APPHYS237/BIO251: Quantitative evolutionary dynamics and genomics

APPHYS237/BIO251: Quantitative evolutionary dynamics and genomics, Winter 2021
Course Meeting Schedule: TTh 10:30am – 11:50am (recordings will also be posted on Canvas)
Course Meeting Location: Zoom. (Link and password are available on Canvas or by request.)
Instructor: Prof. Benjamin Good, *Office:* Clark S231A, *Email:* bhgood@stanford.edu
Office Hours: Th 12:30pm-2pm or by appointment.
Course Website: https://bgoodlab.github.io/courses/apphys237/

Course description:
The genomics revolution has fueled a renewed push to model evolutionary processes in quantitative terms. This course will provide an introduction to quantitative evolutionary modeling through the lens of statistical physics. Topics will range from the foundations of theoretical population genetics to experimental evolution of laboratory microbes. Course work will involve a mixture of pencil-and-paper math, writing basic computer simulations, and downloading and manipulating DNA sequence data from published datasets. This course is intended for upper level physics and math students with no biology background, as well as biology students who are comfortable with differential equations and probability.

Learning goals:
At the end of the course, students will be able to:
1. Formulate canonical models of population genetics in mathematical language, and implement them as computer simulations.
2. Explain basic techniques of experimental evolution & DNA sequencing, and show how to connect these measurements with the population genetic models above.
3. Describe canonical theoretical approaches & approximations for characterizing the behavior of these models mathematically, and use self-consistency arguments & order-of-magnitude estimation to show when these approximations break down.
4. Use branching processes and diffusion theory to predict fixation probabilities and genetic diversity at a single genetic locus, and apply these results in simple multi-locus settings.
5. Reproduce the analysis in three recent studies that use a combination of laboratory experiments, DNA sequencing, & mathematical modeling to measure mutation rates, fitness effects of new mutations, recombination, & epistasis.

Prerequisites:
Biology: None.
Math: Multivariable calculus and a basic familiarity with differential equations & probability. Other mathematical tools will be introduced as needed. Note that this course will be taught at the level of a 200-level course in physics or applied math, so you should be comfortable with mathematical reasoning.
Programming: Basic familiarity with a scripting language (e.g. Python, Matlab, R, etc.) and associated packages for plotting graphs. There are a variety of good tutorials available online (some examples are listed in the Course Materials).

If you are unsure of your background, please contact the instructor to discuss whether the course would be a good fit for you.

Lectures: TTh 10:30am – 11:50am via Zoom. Meeting info and lecture recordings are available on Canvas. Lecture notes will be posted on the course website.

Slack: A dedicated Slack workspace (stanford-1s5y.slack.com) has been created for our course. Feel free to use the **#general** channel to ask (or answer) questions about the course material. We will also need one student volunteer to create a private channel for all the students that can be used for homework collaboration questions, etc., without the instructor present.

Problem Sets: The course will have ***4 problem sets***, due every other Tuesday at the beginning of class. A new problem set will be posted on the course website after the old one is due. You are allowed (and encouraged!) to work with others on the problems, but everyone must turn in their own writeup. Please turn in your homework by emailing a single PDF of your writeup to apphys237@gmail.com using a subject line "Homework 1", "Homework 2", etc.

*Note: if you write solutions on paper, apps like **JotNot** (https://www.jotnot.com/scanner.html) can be used to scan your solutions using a phone or tablet camera and convert them into a single PDF. Programs like **Preview** (Mac) or **PDF Merger and Splitter** (Windows) are also useful for merging separate PDFs or images into a single PDF.*

Problem sets will be graded primarily for effort, so you are encouraged to attempt as many problems as you can and explain why you got stuck (or why your answer doesn't make sense to you if you do not think it is correct). To increase opportunities for continued learning, students may submit ***problem set corrections*** to earn back points lost during the initial round of grading. The maximum number of points recovered will depend on your initial effort: a solid initial effort can be corrected for up to 100% of the original points, while un-attempted or partially attempted problems will be eligible for a smaller amount. Solutions will be posted on Canvas and Slack after each problem set is graded and may be used to help with your corrections. All corrections must be submitted by the last day of classes, and should be submitted using the same procedure above (using the subject line "Corrections for Homework 1", etc.)

Participation: This course is still in active development, so your participation and feedback will be critical for improving the material for future generations of students. It will also allow the instructor to better tailor the material to the specific needs of this year's class (some of whom will be attending lectures asynchronously). We will formalize this process through ***weekly feedback assignments*** which will count toward your final grade. By Monday morning of each week, please send a short email to apphys237@gmail.com, using the subject heading "Week 1 Comments", "Week 2 Comments", etc. These comments could include anything from questions about the previous week's material, portions of the lectures or problem sets that you thought were particularly confusing (or particularly interesting), background knowledge you felt you were missing, comments about pacing, etc. Questions about next week's material are also encouraged if you prefer to read the lecture notes before class. Your email can be brief, but

please try to aim for about 2-3 comments per week. Don't worry if you feel like your comments aren't particularly insightful – everyone's feedback is extremely valuable in an interdisciplinary course like this one.

Final Project: Due to the shortened quarter, we will have to take a different approach for our final project this year. We would ordinarily have you design a research project that is connected to the topics we covered in the course. This year, we will try to achieve some of the same goals by having you suggest a new homework problem based on a recent theoretical or experimental paper. A central goal of this course is to illustrate how quantitative evolutionary reasoning can be applied to "hands on" examples from the recent literature. The goal of the final project is to get you to think through this same process using the skills you will have developed by the end of the quarter.

You should pick an experimental or theoretical paper from the field (broadly defined), and identify at least one aspect of the data analysis (e.g. one figure panel) or one part of the mathematical calculation that you think would be interesting to turn into a homework problem in future iterations of the course. You do not need to formulate or carry out the actual problem, but you should explain roughly what you think a hypothetical student would do, and why you think this would be an interesting problem. Does it illustrate particular concepts that we did (or didn't) cover in the course? Does it highlight an interesting application that you were previously curious about and now have the tools to understand? You should also note any background knowledge that you think the students will need to carry out the problem.

The final product should be a ~1 page (e.g. 3-4 paragraph) writeup to be submitted on the last day of the course. The first paragraph should give a brief summary of the paper you have chosen, and the remaining paragraphs should focus on the portion(s) you wish to turn into a homework problem. Please submit your writeups by emailing a PDF to apphys237@gmail.com with a subject line "Final Project".

There are several potential strategies for finding a good paper for the final project:

(1) Choose one of the papers referenced in the course notes, and look through some of the papers it cites in its bibliography.

(2) Choose one of the papers referenced in the course notes, and look through some of the papers that cite it. (You can do this relatively easily on google scholar: look up the paper and click the "cited by" link listed below it.)

(3) You can often find good papers by scanning recent table of contents from Nature, Science, PNAS, PLoS Biology, Elife, Nature Eco Evo, Genetics, etc. (Note: papers from physics journals like PRL, PRE, etc. are sometimes ok, particularly if they are cited by other biology papers, but I would not recommend starting from their table of contents.)

(4) You can also find good papers by scanning the recent table of contents for preprint servers like Biorxiv (https://www.biorxiv.org/collection/evolutionary-biology) or arXiv (https://arxiv.org/list/q-bio.PE/recent), though the signal to noise ratio can be a little higher there.

Please contact the instructor if you have additional questions about the scope or format of the final project, or if you need more help identifying relevant papers.

Grading: Problem sets 60%, Participation 25%, Final Project 15%.

Tentative list of topics:

1. Experimental evolution and DNA sequencing. Structure of genomes and genetic diversity. Chemostats and serial dilution experiments: measuring fitness differences, mutation rates, and genetic drift. Next-generation sequencing, genome assembly, and alignment; PCR and amplicon sequencing; metagenomics. Genetic signatures of natural selection, parallelism, and entrenchment. Basics of genome editing: transformation, TnSeq, DNA barcoding, CRISPR.

2. Population genetics at a single locus. Microscopic models: mutation, selection, and genetic drift. Mesoscopic models and the diffusion approximation: Fokker-Plank, Langevin, and path integral formulations; universality classes and effective parameters. Non-equilibrium dynamics: fixation probabilities and genetic diversity. Deterministic approximations and branching processes. Heuristic approaches and asymptotic analysis.

3. Population genetics at multiple loci. Genetic linkage, recombination, and epistasis. Two-locus models and the infinite sites limit. Free recombination and quasi-linkage equilibrium. Selective sweeps, clonal interference, and background selection. Backward in time approaches and coalescent theory. Fitness valley crossing and modifiers of evolvability.

Course materials:

Lecture Notes: Lecture notes from the previous iteration of the course are available on the course website. Updated notes will be posted as time allows.

Textbooks: There is no required textbook for the course, and unfortunately there are currently no books that give a good comprehensive coverage of the subject. The following books are useful references for various parts of the course and will be on reserve at the library:

Warren Ewens, *Mathematical Population Genetics*.

Rick Durrett, *Probability models for DNA sequence evolution*.

Charlesworth and Charlesworth, *Elements of Evolutionary Genetics*.

Philipps and Milo. *Cell Biology by the Numbers.* (http://book.bionumbers.org/)

Other useful resources (available on the course website):
1. BH Good (2016), Molecular evolution in rapidly evolving populations, Chapter 1 (PDF)
2. SF Levy, JR Blundell, et al (Nature 2015), Quantitative evolutionary dynamics using high-resolution lineage tracking, Supplementary Information (PDF)
3. DS Fisher (Les Houches Course 11, 2007), Evolutionary dynamics (PDF)
4. Korolev et al (Rev Mod Phys, 2010), Genetic demixing and evolution in linear stepping stone models (PDF)
5. Neher and Shraiman (Rev Mod Phys, 2011), Statistical genetics and evolution of quantitative traits (PDF)

Programming tutorials:
Getting Started with Python for Science (https://scipy-lectures.org/intro/)