APPHYS237/BIO251: Quantitative evolutionary dynamics and genomics

APPHYS237/BIO251: Quantitative evolutionary dynamics and genomics, Spring 2023
Course Meeting Schedule: TTh 10:30am – 11:50am
Course Meeting Location: Wallenberg Hall (Bldg 160) Rm 127
Instructor: Prof. Benjamin Good, *Office:* Clark S244, *Email:* bhgood@stanford.edu
Office Hours: Tu 1pm-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 in Wallenberg Hall (Bldg 160) Rm 127. Lecture notes will be posted on the course website.

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.

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 (**Weds June 7**).

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 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: The goal of the final project is to let you practice designing a research project connected to the topics we have covered in the course. The final product will be a written report that will be due at the end of the quarter **(Weds June 14 at 11:59 p.m)**. You may choose from two main tracks:

Theoretical Track*:* Pick a recent (<20yo) experimental paper from the field. Propose and implement a simulation model that can qualitatively reproduce at least one key finding in the paper (e.g. one of the main figures). Provide order of magnitude estimates for the input

parameters if known, or indicate when they are not known. Choose one limiting parameter regime of the simulation model where you think you can describe some aspect of its behavior analytically.

Experimental Track: Pick a theoretical paper from the field. Propose an experiment to either test one of the predictions of the model, or measure one of the unknown input parameters. Include schematic figures and anticipated figures for what the data will look like. Provide a rough outline of the anticipated materials and personnel, in order to demonstrate that the project can be completed in <10yrs and <$2m.

Detailed instructions for each track of the final project will be posted on the course website.

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*.

Walsh and Lynch, *Evolution and Selection of Quantitative Traits.*

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/)