

Course Description

This course focuses on computational skills for processing data using programming language Python and UNIX environment. No prior programming experience is required, but some basic computer skills will be useful.

With the advancement of high throughput data generation methods, a major challenge that graduate students in life sciences have to face today is to analyze large amount of biological data. The objective of this course is to provide an opportunity for graduate students with no computer science background to learn the basic skills of handling high throughput biological data. It covers the Linux/Unix environment and the importance of the command line interface; the Python programming language; program design, implementation, and testing; BioPython; Strategies for analyzing genome resequencing, RNASeq, and microbiome sequencing data. Students build hands-on skills by analyzing real high throughput biological data through homework assignments and team projects.

Units: 3

Instructor: Jason Stajich (jason.stajich@ucr.edu)

Time and location: Tu 12-2PM / Thur 1-3PM 3365 Spieth Hall (Darwin Room)

Office Hours: Wed 9-10 and best by appointment.

https://biodataprogramming.github.io/GEN220_2026/

Prerequisites

- Coursework in genetics or molecular biology or permission of instructor

Resources

None of these texts are required for completion of the course but they will provide a great deal of helpful background and examples that will improve your ability to master UNIX or Programming in Python.

1. *Bioinformatics Data Skills: Reproducible and Robust Research with Open Source Tools*. Vince Buffalo. 2015 O'Reilly & Associates. Available from [O'Reilly and Associates](#), [Amazon](#) Free to read on UCR network (or use VPN) - [Safari link](#).
2. *Unix and Perl to the Rescue: A Primer*. Keith Bradnam and Ian Korf. [Unix and Perl Primer for Biologists](#)
3. *Unix and Perl to the rescue!* Bradnam and Korf. [Amazon](#)
4. [Rosalind](#) - An online platform to learn bioinformatics and programming in Python.
5. Software Carpentry - <https://software-carpentry.org/> and Data Carpentry - <http://www.datacarpentry.org/>.
6. Berk Ekmekci, Charles E. McAnany, Cameron Mura. An Introduction to Programming for Bioscientists: A Python-Based Primer. PLoS Comp Bio. DOI: [10.1371/journal.pcbi.1004867](https://doi.org/10.1371/journal.pcbi.1004867)
7. Ken Youens-Clark. Tiny Python Projects. <https://www.manning.com/books/tiny-python-projects>
8. Pat Schloss's Riffomonas Code Club has great [videos and links](#) to programming and microbiome analyses.

Grading

- Homework assignments (6 in total, HW0-HW5): 50% of grade
- Project: 50% of grade

Homework

- Homework is due before class on the date listed in the syllabus or if amended when assignment is given.
- There will be a programming assignment every two weeks during the first half of the course. Programming assignments must be prepared along with any necessary input files or documentation to demonstrate program usage.
- Code should be runnable as turned in. You will deposit your code in your github repository or if not possible, by Canvas. You can make one private personal repository to deposit and should organize a folder for each homework assignment (e.g. hw1, hw2, hw3, hw4, hw5). There will be a link to create these through GitHub Classroom and posted in [Canvas](#).

Technology Requirements

Because this course requires use of a computer, you will need the following:

Hardware

Access to a current Mac or PC (with a fast processor and speakers) Webcam and microphone (to participate in any video components, e.g. live sessions, remote proctoring, video presentations)

Operating systems

- OSX is recommended with [Xquartz](#) installed for X11
 - [iTerm](#) is a useful terminal in addition to default Terminal.
- Windows with [MobaXterm](#)
- UNIX or Linux eg [Ubuntu](#)

Other recommended software - a local text editor

- [Visual Studio](#) highly recommended
- [Notepad++](#)

Projects

- Project Topics will be discussed in October and teams will select a project idea to focus on.
- Project will be 2-3 individuals working together.
- A presentation will be made by each team - last day(s) of class.
- A final report with the details will be turned in by the group.
- The report needs to detail what each person's contribution is to the project.

Schedule

Date	Day	Lecture Topic	Notes
Sep-24	Th	Course Intro / UNIX I: Logging into HPCC, command line basics; GitHub account and SSH keys	Homework 0 assigned
Sep-29	Tu	UNIX II: Files, data and running programs on the HPCC cluster (modules, srun; sbatch preview); Git and GitHub: clone, commit, push	Homework 1 assigned
Oct-1	Th	UNIX III: Shell programming - variables, loops and scripts (lab/reading: data processing with cut, sort, uniq, awk)	Homework 0 Due

Date	Day	Lecture Topic	Notes
Oct-6	Tu	Python I: Variables, running, cmdline, strings, math	
Oct-8	Th	Python II - Logic, loops, lists, iterator; I/O reading/writing files	Homework 1 Due; Homework 2 assigned
Oct-13	Tu	Python III - Dictionaries and Functions	
Oct-15	Th	Bioinformatics I - Sequence search: BLAST on the command line	Homework 3 assigned
Oct-20	Tu	Python IV - Modules, BioPython and Pandas	Homework 2 Due
Oct-22	Th	Python V - Regular Expressions. Class Project Info	
Oct-27	Tu	Bioinformatics workshop - parsing BLAST and data tables in Python	
Oct-29	Th	Running analyses on the cluster: SLURM job scripts and job arrays	Class Project Outline/Abstract due
Nov-3	Tu	Bioinformatics II - Short read alignment and genomic ranges (BWA, samtools, bedtools)	Homework 3 Due
Nov-5	Th	Bioinformatics III - RNASeq analyses	
Nov-10	Tu	Building websites with GitHub Pages, Markdown and HTML	Homework 5 assigned
Nov-12	Th	Bioinformatics IV - SNPs and variants	Homework 4 Due
Nov-17	Tu	Bioinformatics V - Protein domains and homology search (HMMER, Pfam)	
Nov-19	Th	Bioinformatics VI - Orthology and phylogenetics	
Nov-24	Tu	Bioinformatics VII - Sequence evolution and statistics: Ka/Ks, simulation, shuffling and p-values	Homework 5 (GitHub website) Due
Nov-26	Th	** NO CLASS ** - Thanksgiving	
Dec-1	Tu	Genome and Statistical Data visualizations	Extra Topics
Dec-3	Th	Class Presentations	
Dec-9	Wed	Final papers due	

*note these dates and topics may changes if illness or conflict arises. Class will also vote to emphasize special topics towards the end of quarter.