information (*i.e.* sequence reads) into knowledge through the use of computational genomics tools and applications. The class will be provided with unassembled genome sequence data from the Centers for Disease Control and Prevention (CDC) and will proceed through five distinct stages of analysis and interpretation of that data: 1-**genome assembly**, 2-**gene prediction**, 3-**functional annotation**, 4-**comparative genomics** and 5-production of a **predictive webserver**. This course will be entirely practical in nature. Students will learn to do the actual work of computational genomics. Expert guest lecturers will be brought in to provide information on state-of-the-art computational genomics tools. Based on this information, other class lectures and their own research, students will be solely responsible for choosing which tools (*e.g.* programs and/or databases) to use, how to implement them and for producing and thoroughly documenting their final results. All results will be integrated into a publicly available predictive webserver.

This class meets on Tuesdays and Thursdays from 8:00am to 9:15am in Klaus 2456. There is no textbook. Required and recommended readings will be made available on the course Wiki page - http://compgenomics2020.biosci.gatech.edu/ - along with any lecture material. Students are required to use online databases and the scientific literature to inform their choice of computational tools to be used. Since there is no textbook and many of the sessions involve class discussion and lab activities rather than formal lecture, attendance and class participation are absolutely mandatory.

Evaluation:

Class participation	**10%**
Exercise sessions (4 x 5%)	**20%**
Group presentations (4 x 5%)	**20%**
Final Results (2 x 20%)	**40%**
Documentation (2 x 5%)	**10%**

Class attendance and participation are mandatory. **Class participation** will be judged by the degree to which each student participates in class lectures and discussions (by asking questions, answering questions, offering ideas and opinions), during group presentations (by asking questions during others’ presentations, by engaging the audience during their own presentation, by connecting their presentation to previous class discussions, by working successfully in a small group), and during computer laboratory activities (by performing analyses and working with other students). Students who show up late or miss class will lose 10% of their class participation grade each time.

Each group will give a series presentations and laboratories/demos. **Group presentations and labs/demos** will be judged by the depth of analysis presented, the clarity of presentation, the utility of the exercises, the appropriateness and justification of the choices made, the validity and robustness of the results and the thoroughness of the documentation. In addition to presentations, results and documentation should be presented on the class Wiki site. All student code and analysis contributions must be shared and documented on Github – https://github.gatech.edu/compgenomics2020/. Specific requirements for the presentations will be provided during class sessions. Contributions of each individual student to the overall group effort must be meticulously detailed and documented.

Please see *www.honor.gatech.edu* for Georgia Tech’s Academic Honor Code, which you are required to uphold.

Date	Day	Topic	Presenter(s)
1/7/2020	Tuesday	Introduction & Logistics	King Jordan
1/9/2020	Thursday	Class groups, Wiki & GitHub	Shashwat Deepali Nagar
1/14/2020	Tuesday	Source Code Management with GitHub	Shashwat Deepali Nagar
1/16/2020	Thursday	Genome Assembly Concept	King Jordan
1/21/2020	Tuesday	Genome Assembly Exercise	Shashwat Deepali Nagar
1/23/2020	Thursday	CDC Enteric Disease & PulseNet	Heather Carleton-Romer, CDC
1/28/2020	Tuesday	**Genome Assembly Background & Strategy**	**Students - Genome Assembly Groups**
1/30/2020	Thursday	Gene Prediction Concept	King Jordan
2/4/2020	Tuesday	Gene Prediction Exercise	Shashwat Deepali Nagar
2/6/2020	Thursday	Bioinformatics at the CDC	Scott Sammons, CDC
2/11/2020	Tuesday	Modern Approaches to Genomic Epidemiology	Lee Katz, CDC
2/13/2020	Thursday	**Gene Prediction Background & Strategy**	**Students - Gene Prediction Groups**
2/18/2020	Tuesday	**Genome Assembly Final Results**	**Students - Genome Assembly Groups**
2/20/2020	Thursday	Functional Annotation Concept	King Jordan
2/25/2020	Tuesday	Functional Annotation at NCBI	Leonardo Mariño-Ramírez, NCBI
2/27/2020	Thursday	Functional Annotation Exercise	Shashwat Deepali Nagar
3/3/2020	Tuesday	**Functional Annotation Background & Strategy**	**Students - Functional Annotation Groups**
3/5/2020	Thursday	**Gene Prediction Final Results**	**Students - Gene Prediction Groups**
3/10/2020	Tuesday	Investigating foodborne outbreaks with genomic epidemiology	Sung Im, CDC
3/12/2020	Thursday	Bioinformatics webservers and visualization	Andrew Conley, IHRC-ABiL
3/17/2020	Tuesday	**Spring Break**	
3/19/2020	Thursday		
3/24/2020	Tuesday	**Functional Annotation Final Results**	**Students - Functional Annotation Groups**
3/26/2020	Thursday	**Comparative Genomics Background & Strategy**	**Students - Comparative Genomics Groups**
3/31/2020	Tuesday	**Webserver Background & Strategy**	**Students - Webserver Groups**
4/2/2020	Thursday	Open lab	Shashwat Deepali Nagar
4/7/2020	Tuesday	Comparative Genomics Concept	King Jordan
4/9/2020	Thursday	Comparative Genomics Exercise	Shashwat Deepali Nagar
4/14/2020	Tuesday	**Comparative Genomics Final Results**	**Students - Comparative Genomics Groups**
4/16/2020	Thursday	**Webserver Final Results**	**Students - Webserver Groups**