Genomics Lab:
Methods in Data Collection and Analysis

BSC4445C-RI
Fall 2024 (Zaragoza)
4 credits

Course Description

The field of genomics focuses on understanding the collective function of all components encoded in an organism's genomic blueprint. In the past decade, there has been an explosion of new and cost-effective methodologies to sequence the genetic material of life. Originally, high-throughput sequencing was slow, costly, and used only to sequence the genomes of model organisms. Today genome-scale datasets are essential to most molecular biology research in any taxonomic group, including phylogenetics and population genetics, metagenomic sequencing of entire biological communities, functional genomics of the expressed portion of the genome, and whole genome assembly and characterization of an organism's complete biological code. An integral, related, and emerging field of study is bioinformatics, which focuses on developing computational tools to analyze these massively large data sets. In this class, we take a hands-on approach to understanding (1) how genomic datasets are generated in the lab, and (2) how they are analyzed computationally using bioinformatics pipelines. We begin, in both lecture and lab, with the fundamental biochemistry of DNA and the tools biologists have developed to isolate and manipulate genomic elements. We then scale up to the "Next Generation" Sequencing (NGS) revolution, including how and why new methodologies have made genome-scale analyses achievable for nearly any organism, and the detailed methodologies and skills necessary to prepare samples for whole-genome, whole-transcriptome, and whole-metagenomic community sequencing. Finally, we spend the last third of the course learning and utilizing bioinformatics pipelines to manipulate and analyze the genomic datasets we have generated. Imparting a working knowledge of practical methods for generating and analyzing genomic datasets is the fundamental goal of this course. By generating, analyzing, and presenting (in both written and oral formats) novel genomic data, students will hone their writing, critical thinking and problem solving skills. The importance of genomics for all facets of life sciences will be emphasized, particularly the impacts of genomic datasets on recent advances in phylogenetics, evolutionary ecology, epigenetics, functional genomics, and health and medicine.

Research Intensive Course Designation

BSC 4445C is designated as a Research-Intensive (RI) course. This designation will be noted on your transcripts. Your active engagement in the research and/or creative scholarship process will be the core of your learning experience in this course. A significant portion of your grade for the Genomics Lab Course will be derived from both your active participation in the research process and the tangible course-related project(s) that comes out of said project. If you have any questions about this designation, please ask your course instructor.

Prerequisites
A grade of B or better in undergraduate genetics or consent of the instructor. Completion of EHS Lab Safety Training online course and practical by end of add/drop period (mandatory in order to conduct research in the Genomics Core Facility where the course is held).

Class Meetings: M/W 10:30 am – 2:20 pm

Lecture: 30-60 minutes at the start of class on Mondays, and occasionally on Wednesdays.

Lab: remainder of session on Monday and Wednesday (sometimes class will end early, but other times it may require additional time outside of class hours).

Research activities often do not fit perfectly into scheduled time blocks due to the trial-and-error nature of conducting protocols on new samples and testing hypotheses where outcomes are unknown. Thus, the Genomics Core Facility will also be open for additional hours every week for students to complete protocols or re-do failed samples, during which time the lab manager, GTA, and/or instructor will be present to answer questions and provide guidance. See office/lab hours listed below.

Instructor: George Zaragoza

Office: Biology Building 135 (within the Genomics and Bioinformatics Cluster)
E-mail: george.zaragoza@ucf.edu
Office/Lab Hours: M/W: 2:30 pm – 4:30 pm. During office hours, George will be available in BIO 414.

Graduate Teaching Assistant: Samuel Greaves

Office: Biology Building 132 (within the Genomics and Bioinformatics Cluster)
Email: samuel.greaves@ucf.edu
Office hours: T: 1:00 pm – 2:00 pm, W: 2:30 pm – 3:30 pm, or by appointment

Genomics Core Facility Manager: Chris Sarkis

Office: Biology Building 414 (during non-class hours)
Email: christine.sarkis@ucf.edu
Chris will help maintain the lab and reagents for students in the course and allow access to the lab on T/Th/F if needed.

The 414 lab manager, the GTA and/or UTA, and/or Instructor will be available for additional access to 414 and help with protocols on non-class days, as needed. However, you will need to speak with and/or email someone in advance if you plan to work during these lab hours to make sure that someone is available to help you during a

particular time. Note that other researchers use the lab on T/Th/Fri so you may need to share space and equipment during those times.

Webcourses Site

There is a course web site available through Webcourses (https://webcourses.ucf.edu) that I will use to post materials for the course, including the syllabus, calendar dates, PowerPoints, and grades.

Class Policies

1. Attendance will contribute to each student's participation grade. A large portion of the grading for this hands-on methodology course will be based on lab participation, and success in this class requires completion of lab procedures.

2. It is the student's responsibility to arrange to make up any missed work outside of class hours. With that said, it is not unlikely that some of us will become sick or exposed to COVID during this semester. If you are sick, please do not come to class. If you test positive for COVID, please call the UCF COVID line at 407-823-2509. Students will be working in groups of 2-3 on research projects and so lab work will be completed even in your absence. Please contact the instructor and we will make arrangements to catch you up over Zoom.

3. The learning environment the last few years has difficult for many of us because of COVID. As your instructor and research mentor, I want to facilitate your success in this course. If you are struggling in a way that impacts your well-being, understanding, or performance in this course, please let me know. We will work together to make sure that you can be successful and that you can get the most out of this course as possible.

4. Exams will be administered either in class or through webcourses. The format for these exams will be a) fill in the blank, b) short answer, and c) essay. You may use course materials and your notes while taking these exams, but you may not consult with each other or anyone else while taking an exam. Please notify me if you are experiencing circumstances that prevent you from completing these online assessments during the defined time-frame so that we can coordinate.

5. Assigned readings should be completed before attending class and will be provided via webcourses or handed out in class.

6. You are encouraged to discuss any and all portions of the class with me. Please feel free to come to my office hours or make an appointment to discuss the class, especially if you are having trouble.

7. This a research intensive course and we are all colleagues in this setting. Respect should be given to fellow students, the TA's, and the instructor. Please avoid arriving late to lab or leaving early.

8. Hateful or offensive speech or writing will not be tolerated.

9. Cell phones, iPods, and other electronic devices should be silenced and put away before class starts.

10. Academic dishonesty (cheating and plagiarism) is strictly prohibited and will be taken very seriously and will result at least in an "F" for that assignment (and may, depending on the severity of the case, lead to an "F" for the entire course) and may be subject to appropriate referral to the Office of Student Conduct for further action. See the UCF Golden Rule for further information.

11. AI can be a valuable learning tool, but it should not be used to misrepresent your own work. Increasingly, scientific reporting requires transparency regarding AI usage. For writing assignments in this class, please include a statement answering the following questions regarding the level of AI usage:

    a. Level 1: Was AI was used for self-study or research purposes?
    b. Level 2: Did AI provide minor writing assistance (e.g., proofreading, grammar, phrasing)?
    c. Level 3: Did AI contribute to significant writing tasks (e.g., generating sentences)?

    Levels 1-2 will generally not affect your assignment grade if you are transparent about your use of AI. If AI contributed to significant writing tasks (Level 3), you are required to specify where this occurred in your submission. Excessive use of AI for major writing tasks exceeding 5% of the total word count will result in an "F" for that assignment.

Course Accessibility

My goal for this class is to provide an accessible and welcoming experience for all students, including those with disabilities that may impact learning in this class. If anyone believes the design of this course poses barriers to effectively participating and/or demonstrating learning in this course, please meet with me to discuss reasonable options or adjustments. You may also contact SAS (Ferrell Commons 185; 407-823-2371; sas@ucf.edu) to talk about academic accommodations.

Respect for Diversity

It is my intent that students from all backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and

that the diversity that students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. Specifically:

- If you have a name and/or set of pronouns that differ from those that appear in your UCF records, please let me know.
- If any of our class meetings conflict with your religious events, please let me know so that we can make arrangements for you.
- If you feel like your performance in the class is being impacted by your experiences outside of class, please don't hesitate to come and talk with me.
- If you prefer to speak with someone outside of the course, you can contact UCF's Office of Diversity and Inclusion (diverse@ucf.edu or 407-823-6479) or the Biology Department Interim Chair, Dr. Eric Hoffman (eric.hoffman@ucf.edu).
- I am always in the process of learning more about diverse perspectives and identities, and recognize that I have limitations and blindspots. If something was said in class (by me or anyone else) that made you feel uncomfortable, please talk to me about it (or contact the resources listed above if you feel more comfortable).

Grading

There will be no traditional exams in this class – instead, there will be four in-class quizzes/assignments, each worth 5% of the total grade (20% total); in-class participation in research (30% of total grade for attending and successfully completing lab protocols); maintaining a lab notebook documenting research to the standards of a professional molecular biologist (19% of total grade); a group research project final report written in manuscript format (19% of total grade) and presented as a 15-minute oral presentation (10%); plus completion of webcourses pre and post assignments (2%).

Grades will be assigned according to the following scale:
A: 90-100%; B: 80-89%; C: 70-79%; D: 60-69%; F: <60%

The grade for this course will be based on six components (presented out of 100% total):

(1) Four semester **quizzes** will be given in class on the dates indicated on the schedule (**5% each; 20% total**). They will consist of multiple choice (30-50 pts) and short answer questions and problem solving (50-70 pts) based on lecture material. Students will be assessed in their ability to understand genomics lab methodologies, the molecular biology behind how those methodologies work, and how genomic datasets are analyzed and interpreted.

(2) Students will be graded on lab research **participation** (**30%**), including asking questions, performing protocols, and generating the necessary data for each week to move forward in the next lab session. Extra lab time outside of the scheduled four hours will be necessary to complete protocols for the many labs. Make-ups during office hours will be offered for up to two missed lab sessions; additional missed lab sessions will each result in a 5% drop in participation grade.

(3) The **lab notebook** (**19%**), to be maintained by each group of three students working on a unique project, will be evaluated every third week and given a letter grade. Students will be given feedback on how to improve their note taking and record keeping of all lab activities and results. At any time before lab notebooks must be turned in, students can also check with the instructor to determine whether lab notebook content is sufficient or lacking, and the instructor will provide feedback. The final lab notebook grade will be based on both the biweekly grades and improvement over the course of the semester to emphasize the trajectory of learning gains. The lab notebook record keeper will rotate weekly among the three group members during the nine weeks of molecular lab, and each student will be graded on their own three entries.

(4) Each group will submit a **final report** (**19%; group receives one grade**) detailing the results of their research project. The report must be written in the format of a scientific journal article, including the following sections: Abstract (5 pts) Introduction (10 pts), Methods (25 pts), Results (with figures/tables; 35 pts) and Discussion (10 pts). Groups must turn in drafts of the introduction, methods and results sections on the dates listed in the schedule below. Drafts will be evaluated and the instructor will give feedback on how to make improvements before the final draft is due. Groups will receive 15 pts of their overall final report grade for turning in drafts of these sections (5 pts for each deadline and section).

(5) During the final exam period, each group will give a 15-minute **oral presentation** (**10%**) on their project, including powerpoint slides illustrating their research. The presentation must include the following sections: overview (10 pts), introduction (20 pts), methods (30 pts), results (30 pts), and conclusions (10 pts). Each student will speak for 5 minutes and be graded individually.

(6) **Week one survey** on Webcourses to gather data and determine semester groups (**1%)** and one **post-survey** at the end of the semester (**1%**)

Schedule:
The following schedule is approximate and dates may be changed at any time.

Week 1:
Lecture 1: Intro and overview of course, overview of project and methods to be used.
Lab 1: Keeping a lab notebook, lab safety and etiquette, and pipetting.
Week 1 webcourses assignment due.

Week 2:
Lecture 2: Basics of molecular biology: Isolating and characterizing nucleic acids
Lab 2: Dissections/digestions and DNA extractions.

Week 3:
LABOR DAY – NO CLASS MONDAY
Lecture 3: Polymerase Chain Reaction (PCR) and electrophoresis.
Lab 3: DNA extractions. DNA quantification using Qubit.

Week 4:
Lecture 4: Genomic and cDNA libraries. **Quiz 1 (Wed).**
Lab 4: Polymerase Chain Reaction (PCR) and gel electrophoresis.

Week 5:
Lecture 5: Overview of DNA sequencing methods
Lab 5: Metabarcoded PCR. Metabarcoded gel electrophoresis. Bead cleanups to remove primer dimers.

Week 6:
Lecture 6: Illumina Library Prep. **Project outline due**
Lab 6: Catch-up/repeat week.

Week 7:
Lecture 7: Analyzing sequencing data: Quality control, NCBI, GenBank, and BLAST;
Lab 7: qPCR on serial dilutions of group library pools to precisely quantify the amount of available material for sequencing, equimolar pooling of group libraries, final prep of libraries for sequencing, QC of group pools by TapeStation.

Week 8:
Lecture 8: Bioinformatics overview; **Quiz 2 (Wed)**
Lab 8: Command Line Tutorial

Week 9:
Lecture 9: Whole genome assembly, genomes and genome browsers
Lab 9: Intro to Geneious and gene identification using GenBank

Week 10:
Lecture 10: Genome-scale analyses: RadSeq, SNPs and GWAS.
Lab 10: Bioinformatics pipeline tutorials for QIIME2

Week 11:
Lecture 11: Microarrays, transcriptomics, and RNAseq **Quiz 3 (Wed)**
Lab 11: Manipulating Illumina data and Quality Control (QC) using FastQC; start running pipeline on lab-generated data (students work in groups).

Week 12:
Lecture 12: Metagenomics and metabarcoding
Computer Lab 12: QIIME on lab-generated data.

Week 13:
Lecture 13: Presenting genomic data
Lab 13: Continue running pipelines on lab-generated data; analyze results/generate summary statistics (students work in groups). **FULL DRAFT DUE (Wed)**

Week 14:
Lecture 14: **Quiz 4 (Wed)**
Lab 14: Finish/troubleshoot lab-generated data analysis; finish analyzing results and work on tables, figures and reports (students work in groups).

Week 15:
Monday, November 25: **FINAL REPORT DUE & Final group presentations**