BIMM170: Genomics Research Initiative Lab II
Spring Quarter 2020
Instructor: Rachel Dutton, PhD rjdutton@ucsd.edu
IA: Cong Dinh cbdinh@ucsd.edu

This course is part of the Howard Hughes Medical Institute (HHMI) SEA-PHAGES program. SEA-PHAGES (Science Education Alliance-Phage Hunters Advancing Genomics and Evolutionary Science, https://seaphages.org/) is a discovery-based undergraduate research course that begins with simple digging in the soil to find new viruses, but progresses through a variety of microbiology techniques and eventually to complex genome annotation and bioinformatic analyses.

When last seen, your phage DNA was in a microfuge tube. Now what?

A significant portion of the course this quarter is to analyze that string of letters (the sequence of your phage genome) and use a combination of comparative tools and bioinformatic algorithms to determine, to the best of your ability, the location and function of genes and features in the genome, and how the genome compares to others that have been discovered previously. Along the way, you will learn about the field of Bioinformatics, and the power and limitations of performing experiments within the computer environment.

What we hope you will learn during this course:
1. Understanding viral genome diversity and evolution
2. Understanding the major components of a virus
3. How to find genes in a genome and annotate their function
4. Understanding The Central Dogma: DNA to Proteins
5. How to process sequencing data (bioinformatics)
   a. How to use high-performance computing (Unix/command line/TSCC/AWS)
   b. Sequence analysis software (assembly, variant calling, gene calling, gene annotation, mapping)
6. Reading and interpreting scientific papers and data
7. Visualizing and presenting data

How this class is going to work:
1. Each week, we will introduce a new topic related to viral biology, and a set of tools we will be using to answer related questions about a viral genome
   a. **Lectures will be held over Zoom on Tuesdays at 2pm**
   b. Lecture zoom sessions **will be recorded** and uploaded to Canvas for viewing later if the course meeting time is not convenient for anyone
   c. Participation in Zoom lecture sessions is encouraged but <u>NOT required</u>
2. For each assignment, you will work in groups (ideally with the same group of people as Fall quarter, but if anyone needs to change groups please let me know ASAP) to answer the questions and accomplish the goals for the week. Each group will be examining a different viral genome for the duration of the course. This includes: phage isolated in the Fall quarter, Pseudomonas phage that may be used in phage therapy, or a genome of SARS-CoV-2, the virus causing the current COVID-19 pandemic
   a. Each group will work together on assignments, but each person will answer assignment questions in their own words, and maintain their own lab notebook documenting their

work.

    b. For each week, each group will build a short presentation on their progress in Google slides (a few slides describing the goals for the week and the results)

    c. Both of these will require some coordination of meeting times for group members or some other way to work together remotely. You can use Zoom or any other technology that works best for you.

3. Each week we will hold office hours for questions on biology and for bioinformatics help on **Thursday from 2-4pm.**

4. Each week you will be asked to fill out a form that will ask you to "self-report" your contributions for the week. You will also be asked to fill out a "Muddiest Point" question, in which you tell us what was most confusing about the class for the week. These will both be on Canvas and will be used for your "participation" points.

How you will be graded in this course:

1. **Participation:** 20% of grade
   a. Based on self-reporting of contributions for each week and turning in a weekly "muddiest point", both on Canvas.
   b. NOT based on attendance in Zoom meetings
2. **Weekly Assignments and Lab Notebook Checks**: 70% of grade
3. **Final Lab Report:**10% of grade

Other important information:

1. Class sessions will be recorded and made available to students asynchronously
2. This syllabus is subject to change, particularly because of campus efforts to contain covid-19. Any schedule changes will be posted on the course website.
3. Please make sure that you check out this website for resources on how to learn remotely:
   a. https://digitallearning.ucsd.edu/learners/learning-remote.html
4. Students are expected to do their own work, as outlined in the UCSD Policy on Academic Integrity. **Academic misconduct** is broadly defined as any prohibited and dishonest means to receive course credit, a higher grade, or avoid a lower grade. Academic misconduct misrepresents your knowledge and abilities, which undermines the instructor's ability to determine how well you're doing in the course. Please do not risk your future by cheating.
5. I understand this is a challenging time and that you may have challenges with accessing the course material, adapting to online-only learning, and taking online quizzes and exams. My goals are to teach you the course material, fairly assess your knowledge of this material, and grade you accordingly, while keeping these challenges in mind.

Course schedule (Subject to change):

Week	Activities
1	Lecture #1: Review of Fall quarter, Plan for Spring Quarter
	Assignment #1: Getting set up! - Downloading and using zoom, and other basic tools (Terminal, FileZilla) - Review your notes from Fall quarter - Coordinating meeting times with your groups

2	Lecture #2 : Viral biology. How do we measure and classify diversity? Bioinformatics: High Performance/Cluster computing

	Assignment #2:
	    - Reading paper on phage diversity (Arthrobacter paper) and NYT virosphere
	    - Intro to TSCC and conda environments

3	Lecture #3: Next-gen sequencing/Illumina sequencing and their applications. Bioinformatics: what is a fastq/fasta file? Sequence quality metrics
	Assignment #3: Sequencing read analysis
4	Lecture #4: Viral genomes/ Bioinformatics: How do you go from reads to a genome?

	Assignment #4: Genome Assembly and Analysis
	    - Assembly with SPAdes
	    - Contig analysis: GC, dinucleotide, trinucleotide, start sites.

5	Lecture #5: Comparative genomics Bioinformatics: NCBI/BLAST intro
	Assignment #5: BLAST the longest contigs of the assembly! What is in there?
6	Lecture #4: Transcription, translation, genes, operons, tRNAs Bioinformatics: Annotating phage parts, in-class examples
	Assignment #6: Genome annotation Check tRNA predictions, Examine gaps, Evaluating Protein Coding Gene Predictions
7	Lecture 7: Annotation continued
	Assignment #7: Finish annotations/ Intro to R
8	Lecture #8: Viral evolution
	Assignment #8: Mapping Reads
9	Lecture #9: SARS-CoV-2
	Assignment #9: Rsamtools and viewing mutations.
10	TBD
	Work on Final Reports