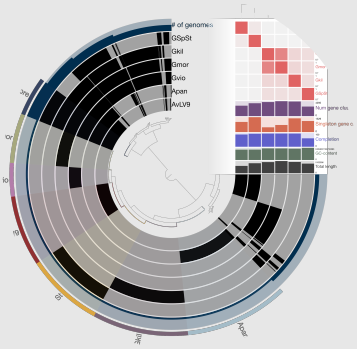

Microbial Genomics Lab

BISC 4234/6234
(3 credits)

Spring 2023

Course Info

- Prerequisites: BISC2336 and basic understanding of Unix/Linux commands OR approval from the instructor.
- Wednesdays
- 12:45-15:15
- Duques 250

Instructor Info

- Jimmy Saw
- Office hours: Thursdays 1-3pm
- Lisner 413
- jsaw@gwu.edu

Course Description

What do you see when you look at the figure below? An eye staring at you? Some sort of sand mandala?

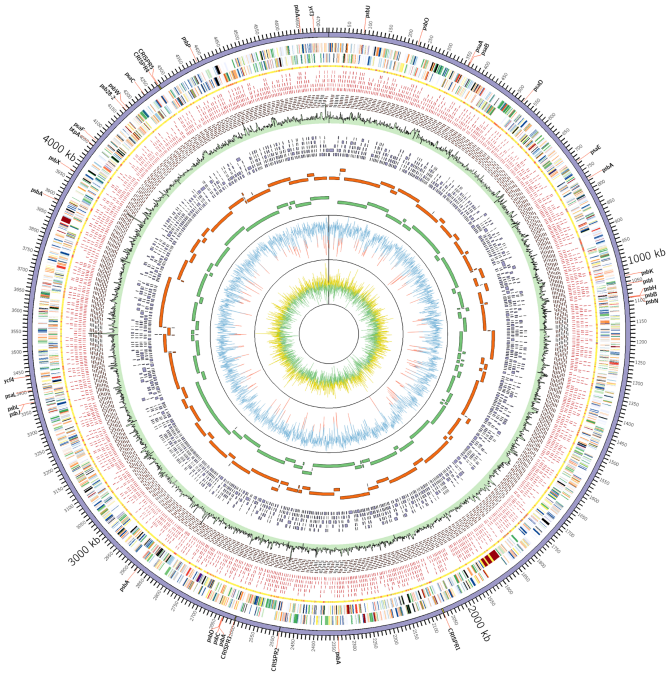

Figure 1: What is this?

This is a circular representation of a bacterial genome! It is just one way of visualizing genomic information. Did you know that most of microbial genomes are circular? Microbial genomics is becoming one of the fastest-growing fields of research, thanks to the advances in sequencing technology and the number of microbial genomes being sequenced are rapidly expanding. Currently, microbial (bacterial and archaeal) genomes outnumber eukaryotic genomes and will continue to do so in coming years.

Consequently, the ability to analyze massive number of microbial genomes will become a crucial skill for anyone who is working in academia, biotechnology, and government labs. This laboratory course will prepare you to gain hands-on computational skills needed to analyze microbial genomes and metagenomes with ease. You will become familiar with techniques such as genome assembly, annotation, metagenome binning, phylogenetic and phylogenomic, and learn how to write simple scripts to help you analyze genomic data. This course will also allow you to not only work on pre-existing data but also your own data if available.

This is a **cross-listed** course designed for seniors and graduate students.

Materials

Required Text

None. Instructional materials and required reading will be provided through **Blackboard** and/or **Slack**. Computational exercises will be provided on this web page:

https://microbialgenomicslab-spring2023.readthedocs.io/.

Instructions for lab activities will be updated on a weekly basis on this page and will be available on each Wednesday.

Required Materials

A **laptop computer** and reliable access to the internet. You should bring your laptop computer to class each lab period and also bring a charger in case your laptop battery is running low. The classroom has power outlets for you to plug in your computer. Please make sure you can log on to GW wireless network during classes.

You will also need to access a GW **high-performance computing (HPC) cluster** known as **Cerberus**. I will let you know how to get access to the cluster.

Additional materials

Additional reading materials and media will be provided on Blackboard and/or Slack. These may include research journal articles, news articles, YouTube videos, etc.

Course Platforms

This computational genomics lab is run fully in-person and you are expected to be present in person each lab period. I will be taking attendance and failure to show up each lab session will lead to a deduction of points for each missed classroom period. Lab sessions will be recorded using the **Echo360** classroom capture technology and the recordings will only be made available to you for valid reasons (such as having to miss the class due to testing positive for SARS-CoV-2 or due to other emergencies. We will not have a remote option for you to participate at this time.

We will also use **Slack** software platform to communicate and you will receive an invitation to join the course Slack group before the class begins.

Course Objectives

The overall goals of this course are to: (1) develop basic understanding of computational methods used to analyze microbial genomes, (2) become familiar with common bioinformatics tools used to (assemble, annotate, analyze, compare) microbial genomes and metagenomes, (3) learn basic techniques to construct phylogenetic trees, (4) learn how to visualize microbial genomic data using Python and/or R, (5) learn basic Python programming.

Specifically, at the end of this course, you should be able to:

- use command line environment
- use basic Unix/Linux commands to process biological data
- process raw biological sequence data into usable forms
- assemble microbial genomes and metagenomes
- annotate, analyze, and compare microbial genomes
- construct phylogenetic trees to assess relationships between different microbes
- identify any given gene(s) of interest and use it for further downstream processes
- learn to write simple Bash, Python, and R scripts
- use Jupyter Lab to document computer exercises
- learn Markdown syntax and write simple Markdown documents
- use git for version control

Assessment of Learning

Attendance/In-class Participation

Attendance will be taken and will count towards the final grade (20%). You will get points for attending live sessions during class hours on Wednesdays. When the class is in person, I expect you to be physically present in the classroom unless you have tested positive for COVID-19 or are sick. If you are sick or have symptoms of COVID-19, you can stay at home and watch the Echo360 recordings at a later time to catch up on learning materials. I will record attendance each class and you will lose **3 percentage points** (of total 20% possible percentage points for attendance) for each unexcused absence.

Assignments

In addition to in-class computer exercises, students will have computer assignments to complete before the next class period. There will also be weekly assigned readings and materials from the assigned readings may be on the midterm or the final examinations. Assignments due dates are listed in the syllabus. Extensions to the due dates will only be given under extreme circumstances and you will need a good reason for it. The amount of work for each assignment is carefully determined to make sure that you can complete them within one week.

You can submit your work on Blackboard multiple times until the due date/time. Any submissions received after the due date (marked late by Blackboard) will receive a 10% automatic deduction of total points possible for that assignment. If an assignment is received after 48 hours of due date, you will receive a deduction of 50% of the total points earned. If they are late for more than one week, you will get a zero for it.

Documentation using Jupyter Lab

You will use Jupyter-lab notebook to document your final project and workflows and it is worth 10% towards your final grade. Detailed instructions are given in the "Final Project/Presentation Guidelines" PDF file you have been given through Blackboard. Briefly, to reiterate the instructions, your Jupyter-lab notebook should document everything that you have done from start to finish. It should be documented in such a way that if I were to take this notebook and rerun the analyses you have done, I should be able to produce the same results you have obtained (This include details of the commands you ran, the tools you used, and any **bash** or **sbatch** scripts used in the process, for example). You should explain in detail some steps you have performed and why you did these steps. Students in BISC 4234 will work in groups (of 2-4 people) to submit group notebooks while the students in BISC 6234 will submit individual notebooks.

Final Project

BISC 4234 students will need to work in groups or in pairs (I will assign and the number of students in each group will be determined later in the semester) and work on a project assigned by the instructor or design their own project of choosing after consulting with the instructor. **BISC 6234** students will work individually. Findings from the project will need to be reported in a form of a written report. This project will count towards the final grade. There may be opportunities for you to work on original unpublished data that I generated in my lab. Depending on how novel the data and findings are, there may be opportunities for publications in peer-reviewed journals. Detailed instructions will be given for the final projects.

Final Presentation

BISC 4234 students will present their final group projects in the form of a Powerpoint presentation or something similar. Each of the group members will need to take turn to present their project in order to get points. **BISC 6234** students will present their final projects individually. Separate instructions will be given for final presentations.

Exams

There will be **two exams**: a **midterm** and a **final**. These exams will test students' comprehension of lectures, computer exercises, and assigned readings. The exams may consist of questions to be answered through Blackboard and computational exercises to be answered using a Jupyter notebook.

Amount of Effort Required

Average amount of direct instruction or guided interaction with the instructor and average minimum amount of independent (out-of-class) learning expected per week

In a traditional 15-week semester (which includes exam week), for each credit, students are expected to spend a minimum of 100 minutes on independent coursework for every 50 minutes of direct instruction for a minimum total of 2.5 hours per week. For this 3-credit course, you will need to spend 3 hours of direction instruction and a minimum of 6 hours of independent learning, totaling a minimum of 9 hours per week or 135 hours per semester. I am available for consultation during office hours (by appointment) or through email if necessary.

Evaluation and Grading

20%	**Attendance/in-class participation**
20%	**Assignments (lowest one dropped)**
10%	**Midterm**
10%	**Final Exam**
20%	**Final Project - Written Report**
10%	**Final Project - Jupyter Lab Notebook**
10%	**Final Presentation**

Grades will follow the following scale:

92.5 - 100%	=	A
88.5 - 92.49%	=	A-
85.5 - 88.49%	=	B+
82.5 - 85.49%	=	B
79.5 - 82.49%	=	B-
76.5 - 79.49%	=	C+
73.5 - 76.49%	=	C
70.5 - 73.49%	=	C-
67.5 - 70.49%	=	D+
64.5 - 67.49%	=	D
60 - 64.49%	=	D-
<60%	=	F

Course Policies

Attendance

We will meet fully in person and remote participation will not be an option for this course. If you tested positive for SARS-CoV-2, then you can stay at home and watch the Echo360 recording at a later time.

Exams

Make-up exams will only be considered for emergencies or university-approved absences, both of which will **require appropriate documentation**. In the case of university-approved activities, documentation will need to be presented to the instructor well in advance of scheduled exam date so that an alternative date before the actual exam date can be scheduled.

Academic Integrity

The course will be conducted according to established GWU guidelines of academic behavior and integrity (https://studentconduct.gwu.edu/). Academic dishonesty is defined as "cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and fabrication of information." Note that plagiarism covers both words and ideas - be sure to use proper citations for both! The minimum penalty for academic dishonesty is to receive a zero on the assignment. Academic dishonesty on the exam will result in receiving a failing grade on the exam. Major offenses will result in failure of the course or report to disciplinary committees.

sectionUniversity Policies

University Policy on Religious Holidays

In accordance with University policy, students should notify faculty during the first week of the semester of their intention to be absent from class on their day(s) of religious observance. For details and policy, see: provost.gwu.edu/policies-procedures-and-guidelines.

Academic integrity code

Academic dishonesty is defined as cheating of any kind, including misrepresenting one's own work, taking credit for the work of others without crediting them and without appropriate authorization, and the fabrication of information. For details and complete code, see: studentconduct.gwu.edu/code-academic-integrity.

Office Hours

Office hours are held in person in my office at **Lisner 413** on **Thursdays** from **1-3 pm**. However, if you are unable to meet with me during this time and need to see me another time, you can make a request to schedule a time that works for both of us. Remote options may be possible.

Support for students outside the classroom

Blackboard

Class announcements, additional readings, lab materials, and assignments will be posted on the course Blackboard site (https://blackboard.gwu
Any PowerPoint slides/lectures will be made available on Blackboard **after** the corresponding class.

The GW Library system

The library system (https://library.gwu.edu/) will be able to procure almost any book or article that you might need, but some materials may take a few days to arrive so you must plan ahead to avoid last-minute problems.

The GW Writing Center

Need help on a paper or other written assignment? Make an appointment (or drop into) the GW Writing Center (https://writingcenter.gwu.edu/).

Disability Support Services (DSS) 202-994-8250

Any student who may need an accommodation based on the potential impact of a disability should contact Disability Support Services in Rome Hall, 801 22nd Street, NW, Suite 102, to establish eligibility and to coordinate reasonable accommodations. For additional information see: disabilitysupport.gwu.edu

Counseling and Psychological Services 202-994-5300

GW's Colonial Health Center offers counseling and psychological services, supporting mental health and personal development by collaborating directly with students to overcome challenges and difficulties that may interfere with academic, emotional, and personal success. For additional information, see: healthcenter.gwu.edu/counseling-and-psychological-services

University Counseling Center (UCC)

The University Counseling Center (UCC) offers 24/7 assistance and referral to address students' personal, social, career, and study skill problems. Services for students include crisis and emergency mental health consultations and confidential assessment, counseling services (individual and group), and referrals. For additional information please visit
https://healthcenter.gwu.edu/.

Safety and security

- In an emergency: call GWPD 202-994-6111 or 911
- For situation-specific actions: review the Emergency Response Handbook: safety.gwu.edu/emergency- response-handbook
- In an active violence situation: Get Out, Hide Out or Take Out: go.gwu.edu/shooterprep
- Stay informed: safety.gwu.edu/stay-informed

Week	Dates	Topics/Activities	Assignments due
1	01/18/23	**Lecture:** Course introduction, brief history of microbial genomics **Lab:** Setting up computer environment, basic Unix commands, install Miniconda, git Jupyter Lab, install bioinformatic tools to use in subsequent weeks. Test some tools.	
2	01/25/23	**Lecture:** Introduction to sequencing technologies and bioinformatic tools **Lab:** Processing and analyzing various sequence data types, download example datasets, learn to use fastqc and bbduk, Jupyter-lab to document notes and run tests, prepare for next lab. *Assignment 1*	
3	02/01/23	**Lecture:** Microbial genome assembly **Lab:** Microbial genome assembly using SPAdes, checking assembly metrics using Quast.	*Assignment 1*
4	02/08/23	**Lecture:** Introduction to High Performance Computing **Lab:** Learn to connect to Cerberus computing cluster and submit jobs using a job scheduler. Run SPAdes assembly on Cerberus. Learn to transfer files between local and remote computers using rsync. Prepare for the next lab (prepare sequences for metagenome assembly) *Assignment 2*	
5	02/15/23	**Lecture:** Metagenomics and uncovering the hidden microbial majority **Lab:** Metagenome assemblies and binning exercises using metaS- PAdes, Megahit, and Metabat2. Assessment of bin qualities and re- finement using CheckM and RefineM. Hybrid assembly using Illumina and Nanopore data.	*Assignment 2*
6	02/22/23	**Lecture:** Inferring evolutionary relationships among Bacteria and Ar- chaea **Lab:** Using the 16S or genomic sequences to build phylogenetic trees, using GTDB-TK to identify unknown microbes. *Assignment 3*	
7	03/01/23	**Lecture:** Predicting gene function and understanding microbial metabolism **Lab:** Genome annotation using Prokka and Interproscan. KAAS and GhostKoala to identify metabolic pathways and investigate microbial metabolism.	*Assignment 3*
8	03/08/23	**Midterm**	
9	03/15/23	**Spring Break (no classes)**	
10	03/22/23	**Lecture:** Microbial genome evolution, HGT, recombination, Compara- tive genomics **Lab:** Learn to compare microbial genomes, pangenomic analysis using Anvi'o *Assignment 4*	
11	03/29/23	**Lecture:** Microbial diversity in various habitats **Lab:** Assessment of microbial diversity and population structure using DADA2 and Phyloseq in R.	*Assignment 4*
12	04/05/23	**Lecture:** Basic Python Programming **Lab:** Learning Python and Biopython to analyze sequence data, Data visualization using Python and R, learn to make publication-quality fig- ures, consult with instructor to decide on projects, start working on projects. *Assignment 5*	
13	04/12/23	**Lab:** Continue working on Python and R related exercises and consul- tation. Work on projects.	*Assignment 5*
14	04/19/23	**Lab:** Continue working on projects *Assignment 6*	
15	04/26/23	**Final Project Presentations**	*Assignment 6*
16	05/03/23	**Designated Monday**	
17	TBD	**Final Exam**	

Note that the schedule may change during the semester for unforeseen reasons.
There will also be weekly assigned reading list of peer-reviewed articles or review papers.