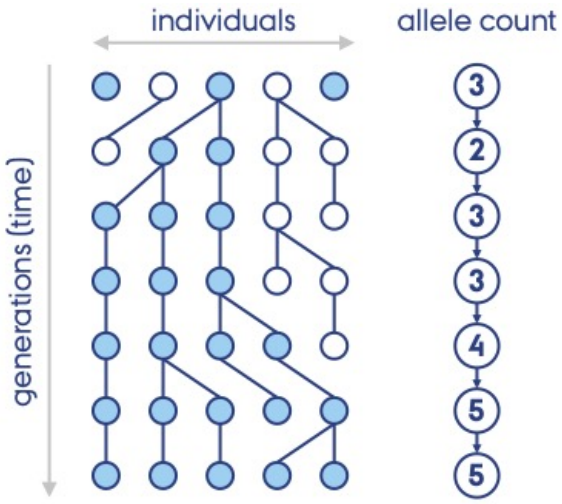
Course Syllabus

Bio 35: Computational Biology



A depiction of a population model, demonstrating changes in the frequency of genetic variation over evolutionary time.

Image credit: Dr. Paula Tataru (**https://www.slideshare.net/PaulaTataru/paulatataruvienna (https://www.slideshare.net/PaulaTataru/paulatataruvienna)**)

Course Description

Biological questions have fascinated and preoccupied our species throughout our history. We have pondered what makes us human, attempted to explain the mechanisms driving biodiversity, and sought to develop technologies that can treat diseases or control our environment. Until quite recently we had relatively little data to help us pursue these goals, but many competing theories and models to explain biological observations. A revolution in biological data generation is underway, with new genomes being sequenced at astonishing rates. Other types of data (including digital images, biological annotations, protein structures, etc) are also rapidly increasing in scale.

This massive size of biological datasets requires automated summarization and responsible interpretation of biological data. In this course, we will ask *how can biological questions be posed and answered using computational approaches?* We will approach this question through the lens of current open questions in evolutionary biology and molecular medicine. We will develop basic programming skills and learn about several widely-used computational methods. More importantly, we will develop a conceptual basis for asking and answering biological questions using computational tools, and for assessing the limitations of these tools.

Learning Objectives

The course will explore several topics in contemporary biology. Each module will follow a general outline of asking a biological question, recasting it as a computational problem, proposing a solution to the computational problem, and finally implementing the solution in computer code. We will introduce real or simulated datasets for each of these problems.

- *Ask a biological question*: We will focus a substantial portion of the course around ongoing discussions in both popular science journalism and the scientific literature. By the end of the course, students should be able to read a journalistic article about a problem in computational biology, identify the research question being posed, and describe why a computational approach may be needed for this problem. They should be able to broadly describe several open biological/evolutionary questions that researchers are currently investigating using computational approaches.

- *Reduce a biological problem to a computational problem*: Many biological questions can be restated as computational problems once the appropriate data have been collected. Students should be able to take a biological question and suggest approaches for recasting it as computational problem (within the realm of a range of topics discussed in the course).

- *Propose a computational solution*: Having identified the biological problem as a computational problem, students should be able propose a computational solution (i.e., an algorithm). This is not a computer science course! We will not focus on correctness proofs or computational complexity. Rather we will attempt to provide intuitive explanations for the operation of algorithms, and we will expect that students should be able to explain how/why the algorithm works.

- *Implement a solution*: We will use Python to implement solutions to the computational problems introduced in the course. Students will use real or simulated datasets to answer the questions posed. While we will emphasize basic coding principles, writing beautiful code will not be a main focus of the course.

Course Instructors & Contact Info

- Instructor: Lawrence Uricchio (he/him/his), **Lawrence.Uricchio@tufts.edu (mailto:Lawrence.Uricchio@tufts.edu)**
- Teaching Assistant: Jacob Dayton, **Jacob.Dayton@tufts.edu (mailto:Jacob.Dayton@tufts.edu)**

We will generally reply to course related emails within 1 business day, but it may occasionally take 2 business days. If we have not replied within 2 business days please resend your message.

Emails about course content should start with a subject line of "BIO35:" followed by any subject you wish to give the email, to ensure they are not lost in our inboxes.

Zoom link

Thursday classes: https://tufts.zoom.us/j/92287073231?pwd=bXJBWmF0NWRCeG5ucGxTVmVRYTBkdz09 (https://tufts.zoom.us/j/92287073231?pwd=bXJBWmF0NWRCeG5ucGxTVmVRYTBkdz09)

Tuesday classes: https://tufts.zoom.us/j/91804797214?pwd=ZGJZR2xBVFVnSmZyV0NVcFlaVUtzUT09 (https://tufts.zoom.us/j/91804797214?pwd=ZGJZR2xBVFVnSmZyV0NVcFlaVUtzUT09)

Office Hours

- Dayton: Mondays (11am-12pm) and Wednesdays (1:30-2:30pm) remotely over Zoom (https://tufts.zoom.us/j/9044684121)
- Uricchio: Tuesdays (4pm-5pm) and Wednesdays (1pm-2pm) at https://tufts.zoom.us/my/lawrence.uricchio (https://tufts.zoom.us/my/lawrence.uricchio)

Software

We will use several software tools in this class, but most importantly we will be writing our software in Python3. You will need access to a computer on which you can code Python3, or alternatively you will need to login remotely to another machine that has Python3.

Below please find a list of software tools that we will use throughout the course. While it may be helpful to download and install each of these tools before the beginning of the course on your own laptop, we will use the Tufts HPC (high performance computing) as our primary resource for programming in this course, which will have these tools pre-installed for your use. If you are trying to install Python3 and have any difficulty please contact the instructors (see "contact" section below). Please note that Python may already be installed on your machine.

- Python3 (https://www.python.org/download/releases/3.0/)
- numpy (https://numpy.org/install/)
- SLiM (https://messerlab.org/slim/)
- stdpopsim (https://stdpopsim.readthedocs.io/en/latest/)

Materials

We will use the textbook *Computing for Biologists* for most of sections in this course. Occasional additional readings will be assigned and distributed as PDFs to the class. The book should be available from the bookstore. If you are having trouble obtaining a copy for any reason please contact the course instructors. This includes both logistical challenges and any hardship that makes obtaining a copy challenging.

- *Computing for Biologists: Python Programming and Principles*, Ran Liebskind-Hadas and Eliot Bush. **link here (https://books.google.com/books/about/Computing_for_Biologists.html?id=E8HOBAAAQBAJ)**
- Course Website: Canvas (https://canvas.tufts.edu/courses/31670) – All course materials will be put on this page. This is the *only* official website for this course.
- Laptop: please bring your laptop to class if you have one and are able. The laptop is not strictly required, but may be helpful for following along with the material, and so that you can ask questions in the zoom chat box. If you do not have a laptop and would like to use one in class please let us know as soon as possible so we can request support.

Additional resources

A few optional resources that may provide insights not available from your textbook. We will update this list throughout the term. If you find a helpful resource please feel free to message your TA or instructor so we can share it with the class. Only free resources will be posted to this list.

- **Tufts Python and data science resources (https://sites.tufts.edu/datalab/python/#setup)**
- **Windows commands (https://uricchio.github.io/CompBio/assets/windows_command_line_sheet_v1.pdf)**

Assignments

There will be three types of assignments in this course:

- *Coding assignments*: There will be 8 coding assignments throughout the semester, approximately one every 1.5-two weeks. With some exceptions, these assignments are adapted from *Computing for Biologists.*
- *Scientist spotlights*: Asking biological questions is a central component of this course. Each week we will cover the work and life of a contemporary scientist working in computational biology. We will read one of their academic papers, discuss how their work relates to the broader computational/biological topics that we are exploring, and learn about their career trajectory and

life. Students will be responsible for writing a short reflection on each scientist. This reflection should be about 150-200 words. We won't penalize you for writing 100 or 400, but students who are consistently way outside the guidelines may get a warning to readjust.

- *Meeting with instructor*: We will have two required one-on-one zoom meetings with the instructor. These are pass-fail, and all you have to do to get credit is show up. Their purpose is to help us assess how we (i.e., the instructional team) are doing, and what we can do to support you in the classroom and beyond.

Grading

Scores for each assignment will be returned within about 1 week of the due date. The course instructor (Prof. Uricchio) will be the primary grader. Additional qualitative feedback may be provided by TA Dayton in some instances.

Generally speaking, our goal in the course is for most students to be able to get a B or better if they consistently complete the assignments. After each coding assignment, the course instructors will post the histogram of scores (i.e., a plot of the distribution). We expect to use a standard grading scale (90-100 for an A, 80-89 for a B, 70-79 for a C, etc). However, please note that this is the first time this course is being taught! We reserve the right to curve grades UP if a substantial portion of students score much lower than expected. We will NOT curve grades down. If/when assignments are curved up we will alert the whole class at the time when the scores are released.

Spotlight assignments are graded on a binary basis (1 or 0 pts).

- Coding assignments: 70% (10 points each, lowest score dropped)
- Scientist spotlights: 20% (2 points each, you can skip up to three with no penalty. Complete all 13 and get two bonus points, i.e. spotlights will be 22% rather than 20% of your grade)
- Midterm meeting with instructor: 10%

Late policy

In general, we will allow a grace period of three business days for the first two late CODING assignments, no questions asked (i.e., 72 hours from the time the assignment is due). In other words, if an assignment is due on Friday and you don't turn it in until Wednesday, that's OK in the first two instances and you will not be penalized. Subsequent late assignments will be subject to a 20 percent penalty per day late. This policy applies to the CODING assignments, not the spotlight assignments. Late spotlight assignments will not be accepted, but note that you can skip up to three without any loss of points.

Group Work

You are always welcome to work with up to two other class members on each assignment. Please indicate which class members you have collaborated with at the top of each submitted assignment. Failure to acknowledge that you worked with a classmate could be considered a violation of academic integrity. If we notice similarly worded assignments that do not acknowledge team effort we will follow up with you. Any students who are looking for partners to work with may contact the course instructors and we will facilitate connections between interested students.

What isn't in this course?

This isn't a course about software design, computer science, statistics, or probability. While we will spend a little time talking about each of these topics, this is fundamentally a biology course, about questions that arise in biology. We will not emphasize software design -- if this an interest of yours, you will find many courses in engineering or computer science that emphasize these concepts. Moreover, it is not a course about rigorous algorithmic analysis or data structures. Statistics and probability will be covered in some sections in an applied, intuitive way, but we will not be proposing theorems or writing proofs.

We include this statement for two reasons: first, in some cases we may ignore good software design principles for the sake of efficiency. We will take the attitude that our software must work "for us" and we are not developing for external users or large development teams. We will highlight some examples where this matters and might affect how we code. If you have taken a CS course you may find some examples yourself, and we will not ding you for inefficient or unnecessarily long code.

Second, we are focused primarily on biological questions in this course. We will cover computational methods as an elegant tool to answer biological questions. In so doing, we will sometimes pass over interesting aspects of the computational approaches themselves.

All of these aforementioned engineering and quantitative skills are valuable and important, and we exclude them only because there is so much interesting biology to talk about! If you are interested in them I highly recommend taking a CS, math, engineering or statistics course that covers them!

Mental Health Information

As a student, there may be times when personal stressors or emotional difficulties interfere with your academic performance or well-being. The Counseling and Mental Health Service (CMHS) provides confidential consultation, brief counseling, and urgent care at no cost for all Tufts undergraduates as well as for graduate students who have paid the student health fee. To make an appointment, call 617-627-3360. Please visit **the CMHS website (http://go.tufts.edu/Counseling)** to learn more about their services and resources. Please note that we have designed the curriculum to accommodate flexibility in variations in well-being and academic performance – see the late-work policies above for more information.

In addition to personal challenges, there may be events that make it challenging for all of us to work. We will try to be sensitive to these types of events and provide leniency for the whole class under such circumstances. We encourage the students to make us aware of any such events if we miss them.

Academic Support at the StAAR Center

The StAAR Center (formerly the Academic Resource Center and Student Accessibility Services) offers a variety of resources to all students (both undergraduate and graduate) in the Schools of Arts and Science, Engineering, the SMFA and Fletcher; **services are free to all enrolled students**. Students may make an appointment to work on any writing-related project or assignment, attend subject tutoring in a variety of disciplines, or meet with an academic coach to hone fundamental academic skills like time management or overcoming procrastination. Students can make an appointment for any of these services by visiting **go.tufts.edu/TutorFinder (https://go.tufts.edu/TutorFinder)**, or by visiting the website (**students.tufts.edu/staar-center (https://students.tufts.edu/staar-center)**).

Accommodations

Tufts University (and specifically, the instructors of this course) value the diversity of our students, staff, and faculty; recognizing the important contribution each student makes to our unique community. Tufts is committed to providing equal access and support to all qualified students through the provision of reasonable accommodations so that each student may fully participate in the Tufts experience. If you have a disability that requires reasonable accommodations, please contact the StAAR Center (formerly Student Accessibility Services) at **StaarCenter@tufts.edu (mailto:StaarCenter@tufts.edu)** or 617-627-4539 to make an appointment with an accessibility representative to determine appropriate accommodations. Please be aware that accommodations cannot be enacted retroactively, making timeliness a critical aspect for their provision. Please email your official accommodations letter by email to **Lawrence.Uricchio@tufts.edu (mailto:Lawrence.Uricchio@tufts.edu)** as soon as possible. Note that there are no exams in this course (timed or otherwise).

We are in the process of finalizing much of the course material. It is our goal to make all content accessible through a screen reader. If we have distributed inaccessible documents we are happy (and grateful) to have students notify us and we will rectify this as fast as possible.

Diversity Statement

This course will address questions that arise in the fields of evolutionary biology, human genetics, and medicine. Due to pervasive social influence of patriarchy and white supremacy, these fields have a long and troubled history with problematic individuals whose ideas have been applied (and misapplied) to justify prejudice. In this course, we will highlight the contributions of diverse scientists whose work is

reshaping our scientific and social understanding of these fields. We will discuss the toxic impacts of problematic individuals and how their social context may have shaped their scientific work. We will also highlight opportunities and initiatives seeking to make science more inclusive.

We acknowledge that some of this material is sensitive and may make the classroom more stressful for some students. We will seek to provide content warnings before class. This course is being taught for the first time. We are very open to student feedback in general, and especially in the context of classroom inclusivity. If we can make the classroom more inclusive for you or if you have suggestions on how to do so we would be very happy to talk with you.

If you prefer to speak with someone outside of the course about issues of equity and inclusion, the Center for STEM Diversity is an excellent resource.

Course Summary:

Date	Details	Due
Thu Sep 9, 2021	Welcome to comp bio (https://canvas.tufts.edu/calendar?event_id=225172&include_contexts=course_31670)	10:30am to 11:45am
Tue Sep 14, 2021	Cluster use and unix basics (https://canvas.tufts.edu/calendar?event_id=225158&include_contexts=course_31670)	10:30am to 11:45am
Thu Sep 16, 2021	Python basics (https://canvas.tufts.edu/calendar?event_id=225173&include_contexts=course_31670)	10:30am to 11:45am
Tue Sep 21, 2021	Identifying Genetic Patterns (https://canvas.tufts.edu/calendar?event_id=225159&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 1 (https://canvas.tufts.edu/courses/31670/assignments/194768)	due by 10:30am
Thu Sep 23, 2021	Functionals, conditionals, loops (https://canvas.tufts.edu/calendar?event_id=225174&include_contexts=course_31670)	10:30am to 11:45am
Fri Sep 24, 2021	Coding Assignment 1 (https://canvas.tufts.edu/courses/31670/assignments/188209)	due by 1pm

Date	Details	Due
Tue Sep 28, 2021	Identifying Genetic Patterns (https://canvas.tufts.edu/calendar?event_id=225160&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 2 (https://canvas.tufts.edu/courses/31670/assignments/194769)	due by 10:30am
Thu Sep 30, 2021	Lists and loops (https://canvas.tufts.edu/calendar?event_id=225175&include_contexts=course_31670)	10:30am to 11:45am
Tue Oct 5, 2021	Molecular evolution (https://canvas.tufts.edu/calendar?event_id=225186&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 3 (https://canvas.tufts.edu/courses/31670/assignments/194771)	due by 10:30am
Thu Oct 7, 2021	Software design, computational experiments (https://canvas.tufts.edu/calendar?event_id=225176&include_contexts=course_31670)	10:30am to 11:45am
Tue Oct 12, 2021	Genetic drift (https://canvas.tufts.edu/calendar?event_id=225162&include_contexts=course_31670)	10:30am to 11:45am
Wed Oct 13, 2021	Spotlight 4 (https://canvas.tufts.edu/courses/31670/assignments/194772)	due by 1pm
Thu Oct 14, 2021	Simulation and random numbers (https://canvas.tufts.edu/calendar?event_id=225177&include_contexts=course_31670)	10:30am to 11:45am
Fri Oct 15, 2021	Coding Assignment 2 (https://canvas.tufts.edu/courses/31670/assignments/194682)	due by 1pm
Tue Oct 19, 2021	Natural selection (https://canvas.tufts.edu/calendar?event_id=225163&include_contexts=course_31670)	10:30am to 11:45am

Date	Details	Due
	Spotlight 5 (https://canvas.tufts.edu/courses/31670/assignments/194773)	due by 10:30am
Thu Oct 21, 2021	Modeling natural selection (https://canvas.tufts.edu/calendar?event_id=225178&include_contexts=course_31670)	10:30am to 11:45am
Tue Oct 26, 2021	Natural selection (https://canvas.tufts.edu/calendar?event_id=225164&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 6 (https://canvas.tufts.edu/courses/31670/assignments/194774)	due by 10:30am
Thu Oct 28, 2021	Detecting selection (https://canvas.tufts.edu/calendar?event_id=225179&include_contexts=course_31670)	10:30am to 11:45am
Fri Oct 29, 2021	Coding Assignment 3 (https://canvas.tufts.edu/courses/31670/assignments/194683)	due by 1pm
Tue Nov 2, 2021	Ortholog detection (https://canvas.tufts.edu/calendar?event_id=225165&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 7 (https://canvas.tufts.edu/courses/31670/assignments/194775)	due by 10:30am
Thu Nov 4, 2021	Recursion, dictionaries (https://canvas.tufts.edu/calendar?event_id=225180&include_contexts=course_31670)	10:30am to 11:45am
Fri Nov 5, 2021	Coding Assignment 5 (https://canvas.tufts.edu/courses/31670/assignments/194719)	due by 1pm
Tue Nov 9, 2021	Functional genomics -- Jacob Dayton guest lecture (https://canvas.tufts.edu/calendar?event_id=225166&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 8 (https://canvas.tufts.edu/courses/31670/assignments/194776)	due by 10:30am

Date	Details	Due
Thu Nov 11, 2021	NO CLASS -- Veterans Day (https://canvas.tufts.edu/calendar?event_id=225181&include_contexts=course_31670)	1pm
Tue Nov 16, 2021	Homology, genome evolution (https://canvas.tufts.edu/calendar?event_id=225167&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 9 (https://canvas.tufts.edu/courses/31670/assignments/194777)	due by 10:30am
Thu Nov 18, 2021	Sequence alignment (https://canvas.tufts.edu/calendar?event_id=225182&include_contexts=course_31670)	10:30am to 11:45am
Fri Nov 19, 2021	Coding Assignment 6 (https://canvas.tufts.edu/courses/31670/assignments/194766)	due by 1pm
Tue Nov 23, 2021	Speciation, Biogeography and computing with Trees (https://canvas.tufts.edu/calendar?event_id=225168&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 10 (https://canvas.tufts.edu/courses/31670/assignments/194778)	due by 10:30am
Tue Nov 30, 2021	Inferring evolutionary relationships (https://canvas.tufts.edu/calendar?event_id=225169&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 11 (https://canvas.tufts.edu/courses/31670/assignments/194779)	due by 10:30am
Thu Dec 2, 2021	Tree inference algorithms (https://canvas.tufts.edu/calendar?event_id=225184&include_contexts=course_31670)	10:30am to 11:45am
Fri Dec 3, 2021	Coding Assignment 7 (https://canvas.tufts.edu/courses/31670/assignments/194722)	due by 1pm

Date	Details	Due
Tue Dec 7, 2021	Viral evolution (https://canvas.tufts.edu/calendar?event_id=225170&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 12 (https://canvas.tufts.edu/courses/31670/assignments/194780)	due by 10:30am
Thu Dec 9, 2021	Tree inference algorithms (https://canvas.tufts.edu/calendar?event_id=225185&include_contexts=course_31670)	10:30am to 11:45am
Fri Dec 10, 2021	Coding Assingment 8 (https://canvas.tufts.edu/courses/31670/assignments/194764)	due by 1pm
Tue Dec 14, 2021	Human genetics (https://canvas.tufts.edu/calendar?event_id=225171&include_contexts=course_31670)	10:30am to 11:45am
	Spotlight 13 (https://canvas.tufts.edu/courses/31670/assignments/194782)	due by 10:30am