Computation for Biologists – MARB-6360
Department of Life Sciences
Fall 2025

A. COURSE & INSTRUCTOR INFORMATION

Course number/section: 6360.1
Class meeting time: F 2-4:30
Class location: OCNR 241
Course Website: https://github.com/tamucc-comp-bio/classroom_repo_2025

Instructor: Dr. Christopher E. Bird
Office location: TH 234
Office hours: M-W 4-5:30, F 4:30-5
Telephone: 361-443-5676
e-mail: cbird@tamucc.edu

Appointments: A student may make an appointment to see me at times other than the scheduled office hours. I am available for consultation and extra help, but it is the student's responsibility to request such help. Book an appointment

B. COURSE DESCRIPTION

Catalog Course Description
This course is designed to prepare and enable students to use computational tools for bioinformatic applications in advanced courses and independent research projects. Students will be introduced to powerful open-source computing tools used in biological research for creation, organization, manipulation, processing, analysis, and archiving of big data. While not a formal requirement, it is assumed that students have a firm command of basic algebra.

Extended Course Description
This is a 3-credit course that introduces the powerful open-source computing tools that are used in biological research for the creation, organization, manipulation, processing, analysis, and archiving of "big data". This course is designed to prepare and enable students to use computational tools for bioinformatic applications in advanced courses and independent research. The primary topics covered are: data formats, organization, and repositories; command line Linux computing and scripting with BASH; regular expressions; super-computing; data wrangling with BASH and R; data visualization with R & Python; version control and dissemination of scripts and programs with GIT and GitHub; and typesetting with Markdown. The usage of large language models (e.g. ChatGPT) will be incorporated throughout the course.

Whether you want to learn basic data handling skills for your research project or you are curious about a career in bioinformatics and "big data", this course will provide you with the proper foundations.

C. PREREQUISITES AND COREQUISITES

Prerequisites
"None"

Corequisites
"None"

D. REQUIRED TEXTBOOK(S), READINGS AND SUPPLIES

Required Textbook(s)
R for Data Science. Grolemund and Wickham. (Free), https://r4ds.hadley.nz/, https://r4ds.had.co.nz/

Associated Website, Optional Textbook(s) or Other References
Computing skills for biologists: a toolbook. Allesina & Wilmes 2019.

https://github.com/tamucc-comp-bio/classroom_repo_2025/blob/master/resources

Supplies
Windows, Mac, or Linux computer (not chrome, not iOS, not Android). Recommended at least 8gb RAM, 512 GB storage. If you don't own a computer, make arrangements with Dr. Bird – there are resources available.

E. STUDENT LEARNING OUTCOMES AND ASSESSMENT

Assessment helps improve learning by providing feedback to both students and instructors. It starts with clearly defined student learning outcomes, which outline what students should achieve to succeed in the course. Sharing assessment data helps students focus their efforts and allows instructors to adjust teaching to address learning challenges.

By the end of this course, students should be able to:

1. ***Recognize, describe, and organize data into "tidy" data structures***
2. ***Locate scientific data repositories and download data***
3. ***Operate UNIX/LINUX (super)computers from command line***
4. ***Construct and modify computer programming/scripting logic structures for processing biological data***
5. ***Describe and use regular expressions to query data***
6. ***Use version control software (git) in coordination with GitHub to organize and manage projects***
7. ***Typeset with LaTeX or MarkDown***
8. ***Use the most popular open-source tools for biological data manipulation***
    a. ***Shell scripting (bash)***
    b. ***Statistical computing (R, tidyverse)***
    c. ***Scientific computing (python)***
9. ***Use large language models effectively to assist in data processing and teach others how you automated your data processing and analysis.***
10. ***Automate the processing/analysis of a data set and create a GitHub repository that could be submitted as supplemental material for a manuscript to a peer-reviewed journal***

F. INSTRUCTIONAL METHODS AND ACTIVITIES

Classroom time involves a mix of lecture and interactive hands-on computing exercises. Weekly assignments are due at the beginning of the weekly lecture.

G. MAJOR COURSE REQUIREMENTS AND GRADING

Computation for 21st Century Biologists will convene once a week for 2.5 hours. Class periods will involve interactive lectures that require each student to have a computer designed for content creation (Linux, OSX, Windows, not chrome, not iOS, not Android). Homework exercises will embellish upon concepts addressed in lecture. Participation involves attending lectures and performance on unannounced quizzes. Weekly Assignments will be given to reinforce concepts covered in lectures and encourage students to start using computational tools. Exams will be used to evaluate comprehension of the materials covered in lectures and assignments.

Rather than having a final exam, graduate students are expected to complete a Final Project involving the automation of the manipulation and/or analysis of data, This project, including the code should be archived on GitHub. A report written in Latex or Markdown will be due during the final exam period (this can be a markdown document in your GitHub repo, such as the README.md. The report should be concise in stating what the problem is, describing the strategy used for the solution, and describing how the code works (be sure to include a flow-chart or outline describing what code does). Those taking MARB 6360 will give a 10-15 minute presentation during the Final period on their project.

Project examples: automatically process data from experimental apparatus; image analysis; automated reporting of experimental results; downloading and organizing data from online repositories; etc…

Student learning outcomes will be assessed using in-class exercises, assignments, and exams. Your final grade will be based on the percentage you earn out of the total possible points, extra points may be built into exams or other assignments. It is also possible to lose points by turning in assignments late. Statistical manipulations to adjust grades, if used (at the Instructor's discretion), will be performed for each exam individually and all assignments in aggregate. A standard grading scale will be used:

A = 90 - 100 %
B = 80 - 89.9 %
C = 70 - 79.9 %
D = 60 - 69.9 %
F = 0 - 59.9 %

ACTIVITY	% of FINAL GRADE
Participation	10
Assignments	30
Exam 1	10
Exam 2	10
Final Project Report Final Project Presentation	30 10

H. COURSE CONTENT/SCHEDULE

Date	Lecture Topic	HW Due
	Theme I: Welcome to the Matrix	
Wk 0	1. Course overview 2. Biological Data Repositories, Structures, Formats 3. Computer set up	
Wk 1	Linux Boot Camp I 1. UNIX philosophy 2. Navigating/creating/manipulating directories & files	Assignment 0

	3. How to get help: man pages 4. Basic commands: cd, ls, cp, mv, mkdir, rm, tr, cut, cat, head, tail, 5. Commands useful for manipulating data files in text streams	
Wk 2	Linux Boot Camp II 1. Wildcards, substituting characters, permissions, sudo 2. Pattern matching with grep & regex 3. Intro to Computer Programming: Shebang!, Scripting, For Loops	Assignment 1
Wk 3	Linux Boot Camp III 1. More Computer Programming with bash 2. Logic: if-then-else; looping with while, and GNU parallel 3. Functions: diy commands 4. Advanced text stream manipulation: sed, paste, …	Assignment 2 SuperComputer Acct
Wk 4	Version Control & Supercomputing 1. Linux repositories and tools for biologists 2. Version control with git 3. Super computing, ssh	Assignment 3
	Theme II: Wrangling and Visualizing Data With R	
Wk 5	R Boot Camp I 1. R Philosophy 2. Command line R 3. R data types & structures 4. Math, equalities, logic 5. Basic statistical functions 6. Reading and writing data	**Exam 1** Install R & R Studio
Wk 6	R Boot Camp II 1. Scripting & writing good code 2. Loops & if-then decision logic 3. R Studio 4. Functions 5. Libraries 6. Random numbers 7. Vectorized loops 8. Debugging 9. More basic stats 10. Base R plots	Assignment 5
Wk 7	R Boot Camp III 1. tidyverse 2. Basic reading & manipulating data 3. Basic computing statistics 4. Visualization of data w ggplot2	Assignment 6
Wk 8	R Boot Camp IV 1. Advanced tidyverse, pipelines 2. Manipulating & wrangling data	Assignment 7
	Theme III: Programming the Matrix	
Wk 9	Python Boot Camp I 1. Intro to Python	**Exam 2** Install Miniconda

	2. Data structures 3. Functions 4. Decision logic and loops 5. Reading and writing files	
Wk 10	Python Boot Camp II 1. Writing code 2. Modules & Program Structure 3. Errors and exceptions 4. Debugging 5. Testing & Profiling	Assignment 9
Wk 11	Scientific Computing w/ Python 1. NumPy and SciPy 2. Pandas 3. Biopython 4. Other modules	Assignment 10
Wk 12	Scientific Typesetting w/ Latex 1. Latex document structure 2. Typsetting 3. Latex packages for biologists	Assignment 11
Wk 13	Putting It All Together	Assignment 12
Final	**Final Exam: Becoming THE ONE**	

Date	Graduate Students: Final Project Schedule
	Theme I: Welcome to the Matrix
Wk 0	
Wk 1	
Wk 2	Submit Project Idea (2.5%)
Wk 3	
Wk 4	Submit Project Plan/Outline (2.5%)
	Theme II: Programming the Matrix
Wk 5	Link to GitHub repository for project w/ readme (2.5%)
Wk 6	Commit at least 1 working function to GitHub (2.5%)

Wk 7	Commit at least 2 working functions with data I/O to GitHub (2.5%)
Wk 8	
	Theme III: Becoming THE ONE
Wk 9	
Wk 10	Latex or Markdown draft/ progress report (see syllabus section G for description of report); include a description of tasks left to achieve. Code and data committed to GitHub (2.5%)
Wk 11	
Wk 12	
Wk 13	
Last Day of Classes (Dec 3)	Final Report in Latex or Markdown; Working code and data committed to GitHub. (25%)
Final	MARB 6360 Final Presentations

Note: Changes in this course schedule may be necessary and will be announced to the class by the instructor. The assignments and exams shown are directly related to the Student Learning Outcomes described in Section E.

I. COURSE POLICIES

Attendance/Tardiness

Attendance is expected. If you are late, don't make a disturbance and you will be responsible for catching yourself up to where we are. If you cannot attend a lecture, make arrangements with Dr. Bird

Late Work and Make-up Exams

10% of total possible score is deducted per day late. Inform professor as soon as you find out that you will miss and exam. Make arrangements with professor for make-up.

Extra Credit

None

Cell Phone Use

Required

Laptop Use

Required

Food in Class
Not allowed in computer lab

Participation
Attendance at lecture is mandatory. Make arrangements with Dr. Bird if you cannot attend.

J. UNIVERSITY & COLLEGE POLICIES

For information on:

- Campus Emergencies
- Statement of Academic Continuity
- Academic Integrity/Plagiarism
- Classroom/Professional Behavior
- Statement of Civility
- Civil Rights Reporting
- Disabilities Accommodations
- Mental Health and Well-Being Services
- Academic Advising
- Dropping a Class
- Student Grade Appeals
- Copy of Academic Calendar

Please visit https://www.tamucc.edu/science/student-information/syllabi-policies.php or scan:

K. AI STATEMENT

Students are expected to use "artificial intelligence" (e.g. large language models) in this course. When you submit work created using artificial intelligence, you are responsible for the content, not the ai. If you used ai in a novel fashion, you should describe how it was used to accomplish the task so that Dr. Bird can learn and incorporate it into future lectures/assignments.

L. GENERAL DISCLAIMER

I reserve the right to modify the information, schedule, assignments, deadlines, and course policies in this syllabus if and when necessary. I will announce such changes in a timely manner during regularly scheduled lecture periods.