

Bioinformatics Applications (PLPTH813)

Semester: Spring 2021

Instructor: Sanzhen Liu, Plant Pathology, Kansas State University

Contact information: liu3zhen@ksu.edu

ZOOM: <https://ksu.zoom.us/j/95782921551>

Time: Tuesday, Thursday 10:30am-11:20pm (lectures); Thursday 12:30-2:00pm (lab)

Office: <https://ksu.zoom.us/j/8468443307> (appointment is required)

Office hours: Tuesday 12:30-1:30pm

Exam: midterm from 10am on 3/11/2021 Thursday; final in the final week

Goal of the course

PLPTH813 will cover the basic principle of regular bioinformatics applications and emphasize the practice of bioinformatics in biological research. The course will introduce some useful computation skills, such as Unix commands and R programming, in data analysis at the beginning of the semester, and then will apply these skills throughout the course. This course will be focused on the application of bioinformatics in next-generation sequencing (NGS) data analysis, including genetic mapping, genome assembly, variant discovery, genotyping, RNA-Seq, and so on. The ultimate goal of this course is to help students to be prepared for the next-generation biological research that often generates large data and requires computational skills in data management and data translation.

Tentative course content

(The course content will be adjusted according to the advance of computational tools and the interests of course audience.)

PART 1. Useful Unix tools and R programming

1. Unix/Linux system and useful tools
2. R introduction and useful R packages

PART 2. Basic tools for DNA sequence analyses

1. Introduction of NGS technologies
2. Standard data formats
3. Tools for NGS data process and visualization

PART 3. Application of NGS in Genomics

1. Genomic variant discovery
2. Phylogeny
3. Genetic mapping (QTL, bulked segregant analysis, genome-wide association)

4. Genome assembly (assembly tools and evaluation)
5. Comparative genomics
6. Metagenomics
7. RNA-Seq (technology, experimental design and differential expression)
8. Applications of machine learning in genomics

PART 4. Student project presentations

Grading

Participation 10%, Homework 30%, Midterm Exam 20%, Student project 15%, final Exam 25%

1. Participation

Students are anticipated to be present in each Zoom session. Absence needs to be informed prior to the class. Without notifications, the score for the participation will be compromised.

2. Midterm and final exam

The midterm exam will cover Part 1 and 2. The final exam will cover all teaching materials through the semester. Students are allowed to use books and laptops. Two hours will be given to students to return their answers. Communication is NOT allowed among students during the exam course.

3. Student project

Students are expected to start to design their projects after February. The project can be related to students' own research projects or utilize public data for a meaningful analysis. Ideas will be proposed in a 5-minute presentation during the class. The presentation should include the goal, the rationale, the data source, and the expected result. In the last 1-2 weeks, students will present their projects. Each presentation will take ~15 minutes.

Other notes

We will use the software R throughout the course. Students in the class are required to have the Beocat account that allows you to access the Beocat server and use R-studio through Beocat.

Acknowledgements

The development and improvement of this course was contributed by Hairong Wei, Ying Hu, Lei Li as well as students taking this course in the past.