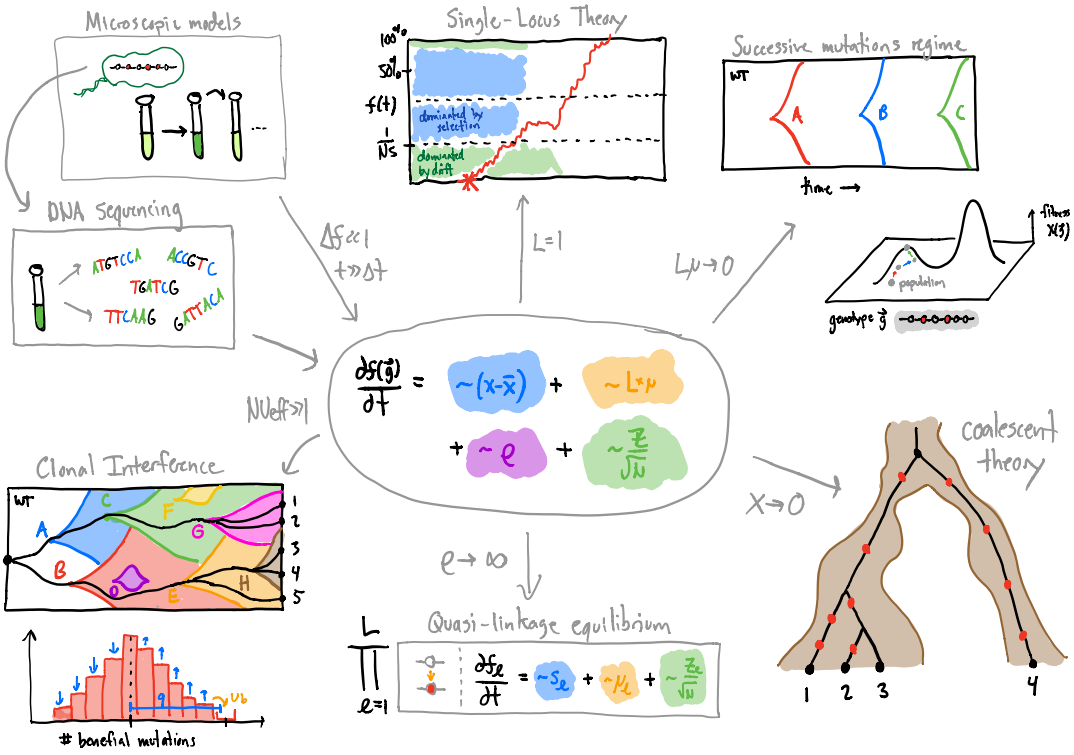
Course notes for APPHYS237 / BIO251:

Quantitative Evolutionary Dynamics & Genomics



Benjamin H Good
Department of Applied Physics
Stanford University

Spring 2023

Contents

Chapter 1

Introduction

1.1 Preface

The goal of this course is to provide an introduction to quantitative evolutionary modeling through the lens of statistical physics. Why is such a course necessary, and why should you take it?

At its core, physics is the quantitative study of how matter and energy change over time. In the living world, many of these changes are driven by Darwinian evolution, which acts on populations of organisms and the information encoded in their genomes. The study of this process — often known as ***evolutionary dynamics*** or ***population genetics*** — has become one of the fastest growing subfields of biophysics, which is itself one of the fastest growing areas of physics[1]. Technological advances in our ability to read and write genomes are fueling a lot of exciting progress in this area, in which interactions between quantitative theory and experimental data are playing an important role. Physicists and engineers are uniquely poised to contribute at this interface, given their extensive training in both theoretical and applied problems.

Unfortunately, it can be hard to find a dedicated set of courses where one

[1] See the recent report, *Physics of Life*, from the National Academy of Sciences, `https://nap.nationalacademies.org/resource/26403/interactive/`

can learn this material, despite the fact that it's now a relatively established subfield. This is particularly true for evolutionary biology and population genetics, where the underlying mathematical models are sufficiently complicated that they are rarely covered – even in graduate-level courses – in the traditional biology curriculum. There are lots of great courses in population genetics that are now available [2], but they tend to be geared toward "consumers" of population genetic methods, and assume that students have little familiarity with the mathematical tools (e.g. PDEs, series expansions, probability distributions) that are a core part of the undergraduate physics curriculum. As a result, students are often left to comb through the primary literature, which can be quite challenging given the long history of the field.

The this course is an attempt to fill this gap. It aims to provide a mathematically rigorous but biologically naive introduction to the field of evolutionary dynamics and genomics. It is targeted both to physicists and engineers who are curious about evolution, and want to get up to speed on modern theoretical and experimental approaches, as well as biologists who might want a deeper understanding of the theoretical tools we can use to model evolution mathematically. The course covers topics ranging from the foundations of theoretical population genetics to experimental evolution in laboratory microbes, while emphasizing techniques like order-of-magnitude estimation and the method of successive approximations. For physics students, it might also provide a first exposure to non-equilibrium approaches in statistical physics (e.g. stochastic differential equations and continuous-time branching processes) which have widespread applications beyond this course. As we will see throughout the course, evolutionary phenomena will turn out to provide a fantastic setting in which to explore many of these ideas.

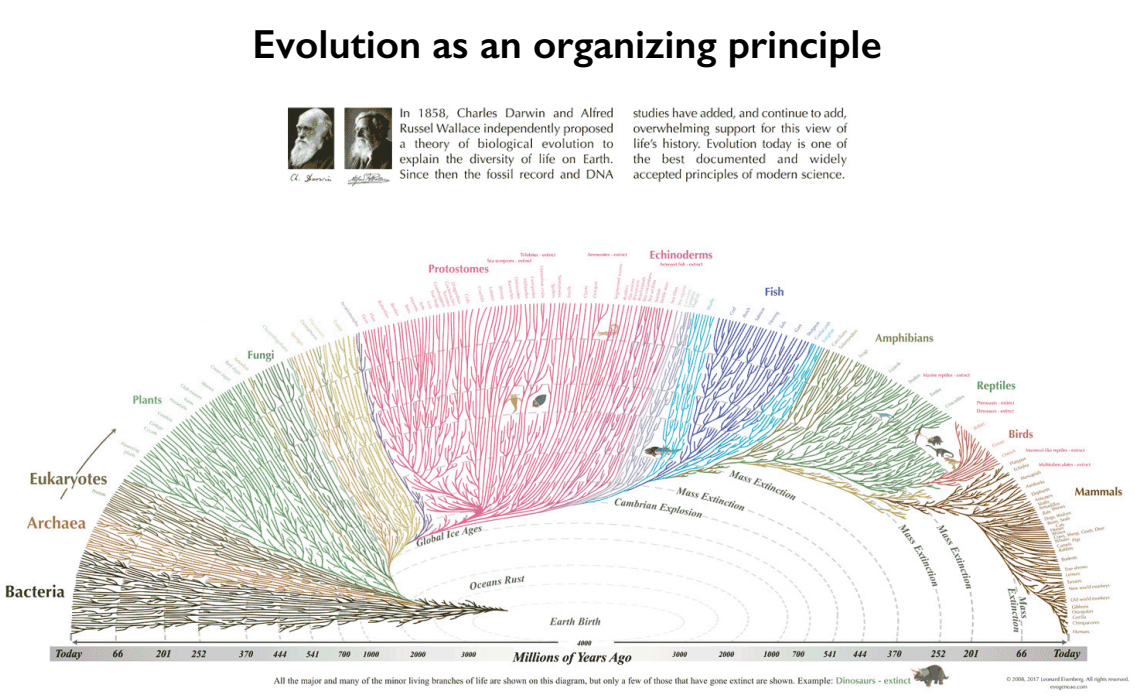
In the remainder of this section, we will provide a brief overview of what we mean by "quantitative evolutionary dynamics", and start to introduce some of the key questions that we will be interested in during the course. The next

[2] One of my favorites is Graham Coop's "Population and Quantitative Genetics" course, which is available online: `https://github.com/cooplab/popgen-notes/releases`.

two chapters will quickly review some of the mathematical and biological background, and then we'll start with or first model of evolution in Chapter 4.

1.2 Evolution as a statistical mechnical process

What do we even mean by the phrase "quantitative evolutionary dynamics"? Traditionally, I think a lot of us are used to thinking about evolution as a historical process – that is, the story of how life came to be the way it is today.



In this view, a major goal is to figure out what these historical relationships are, what happened at the major transitions, and so on.

We're also probably used to thinking about evolution as the world's best optimization scheme, which is able to generate some exquisitely fine-tuned biological structures when compounded over millions and billions of years. Here is one of my favorite examples that you might have heard about from *Planet Earth*:

Evolution can produce exquisitely fine-tuned structures over long (geological) timescales

This is a picture of a fungus named Cordyceps, which infects a particular species of ant, and manages to control the ant's behavior by taking over its brain.[3] It makes the ant climb onto a leaf that is 25 ± 2cm off the ground, and then a fruiting body bursts out of the ant's head, in order to rain down spores onto the generation of ants.

This behavior is really fine-tuned: if the leaf is a bit higher up or a bit farther down, then the temperature and humidity are such the spores don't grow as well. This particular species of fungus also has difficulty growing in closely related species of ants. Evolution provides us with a story for how this fine-tuned behavior could arise — something we might call the ***"Jurassic Park" Theory of Evolution***: life just finds a way.

At the same time, this process seems to be entirely constrained by the precise biological mechanisms that allow this sort of mind control to occur, and the

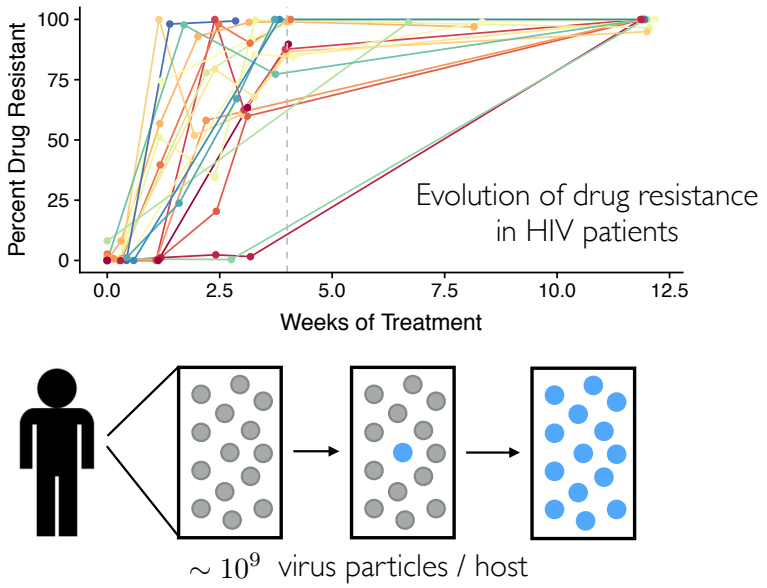
[3] Ed Yong wrote a nice popular science article about this species for *The Atlantic*: `https://www.theatlantic.com/science/archive/2017/11/how-the-zombie-fungus-takes-over-ants-bodies-to-control-their-minds/545864/`

chance events that allowed it to happen for this particular pair of species, and not others. It doesn't seem like physics would be particularly helpful for predicting this sort of behavior.

However, if we turn our attention to microbial organisms, we'll notice that not all of evolution involves these micraculous innovations that take place over geological time. Instead, there are many smaller-scale examples of evolution that take place on human-relevant timescales — some of which have important practical consequences that we might want to predict or control.

Here is just one real-world example, showing the evolution of drug resistance in a cohort of HIV patients during a clinical trial.

Evolution can also occur on *human-relevant* timescales in fast growing microbial populations



Feder et al (*PLoS Genetics*, 2021)

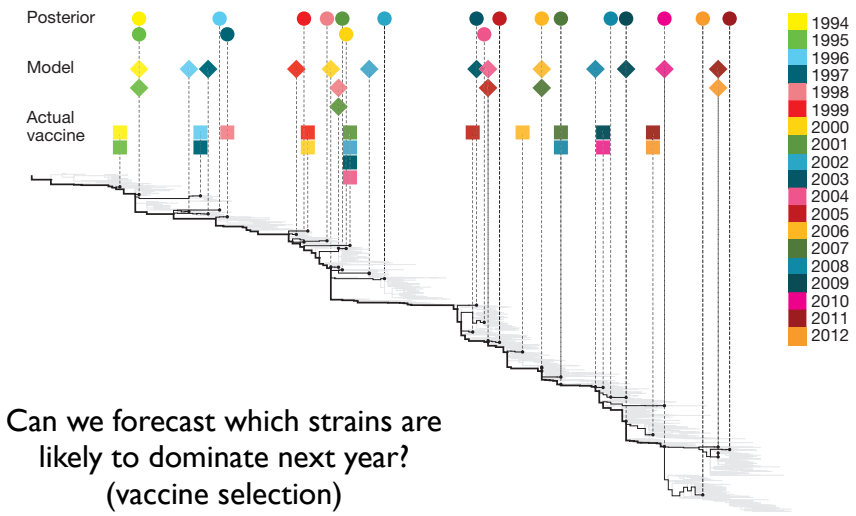
Each one of the colored lines represents a different patient. So we can see that all 10 or so acquired resistance to this particular drug within about 12 weeks. The

initial HIV strains were not resistant to begin with, so this means that in each host, one of the billion or so viral particles acquired a random mutation that allowed it to evade the drug, and this its descendants to rapidly take over the population.

At some level, this process is just as random as the zombie ant example above. It still relies on a random mutation occurring in a single random individual, which just happens to provide resistance to this particular drug. In this case, however, we can see that these random events lead to much more repeatable behavior at the population level — enough that we can start asking some *quantitative* questions: For example, is there something special about the two patients at the bottom that caused them to acquire drug resistance anomalously late? Or is this just the typical variation we'd expect in a random ensemble of 10 patients?

There are many other examples like this. This is a genealogical tree showing the worldwide evolution of the influenza virus over the last 30 years, as it evolves to evade the collective effects of all of our immune systems.

Example: antigenic evolution of the global influenza pop'n



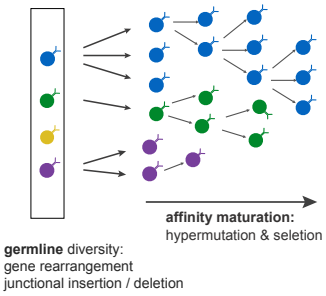
Luksza and Lassig (*Nature*, 2014)

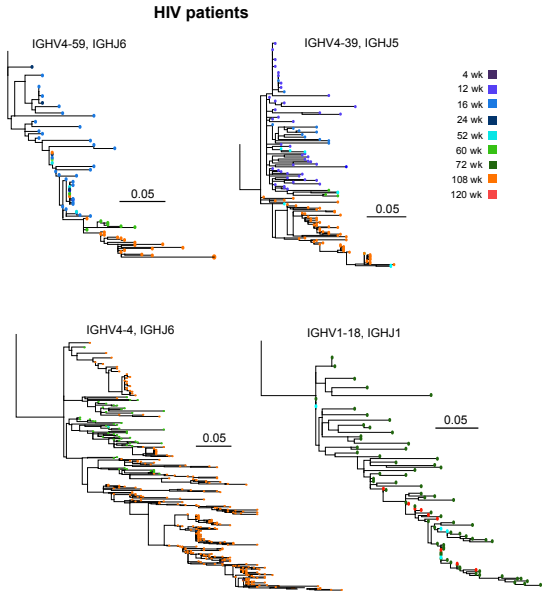
We could construct an analogous (and much bigger tree) for SARS-CoV2 over the last few years[4]. A big challenge is that every year, we have to choose one or two of these strains to use to serve as the basis for that season's influenza vaccine. So something we might want to know — and which researchers are actively trying to do right now — is to determine whether we can use real-time genome sequencing to forecast which strains are likely to dominate in the next flu season, and to use this information to inform vaccine selection.

Here's another example, showing the evolutionary processes that occur in our immune cells as they respond to antigenic pressure.

[4]You can play around with these trees yourself using the NextStrain tool: `https://nextstrain.org/ncov/gisaid/global/6m`

Example: somatic evolution of immune repertoires





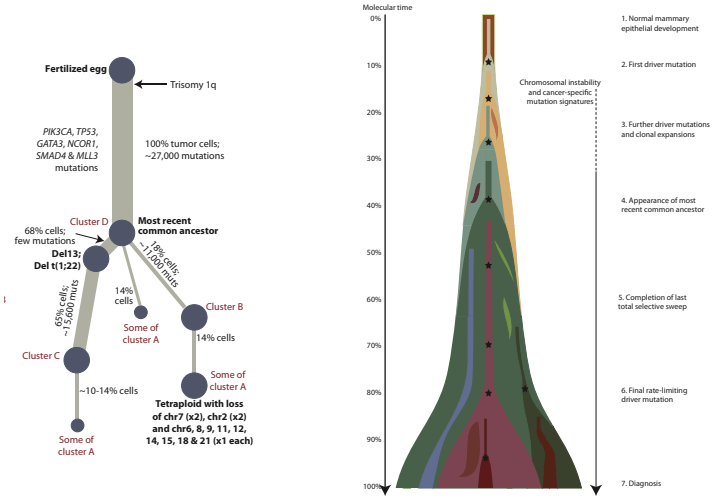
Can we guide the evolution of specific antibodies with the right vaccination strategy?

Nourmohammad et al (*MBE*, 2019)

In this case, we might want to figure out whether we can guide our immune system to evolve a particular desired antibody over another, by designing the right vaccination cocktail.

Cancer is another example. This is fundamentally an evolutionary disease, in which some of our cells acquire a sequence of mutations that allow them to proliferate out of control.

Example: somatic evolution of cancer tumors



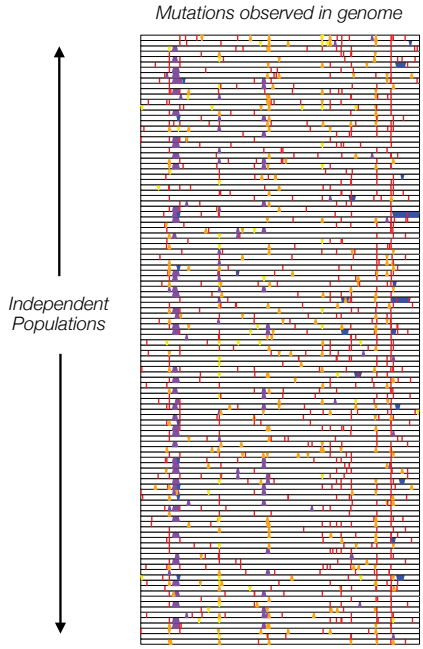
- How long does it take for cancer to emerge? 1 yr? 1000yrs?
- How rapidly do tumors acquire resistance to treatment?

Nik-Zainal et al (*Cell*, 2012)

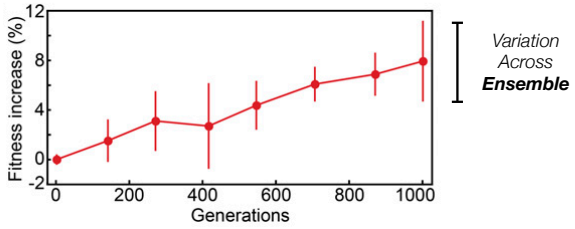
In this case, we might want to know things like how long we expect it to take for a particular cancer to emerge in a given individual. From the organism's perspective, it makes a big difference whether this process takes <10 years or 100's to 1000's of years. This illustrates that even order-of-magnitude predictions could be extremely useful.

Finally, there are a lot of interesting examples in the field of ***experimental evolution***, in which large numbers of independent populations can be evolved in parallel in controlled laboratory conditions.

Example: high-throughput evolution in the laboratory



Tenaillon et al (*Science*, 2012)

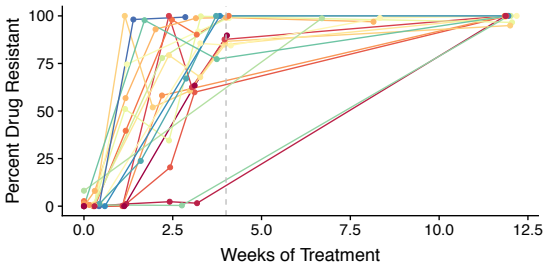


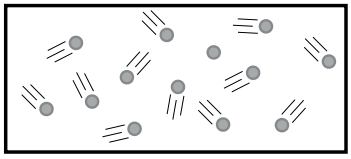
Lang et al (*Genetics*, 2011)

These experiments offer an opportunity to move beyond merely speculating about what might happen if we replayed the tape of evolution, and instead start to map out the entire statistical ensemble of outcomes in a given environment.

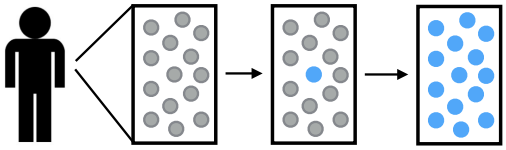
To make progress on these questions, it's clear that we'll have to move beyond thinking about evolution as a historical process or a perfect optimization machine, and instead start thinking about ***evolution as an algorithm***, or a ***statistical mechanical process***.

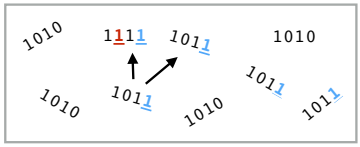
Evolution as a statistical mechanical process





$$PV = nRT$$





Goal: understand the ***mathematical models*** and ***experimental data*** that help us think about this process in a quantitative way

The catch is that this will require us to move beyond the statistical mechanics of billiard balls that we're used to thinking about in physics. Instead, we'll have to deal with the statistical mechanics of noisy self-replicating bit-strings. The goal of this course is to introduce the mathematical models and experimental data that help us think about this process in a quantitative way.