

Computational Genetics of Big Data

Online classes

The class will be entirely online this semester. We will use Zoom for online instructions. All meetings will also be recorded and made available on a YouTube channel.

Website

A GitHub repo will serve as the course syllabus. Course materials and computing resources for this course will be available at Amarel. Grades will be posted on Canvas.

Communication

The primary mode of communication in this course (including major announcements) will be through Canvas.

Emails

Although the bulk of the communication will take place via Canvas, at times (rarely), we will send out important course information via email. This email is sent to your Rutgers email address. You are responsible for all information sent out to your University email account, and for checking this account on a regular basis.

Virtual office hours

Virtual office hours will be held via whereby.com or Zoom by the two teaching assistant on Wed and Fri. The instructor will hold office hours by prior appointment.

Course Activities

In-class exercises

Each class will contain 2-4 modules. Each module will involve a ~15 min lecture followed by a programming exercise lasting ~20-30 min. Programming exercises will typically involve 2-5 questions. You will be asked to solve these questions on your own, which will involve applying the functions taught in the lecture part of the module. At the

end of the individual exercise portion, I will go over the exercise as a group. Any piece of code written by me will be annotated and commented before being uploaded to the course website.

DataCamp for Students

All students taking this course will be assigned DataCamp]modules every week. The modules will be parallel to the materials covered in the class. All students will be required to complete DataCamp modules by the assigned due date.

Each module is worth 10 points.

Homeworks

In every lab, you will receive a homework assignment, typically a series of programming exercises that you must code based on the material you learn in class that day. You will upload solutions to these homework as a single R Notebook file before the beginning of the next lab. **NO LATE SUBMISSIONS WILL BE ALLOWED.** If you miss your due date for the homework, you will receive a 0 points for that assignment.

Each homework is worth 10 points.

Quizzes

Quizzes will be conducted once every 4 classes. They will last 1-hr and be similar to in-class exercises. Quizzes will typically be "open-book", i.e., you will be able to refer to your class notes, presentations, and my codes.

Each Quiz is worth 10 points.

Final projects

Towards the end of the course, all students will be split into teams of 2-3 individuals. Each team will complete a finals research project. Each team can select one of the two broad research topics listed below:

1. *Comparative genomic analyses:*

This project will involve downloading genome sequences and annotations of 2 or more organisms and performing various sequence

level analyses.

2. *RNA-seq analyses:*

This project will involve downloading raw sequencing datasets from GEO database and identification of differentially expressed genes and their functional characterization.

Detailed instructions for the finals projects will be given 4 weeks prior to the last class.