

Biology 399: Computational Genetics

Biology is now a computational field, where large, multidimensional datasets are the norm. In this class we introduce quantitative approaches to problems that arise naturally in genetics and genomics through a computational lens. The course is project based, with each week centering around a new problem in biology. Students will write code in Python to visualize and statistically analyze datasets that result from modern experiments such as genome sequencing and RNA transcription profiling. Students are **not required** to have any previous experience with programming– if you do have previous experience consider this an excellent opportunity to practice “Beginner’s mind”. The course will be taught in an interactive, workshop-like environment where we will go over code together. This environment will benefit the active participant, so please do your best during class to concentrate on the material at hand rather than blowing it off until later (even if Andy is terribly boring). Below is the tentative schedule for the course. Depending on the speed we go through topics there might be shifts in the schedule.

Learning objectives for this class

- Gaining beginner level fluency with the Python programming language
- Develop skills at dealing with biological data, from importing large datasets, to visualization, to fitting
- Gain appreciation for a broad survey of quantitative methods in genetics / genomics
- Do some actual research on SARS-CoV-2!

Virtual Office Hours

So this is new, and we’ll see how it goes, but I’m planning on putting up a Zoom link in the Canvas announcements for office hours. The current schedule is Tuesday at 2-3pm and Friday from 3-4pm, but I’m also available at other times so it might be equally easy to just arrange an appointment with me.

Requirements and Evaluation

The class is organized around weekly problem sets which our in class work and discussion will prepare you for. Problem sets will be a combination of coding and analytical work, and will be turned in via preparing a jupyter notebook and sending either a path on our server or the notebook itself via email to Andy. My goal is to have the problem sets be challenging and interesting for you but not evaluative– think pass/fail. There will be no quizzes or tests. There will be a final group project where you will partner with 1 or 2 other students to do a bit of research on your own and then present to the class.

Discussion Section

Discussion will be a chance for further review of the programming material with Elliott (eabe@cas.uoregon.edu), our talented TA. Material in discussion section is meant purely to fill in details of what we cover in lectures, not go beyond what we have learned in class.

Final Project

The last week of class will be small groups presenting their final projects to the class. We will organize into groups for this in Week 5. The idea will be for the group to come up with it's *own research idea*, centered around the themes of the class and then take that idea to Andy for approval. With approval in hand the group will go off, do their bit of research, and then present it to the class in a 15min presentation during the last week of class. Some example ideas for the final project might be like: creating a phylogenetic tree of a SARS-CoV-2 and interpreting it, doing simulations of the SIR model, developing a classifier to differentiate genetic variation, etc.. This should be fun, not a slog.

Class Courtesy

This is going to be key to the success of this Zoom format. Please “arrive” on time ready to be engaged. Ask questions if you did not hear or understand something. Classroom courtesy and sensitivity are especially important with respect to individuals and topics dealing with differences of race, culture, religion, politics, sexual orientation, gender, gender variance, and nationalities. Class rosters are provided to the instructor with the student’s legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the quarter (or before) so that I may address you properly. Our classroom is a learning environment, and as such should be a safe, inclusive and respectful place. Being respectful also includes using preferred pronouns for your classmates. Disrespecting fellow students as well as combative approaches, tones and/or actions are not acceptable. Please make me aware if there are classroom dynamics that impede your (or someone else’s) full engagement.

Students with Disabilities

I strongly encourage students with disabilities, including “invisible” disabilities like chronic diseases, learning disabilities, and psychiatric disabilities to discuss with me as soon as possible what appropriate accommodations might be helpful to them. You are also encouraged to contact the Accessible Education Center in 164 Oregon Hall at 541-346-1155 or uoaec@uoregon.edu.

Week 1. Computational Genetics Intro

Basics of nomenclature, our server, jupyter notebooks, python baby steps, Printing, data structures, indexing, control flow, conditionals, looping, functions. Case study: plotting COVID cases

Week 2. Tracking the spread of SARS-CoV-2

Using numpy and scipy to boost our productivity. ndarrays, indexing, etc. Case study: fitting exponential models to growth of SARS-CoV-2

Week 3. Dealing with sequence data

IMPORTANT – no class on 4/15. Using Biopython, managing sequences, alignment Case study: aligning SARS-CoV-2 genomes, extracting genomic features

Week 4. Computational Phylogenetics

Trees and tree thinking, clustering, Maximum likelihood principle, ML phylogenetics Case study: building trees from SARS-CoV-2

Week 5. Unsupervised clustering

Non tree-like ways to cluster, UPGMA, PCA, UMAP? Case study: clustering SARS-CoV-2 and visualizing these clusters

Week 6. Modeling evolution

Using our skills to simulate some the population genetic model of genetic drift and natural selection. Case study: calculating time to fixation of neutral mutations vs. selected mutations

Week 7. Modeling infectious disease

IMPORTANT – may cancel class on 5/13. TBD. SIR and SEIR models, numerical integration of ordinary differential equations, SciPy, spatial SIR model? Case study: simulating SIR model, estimating COVID params

Week 8. Genome-wide association studies

Genotype to phenotype mapping, linear models, GWAS Case study: host genetic variation and its impact on COVID

Week 9. Single cell RNA-seq

Clustering con't, applications to high dimensional data and in particular scRNA Case study: scRNA-seq from COVID-10 infected cells

Week 10. Final Project Presentations

Yay let's see everyone's cool work.