APHYS 237/BIO251: Quantitative evolutionary dynamics and genomics

APPHYS 237/BIO251: Quantitative evolutionary dynamics and genomics, Winter 2020
Course Meeting Schedule: TTh 10:30am – 11:50am
Course Meeting Location: Clark Center S361
Instructor: Prof. Benjamin Good, *Office:* Clark S231A, *Email:* bhgood@stanford.edu
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 Clark Center S361. Lecture notes will be posted on the course website after class. *Note: there will be no lecture on Jan 28, March 3, or March 5.*

Problem Sets: Biweekly 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 may work with others on the problems, but everyone must turn in their own written copy.)

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. 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 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 budget, in order to demonstrate that the project can be completed in <10yrs and <$2m.

To help you choose an appropriate topic, a 1-2 paragraph "preproposal" will be due on March 24th.

Detailed instructions for each track (as well as the preproposal) are available on the course website.

Grading: Problem sets 75%, Final Project 25%.

<u>Tentative</u> 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:
Textbooks: There is no required textbook for the course, and unfortunately there are currently no books that give 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:

John Gillespie, Introduction to population genetics.

Warren Ewens, Mathematical Population Genetics.

John Wakeley, Coalescent Theory.

Suggestions about which sections of these books are relevant background for a given lecture will be posted on the course website.

Other useful resources:
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/)