Molecular evolution in rapidly evolving populations

A DISSERTATION PRESENTED
BY
BENJAMIN HARMAR GOOD
TO
THE DEPARTMENT OF PHYSICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
PHYSICS

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS

MAY 2016

Thesis advisor: Professor Michael M. Desai Benjamin Harmar Good

Molecular evolution in rapidly evolving populations

Abstract

Advances in DNA sequencing are creating new opportunities for studying the process of evolution. These measurements can be particularly useful for rapidly evolving microbial organisms, whose small size and fast generation times make them ideal for controlled laboratory experiments and for tracking replicate populations in vivo. However, the interpretation of this new source of data is complicated by the unique ways in which these large microbial populations evolve.

The basic problem is that natural selection is forced to do too many things at once. Unlike the classical picture, where new mutations arise one-by-one, rapidly evolving populations often harbor many selected variants at the same time. When recombination is limited, selection cannot act on these mutations individually, but only on combinations of mutations that happen to arise on the same genetic background. These effects, known as *clonal interference*, create correlations along the genome that are difficult to disentangle. Existing population genetic models often neglect these effects, which leaves us at loss when interpreting data from these populations.

In Chapters 2-5, we analyze the effects of clonal interference in a simple "null model" of microbial evolution. We focus on the simplest model that is consistent with two empirical observations: (1) many fitness-influencing mutations are created every generation and (2) mutations have a broad range of fitness effects. After analyzing the basic dynamics of this model, we obtain predictions for the substitution rates of individual mutations and the patterns of linked neutral diversity, and we show how these quantities depend on the population size, mutation and recombination rates, and the fitness effects of new mutations.

In Chapters 6 and 7, we apply this null model to data from laboratory experiments in *S. cerevisiae* and *E. coli*. We develop a statistical framework to infer the underlying parameters (the fitness effects of new mutations), which allows us to quantify deviations from the model over longer evolutionary timescales.

Finally, in Chapters 8 and 9, we investigate the behavior of the model when some of the parameters are allowed to evolve or change in time.

Citations to previously published work

Chapter 2 appears in its entirety as

> "Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution," B. H. Good and M. M. Desai, *Theoretical Population Biology*, 85: 86-102 (2013).

Chapter 3 appears in its entirety as

> "Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations, " B. H. Good, I. M. Rouzine, D. J. Balick, O. Hallatschek, and M. M. Desai, *Proceedings of the National Academy of Sciences*, 109: 4950-4955 (2012).

Chapter 4 appears in its entirety as

> "Deleterious passengers in adapting populations, " B. H. Good and M. M. Desai, *Genetics*, 198: 1183-1208 (2014).

Chapter 5 appears in its entirety as

> "Genetic diversity in the interference selection limit," B. H. Good, A. M. Walczak, R. A. Neher, and M. M. Desai, *PLoS Genetics*, 10: e1004222 (2014).

Chapter 6 appears in its entirety as

> "The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations," E. M. Frenkel, B. H. Good, and M. M. Desai, *Genetics*, 196: 1217-1226 (2014).

Chapter 7 appears in its entirety as

> "The impact of macroscopic epistasis on long-term evolutionary dynamics," B. H. Good and M. M. Desai, *Genetics*, 199: 177-190 (2015).

Chapter 8 appears in its entirety as

> "The evolutionarily stable distribution of fitness effects," D. P. Rice, B. H. Good and M. M. Desai, *Genetics*, 200: 321-329 (2015).

Chapter 9 appears in its entirety as

> "Fate of a mutation in a fluctuating environment," I. Cvijović*, B. H. Good*, E. R. Jerison, and M. M. Desai, *Proceedings of the National Academy of Sciences*, 112: E5021-E5028 (2015).

1 Introduction

Evolution is a fundamental organizing principle in biology. It allows us to reconcile the diversity of the living world with the exquisite level of design and purpose that many organisms seem to possess. And it explains how such improbable levels of design could emerge from an otherwise random struggle for existence. Physicists have long been intrigued by evolution, perhaps because they see in it a semblance of their own great organizing principle, the theory of statistical mechanics. The two theories do share some common features. Both are products of the mid-19th century; both are concerned with populations (either of molecules or organisms); and both attempt to explain the collective behavior of these populations by appealing to a simple variational principle, like the minimization of free energy or the maximization of fitness. Both theories have been extremely successful in this endeavor, providing a conceptual framework and vocabulary for reasoning about the world around us. However, the modern versions of these theories can also seem quite different. In particular, while statistical mechanics eventually developed an applied, predictive component, our ability to predict the course of evolution still remains limited.

This does not mean that evolutionary biologists have been idle during this time. On the contrary, biology has undergone a revolution of its own in the 20th century, and has made great strides in understanding the molecular machinery that controls each living organism [160]. We now know the precise mechanisms governing the storage and transmission of genetic information, and sequencing and gene-editing techniques make it possible to read and rewrite this information at will [2]. Thanks in part to these advances, and to the work of many naturalists and geneticists, we have a much better picture of what *can* happen during evolution, and in many specific cases, what *did* happen. Nevertheless, it is still difficult to translate

this knowledge into quantitative predictions about the future. Even in the simplest microevolutionary settings, we usually have much less of an idea of what *will* happen, how long it will take, or what else could have happened if we "replayed the tape" from the beginning. To answer these questions, it is necessary to understand evolution not just as an organizing principle, but as a dynamical process that specifies how populations change with time.

These sentiments are, of course, not new. The search for a predictive theory of evolution is almost as old as the theory itself. So before we commit the cardinal sin of a physicist [241], it is useful to pause and ask: what has made this task so difficult in the past? and what has changed that might give us hope for the future?

For many years, our limited predictive ability could be traced to a shortage of empirical data. Most evolutionary changes are too subtle, or occur much too slowly, to be observed on human-relevant timescales. Instead, we have had to rely on morphological observations from the fossil record, or the inferred relationships between present-day species. This situation has started to improve in recent years, as advances in sequencing technology have enabled precision measurements of the evolutionary process that would have been unimaginable just a few decades ago. Instead of relying on morphological features, which may be poorly preserved in the fossil record or difficult to compare across species, it is now possible to directly observe the genes and mutations that evolution acts on. Low-cost sequencing is a particularly powerful tool for microbial organisms like viruses, bacteria, and populations of somatic cells, whose small size and rapid generation times make them ideal for controlled laboratory experiments [11], and tracking replicate populations in vivo [260, 13, 213, 307, 364]. These populations may have little direct relevance for traditional evolutionary questions such as the origin of flight [209] or the evolution of the eye [196]. However, they can be much more useful for learning about the evolutionary process itself, since they can provide the raw material from which predictive, phenomenological theories can be constructed and tested.

Yet even with the rosiest experimental outlook, there are fundamental theoretical reasons why evolution can be difficult to predict. One reason is that evolution is inherently random. Mutation and recombination — the sources of genetic variation — are ultimately caused by nanoscale cellular machinery acting on a single pair of DNA molecules. These processes are strongly influenced by thermal fluctuations, so the timing and location of these events are essentially random. In addition, biologists have long supposed that the reproductive success of an individual is also influenced by chance (whether of a molecular nature or not), and this leads to an additional source of noise known as *genetic drift*. Given this inherent stochasticity, it is impossible to predict when or whether any specific event will occur. Instead, we can only speak about the relative probabilities of these events.

A second major difficulty is that evolution is inherently messy, i.e., it cannot be divorced from biology. A genetic variant may start as a simple change in a DNA molecule, but its phenotypic effect is a product of

development and physiology, and its eventual reproductive success depends on the interactions between the phenotype and the ecology of the organism. These are entire subfields of biology, which, even in the best studied organisms, are only partially understood. A list of all the factors that *could* influence evolution would quickly overwhelm an army of experimentalists. To make matters worse, many of these factors are not fixed constants. Evolution can alter the values of these parameters, and even the makeup of the list itself.

Yet despite these fundamental difficulties, there are still some reasons for optimism, provided that we take a more nuanced view of what we are trying to predict. In many evolutionary scenarios, we are most interested in predicting the order of magnitude of a quantity, rather than its precise value. E.g., will a population of bacteria acquire resistance to a particular antibiotic in a matter of months, decades, or millenia? Fortunately, there are many examples where orders of magnitude can be predictable even when the underlying events are highly random.

For example, consider the textbook example of randomness in physics: the decay of radioactive isotopes. There are few physical constraints on the timescales of this process. Half-lives of different radioactive isotopes vary over 55 orders of magnitude [5], and even when the half-life of an isotype is known, it is still impossible to predict when a particular atom will decay. Nevertheless, a pair of Carbon-11 and Carbon-14 atoms are not completely indistinguishable from each other. Although we cannot predict whether the Carbon-11 atom will decay in 10 minutes or 30 minutes, we are fairly confident that it will decay within 200 minutes, and even more certain that it will decay within 2000 minutes. Similarly, we are quite certain that the Carbon-14 atom will not decay in fewer than 10^8 minutes, so that it will essentially always be the second of the two to decay. And perhaps most importantly, while we cannot calculate half-lives from first principles, there is an established body of nuclear theory that relates these decay times to other physical quantities (e.g., masses of products, density of states, etc.), which can be measured and used for predictions [31].

Radioactive decay is one particularly well-studied example, but similar order-of-magnitude questions abound in evolutionary biology. Some are grand in scope, e.g. why did multicellularity require $\sim 10^9$ years to evolve, rather than $\sim 10^7$ or $\sim 10^{11}$? Others are potentially more answerable, like the timescale of antibiotic resistance above. But even some of the most basic evolutionary timescales, which arise in the context of laboratory experiments, are still surprisingly difficult to predict.

For example, experiments show that a particular strain of yeast acquires ~ 10 mutations after ~ 1000 generations of evolution in rich media [199]. This sounds like a reasonable number. But why is ~ 10 mutations the correct scale rather than ~ 100 or some number $\ll 1$? Would we be more or less surprised to learn that the same 10 mutations were shared across the population, rather than each cell having its own set? How many more mutations would we expect if the evolution was carried out in a bioreactor, with a

million-fold increase in the number of cells? Or in an environment that fluctuates between high and low temperatures? And most importantly, what biological parameters would we have to measure in order to predict these outcomes ahead of time?

It seems unlikely that we could predict the emergence of antibiotic resistance (let alone multicellularity) if we cannot also answer such basic questions in the laboratory. Borrowing hope from physics, we might expect that these order-of-magnitude questions will depend on a few key biological parameters, which, like the masses of the decay products above, can be measured directly if necessary. But at present, it is not always clear *which* combinations of parameters are relevant for such questions, which makes it difficult to know what to measure. This is as true for the timescales of laboratory evolution as it is for antibiotic resistance in the wild. In both cases, determining the relevant "knobs" will require a careful mix of theoretical exploration and experimental calibration. This is one of the main goals of the present work.

Most importantly, we will not try to predict evolution from first principles, or even from higher-level cellular processes like metabolism and gene regulation. Instead, we will take a phenomenological approach and assume a perfect knowledge of biology, such that the range of possible outcomes, their mutation rates, and their phenotypic consequences, are specified a priori. Our goal then is to predict how evolution acts on this spectrum of possibilities to produce the outcomes that we measure experimentally, particularly those outcomes that can be measured in sampled DNA sequences.

By abstracting away the biology, what remains is primarily a question about dynamics — the "bean-bag genetics" that describe how a mutation comes to dominate a population. Theoretical techniques from physics will naturally be useful for attacking such questions. Perhaps surprisingly, one of the main findings of this and other work is that these dynamical considerations can be just as important as the biology for determining the fates of mutations, particularly in the rapidly evolving microbial populations that are our focus here. Ideally, a better understanding of the dynamics would allow us to turn our predictions around, so that we could utilize patterns of DNA sequence evolution to make inferences about biological parameters that might be difficult to observe directly.

We are not the first to try to model the dynamics of mutations within a population. Similar questions have been studied for more than 100 years in the field of *population genetics*. One of the early triumphs of this field was the *modern evolutionary synthesis*, which showed how the sweeping macroevolutionary changes of Darwin's theory could be reconciled with microevolutionary selection and Mendelian genetics [286]. These contributions still form the basis of our modern conception of evolution today. Although population genetics eventually came to be viewed as somewhat of a mathematical curiosity among biologists [226], it has recently witnessed a resurgence in popularity, as genetic data have become more readily available [85]. In this way, the original theoretical models of Fisher, Wright, Haldane, and others have now become the primary means for interpreting empirical patterns of DNA sequence variation in natural

populations.

For historical reasons, these early population genetic models mostly focused on genomes with no more than a single selected site, with the assumption that an actual genome can be regarded as an independent collection of such sites [305]. These models have provided important intuition about the interactions between mutation, selection, recombination, and genetic drift. But recent work has shown that they may be less well-suited to describing modern DNA sequence measurements [123, 61, 236], particularly in the rapidly evolving microbial populations that are our focus here. In the remainder of this chapter, we will briefly review these single-site models and their connection to DNA sequence data. We will not attempt to give an exhaustive review; several textbooks are already available [133, 339, 84], and the remaining chapters of the thesis will review the relevant literature as necessary. Our intent here is to provide a more pedagogical introduction aimed at physicists and quantitative biologists who are new to the discipline, but who are already familiar with the basic mathematical machinery. After reviewing these ideas, we will briefly describe how this thesis builds on existing population genetic theory to gain a better understanding of evolution in rapidly evolving populations.

1.1 Single-site models: tracking the fate of a mutation

One of the greatest difficulties in studying evolutionary dynamics is deciding on an appropriate mathematical model. On the one hand, we know that the underlying biology is complex. The reproductive success of an individual depends on a multitude of physiological and ecological details, which are at best only partially understood. On the other hand, experience from physics suggests that some *universality* might arise in systems with many individuals, such that the long-time, population-level behavior is controlled by only a few key features of the biology. This is an appealing premise, but it is difficult to prove in practice, and it provides little insight into which biological details will be the relevant ones. This suggests that it will be difficult to guess an appropriate model by working backwards from populations in nature.

At the opposite extreme, the field of population genetics has historically focused on specific toy models of evolution inspired by the ball-and-urn models of classical probability theory. Notable examples include the Galton-Watson process, the Wright-Fisher model, or the Moran model [84]. Although these were originally favored more for their mathematical rigor than their realism, the subsequent arrival of genetic data has lead to the widespread application of these models in many practical contexts as well. Here population geneticists appeal to their own sense of universality, in which the complexities of the actual evolutionary process can be absorbed into a few *effective parameters* that govern these toy models.

For our purposes, we will attempt to find a compromise between these two extremes, focusing on a particular toy model designed to mimic experimental evolution of microbes in the laboratory. The motivation

for this choice is partly practical: microbial evolution experiments will serve as the primary application of the theoretical ideas developed in this thesis, and it will be useful to have a model of these experiments in mind to facilitate such a comparison. However, this choice will also allow us to show how certain "universal" population genetic limits arise in a concrete example, which can be implemented with free-living cells. This will provide some explanation for the success of the traditional ball-and-urn models above, despite their abstract nature.

Microscopic dynamics

Our model attempts to approximate a batch culture experiment, in which a population of bacteria is grown in a well-mixed flask* of rich laboratory media [11]. The ancestral (or *wild-type*) strain is assumed to grow exponentially at rate r, such that the number of cells in the flask at some time Δt later is given by

$$n_{\text{wt}}(t + \Delta t) = n_{\text{wt}}(t)e^{r\Delta t}\,. \tag{1.1}$$

Note that if we measure time in generations (i.e., doublings), then $r = \ln 2$.

Eventually, the bacteria will exhaust the finite supply of nutrients in the flask, and the growth of the population will saturate. Before this happens, however, we assume that the experimenter transfers a small volume of cells into a fresh flask. This volume is chosen such that on average $\overline{N}_b$ cells are deposited into the new flask, although the actual number will follow a Poisson distribution around this mean value. By monitoring the density of cells, the experimenter can choose the volume to ensure that an average of $\overline{N}_b$ cells are still transferred to the next flask regardless of the fluctuations in the initial number of cells. The bacterial population dynamics will then resemble a succession of boom and bust cycles (Fig. 1.1), with a post-bottleneck size of $N_b \sim \overline{N}_b \pm \sqrt{\overline{N}_b}$.

We can introduce evolution into this model by imagining a competition between two genetically distinct strains. Suppose we have a mutant strain of bacteria with a slightly different growth rate, $r + s$, which will grow to size

$$n_{\text{mut}}(t + \Delta t) = n_{\text{mut}}(t)e^{(r+s)\Delta t} \tag{1.2}$$

over the course of the cycle. Since the mutant and wildtype cells are equally likely to be transferred into

*A similar approach can be carried out for evolution in a chemostat [263]. We leave this as an excercise for the reader.

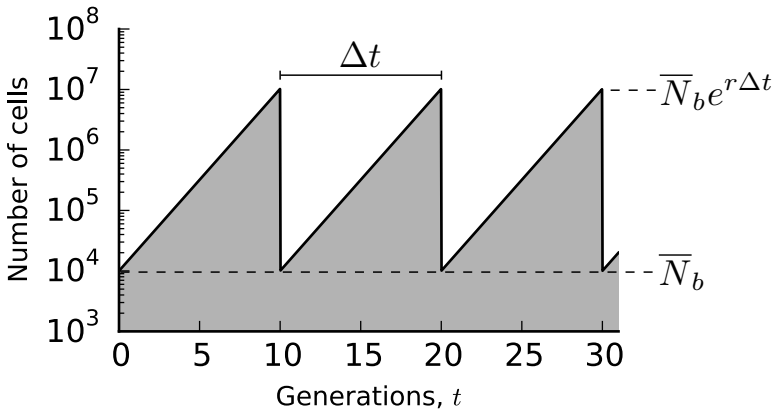

Figure 1.1: Population growth dynamics in a simplified batch culture model. In this example, $\Delta t = 10$ genenerations and $\overline{N}_b = 10^4$ cells.

the next flask, the fraction of mutant cells immediately after the transfer is simply

$$f(t+\Delta t) = \frac{1}{N_b(t+\Delta t)} \cdot \text{Binomial}\left[N_b(t+\Delta t),\ \frac{f(t)e^{s\Delta t}}{f(t)e^{s\Delta t}+1-f(t)}\right], \tag{1.3}$$

where the population size $N_b(t)$ is still a Poisson random variable with mean $\overline{N}_b$. Equation (1.3) provides a concrete rule for updating the frequency of the mutant strain after successive transfers. In this model, the cells are essentially competing for the chance to make it into the next flask. Those cells that grow faster have a natural advantage in this respect, with a magnitude determined by the product $s\Delta t$.

The continuum limit

We could analyze Eq. (1.3) directly, but additional simplifications arise in the limit that $\overline{N}_b$ is large and $s\Delta t$ is small. In this case, the change in $f(t)$ over any single transfer cycle is small, and we can coarse-grain over a large number of cycles to arrive at the continuum limit,

$$\frac{\partial f}{\partial t} = sf(1-f) + \sqrt{\frac{f(1-f)}{N}}\eta(t)\,, \tag{1.4}$$

where $N = \overline{N}_b \cdot \Delta t$ and $\eta(t)$ is a standard Brownian noise term. This stochastic differential equation should be interpreted in the Itô sense [99], which means that, in some infinitesimal time δt,

$$f(t+\delta t) = f(t) + s\delta t f(t)(1-f(t)) + \sqrt{\frac{\delta t f(t)(1-f(t))}{N}} \cdot \text{Normal}(0,1)\,. \tag{1.5}$$

The parameter s is known as the *selection coefficient* or the *fitness effect*, and N is the *effective population size*. In the population genetics literature, it is more common to write the Langevin dynamics in Eq. (1.4) in the form of a Fokker-Planck equation for the probability density, $\phi(f,t)$:

$$\frac{\partial \phi(f,t)}{\partial t} = \frac{\partial^2}{\partial f^2}\left[\frac{f(1-f)\phi(f,t)}{2N}\right] - \frac{\partial}{\partial f}\left[sf(1-f)\phi(f,t)\right]\,. \tag{1.6}$$

In this form, we see that our model is equivalent to the diffusion limit of population genetics [90, 84], in which the frequency of the mutant strain is modeled as a form of biased random walk. In retrospect, this equivalence is not that surprising: Eq. (1.4) is known to be the continuum limit of a large class of models [69], including the traditional Wright-Fisher model, the Moran model, and continuous-time branching processes, which differ only in their definition of the effective population size N. This provides a concrete example of the universality principle discussed above.

Note that in passing to the continuum limit, we have also lost the dependence of one of the parameters of the model. Examination of Eq. (1.6) shows that the quantities N, s, and t do not influence the model independently, but only through the scaled combinations Ns and t/N (or alternatively, Ns and st). This induces an effective "unit system" into the model, in which t, N, and $1/s$ all have effective units of time. Like any other unit system, this implies that observable quantities can only depend on unitless combinations of these parameters, which will provide a useful check for many of our calculations.

At first glance, Eq. (1.4) appears to resemble the diffusion processes that arise in the physical sciences, like the Brownian motion of a particle subject to a potential $V(x)$:

$$\frac{\partial x}{\partial t} = -\frac{\partial V(x)}{\partial x} + \sqrt{D}\eta(t)\,. \tag{1.7}$$

The key difference in Eq. (1.4) is that the effective diffusion constant (i.e., the step-size of the random walk) is itself a function of the mutant frequency, and it grows increasingly small when either the mutant or the wildtype become rare. In the extreme case where $f = 0$ or $f = 1$, the effective step size vanishes and the mutant frequency becomes frozen for all time. Of course, this is exactly the behavior we expect in an evolutionary setting: at long times the mutation must eventually either take over the population ($f = 1$) or go extinct ($f = 0$), after which no further evolution can occur. Note however, that although the

diffusive force grows weaker in an absolute sense, the *relative* size of the fluctuations become increasingly large compared to f when $f \to 0, 1$. Thus, even in an infinitely large population, number fluctuations play an important role whenever the mutant or the wildtype becomes rare. As we will see below, these two features lead to dramatically different dynamics than one would expect based on the analogy to Brownian motion alone.

Although the evolutionary model in Eq. (1.4) is simple to write down, it has no exact closed-form solution. The partial differential equation in Eq. (1.6) is linear in $\phi(f,t)$, so it does admit a formal solution in terms of an eigenfunction expansion,

$$\phi(x,t) = \sum_{k=1}^{\infty} \phi_k(x) e^{-\lambda_k t}\,, \tag{1.8}$$

subject to the correct boundary conditions (see [175, 323] for more details). However, in practice, this formal solution is somewhat opaque, and it provides limited insight into the evolutionary questions we might wish to answer. Fortunately, it is often possible to bypass Eq. (1.8) and obtain predictions for many other quantities more directly.

1.2 Genetic divergence: which mutations will fix?

One of the most basic quantities we can consider is the probability that the mutant strain will eventually take over the population rather than going extinct. This is known as the *fixation probability*, p_{fix}, and it depends on the dimensionless parameter Ns and the current frequency f. Predictions for p_{fix} are easiest to obtain from the moment generating function, $H(z,t) = \langle e^{-zf} \rangle$, whose equation of motion can be obtained by expanding $H(z, t+\delta t)$ and applying Eq. (1.5), or by taking the Laplace transform of Eq. (1.6):

$$\frac{\partial H(z,t)}{\partial t} = \left[sz - \frac{z^2}{2N} \right] \left[\frac{\partial H(z,t)}{\partial z} - \frac{\partial^2 H(z,t)}{\partial z^2} \right]. \tag{1.9}$$

This partial differential equation is no easier to solve than Eq. (1.6), but it has the advantage that $H(z,t)$ will remain smooth even when the probability density $\phi(f,t)$ contains δ-functions. By definition, the mutant strain starts at frequency f, so the generating function is initially given by $H(z,0) = e^{-zf}$. At long times, the mutation will either take over the population or go extinct, so that

$$\lim_{t\to\infty} H(z,t) = 1 \cdot (1 - p_{\text{fix}}) + e^{-z} \cdot p_{\text{fix}}\,. \tag{1.10}$$

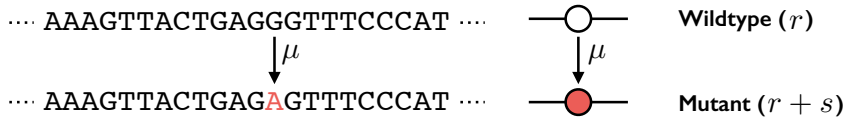

Figure 1.2: A toy model of spontaneous mutation. The mutant phenotype is controlled by the value of a particular site in the genome which mutates at rate μ. Neglecting the remaining sites, we can approximate the population as a collection of single-site genomes, each of which can be in the mutant or wildtype state.

In order to predict p_{fix}, we need to find a way to connect this equilibrium function with its initial condition at $t = 0$. Ordinarily this would be a hard problem, requiring the full time-dependent solution of Eq. (1.9). However, in this case we can appeal to a simple trick [142]. At the special value of $z = 2Ns$, the time derivative of $H(z, t)$ vanishes, which allows us to connect the initial and final values of $H(z, t)$ at this special point:

$$H(2Ns, 0) = \lim_{t\to\infty} H(2Ns, t)\,. \tag{1.11}$$

Solving for p_{fix}, we find that

$$p_{\text{fix}}(s, f) = \frac{1 - e^{-2Nsf}}{1 - e^{-2Ns}}\,. \tag{1.12}$$

This expression has all of the intuitive properties we would expect for a fixation probability: it vanishes when $f = 0$, approaches unity when $f = 1$, and increases monotonically with Ns. If there is no fitness difference between the strains, then $p_{\text{fix}} = f$, as expected from symmetry considerations.

Equation (1.12) predicts the outcome of a mutant/wildtype competition when the mutant is introduced at a given frequency. Although this is potentially useful in an experimental setting (see Chapter 6), in an actual population these mutants must be produced by spontaneous mutations in the wildtype strain. To model this case, we can imagine that the $r + s$ phenotype is controlled by the value of a single nucleotide in a particular site in the genome, and that mutations from the wildtype to the mutant state occur at rate μ per cell per unit time. In the absence of other mutations, we can ignore the remaining sites in the genome and model the population as a collection of single-site genomes, which can either be in the mutant or wildtype

state (Fig. 1.2). It is a straightforward exercise to show this changes our continuum model to

$$\frac{\partial f}{\partial t} = sf(1-f) + \mu(1-f) + \sqrt{\frac{f(1-f)}{N}}\eta(t)\,. \tag{1.13}$$

In many cases of interest, the per-site mutation rates are small compared to the other relevant timescales of system, which means that we can neglect the $-\mu f$ term in Eq. (1.13) and model the fate of each mutation independently. In other words, we assume that new mutations arise as a Poisson process with $N\mu$, and that the subsequent dynamics of these mutations is described by Eq. (1.4). Each of these mutations starts at frequency $f = 1/N$, and fixes with probability

$$p_{\text{fix}}(s, 1/N) = \frac{2s}{1 - e^{-2Ns}}\,, \tag{1.14}$$

so that *successful* mutants are produced at rate

$$\lambda = N\mu p_{\text{fix}}(s, 1/N) = \frac{2Ns\mu}{1 - e^{-2Ns}}\,. \tag{1.15}$$

Equation (1.15) is a key result in population genetics, since it allows us to predict the genetic changes that will accumulate in a population over time. These changes, known as *genetic divergence*, can be measured by comparing a DNA sequence sampled from the population with the DNA of the ancestor or a suitable outgroup (Fig. 1.3). For the single-site model in Eq. (1.4), all of the information about genetic divergence is encoded in the quantity d_N, which is the probability of observing a mutation at the site in question. In practice, it is difficult to measure a per-site d_N based on data from a single population, since a given site will either contain a fixed mutation or not. Instead, divergence is typically estimated for a class of putatively similar sites across the genome by calculating the fraction of sites that have mutated (see Fig. 1.3). For example, the average divergence between the human and chimpanzee genomes is of order $d_N \sim 1\%$ [313], while the divergence between human and mouse is much larger ($d_N \sim 50\%$ [53]).

To relate d_N to the model in Eq. (1.4), we note that a randomly sampled individual will contain the mutation with probability $f(t)$, where the value of $f(t)$ will vary from population to population. Averaging over the ensemble, this implies that

$$d_N = \langle f(t) \rangle\,. \tag{1.16}$$

We cannot evaluate Eq. (1.16) exactly, but we can approximate it in the limit of long times and low mutation rates. In this limit, most populations in the ensemble will either have fixed the mutation or are still waiting

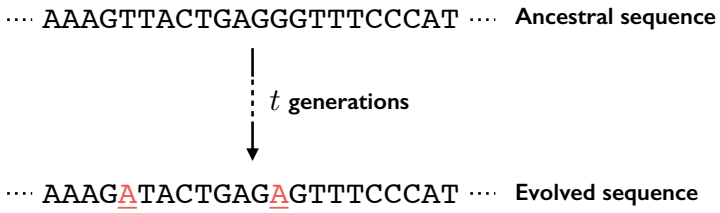

Figure 1.3: Genetic divergence between ancestral and evolved DNA sequences. In this example, the average divergence is $d_N \approx 2/22$.

for it to arise, so that

$$d_N \approx \lambda t = \frac{2Ns\mu t}{1 - e^{-2Ns}} \,. \tag{1.17}$$

In practice, we often do not know the divergence time (or the mutation rate) when comparing DNA sequences from natural populations. It is common to remove the dependence on t by normalizing the observed divergence by the average value at a class of putatively neutral sites (e.g., those leading to synonymous codon changes), for which $d_S = \mu t$. This yields a second quantity

$$\frac{d_N}{d_S} = \frac{2Ns}{1 - e^{-2Ns}} \,, \tag{1.18}$$

which no longer depends on μ or t. This quantity then serves as a measure of natural selection: values of $d_N/d_S > 1$ indicate adaptation or positive selection (i.e., selection acts to change the wildtype sequence), while $d_N/d_S < 1$ indicates negative or purifying selection (i.e., selection acts to preserve the wildtype sequence). For example, when comparing the human and mouse genomes, protein coding sites have $d_N/d_S \sim 0.1$, indicating predominantly negative selection [53]. Values of d_N/d_S for the human-chimpanzee and mouse-rat comparisons are similar, even though the timescales involved are much shorter [313].

1.3 Asymptotics and heuristic analysis

Equation (1.12) is "exact" in the sense that it can be derived Eq. (1.4) without[†] approximation, and is valid for arbitrary values of Ns. Closed-form expressions like Eq. (1.12) are highly prized in population genetics,

[†] We must be careful to remember that "exactness" is relative, as Eq. (1.4) is merely an approximation to the discrete model in Eq. (1.3), which is itself just a toy model of competition between two strains in the laboratory.

since they can easily be inverted into inference methods for estimating parameters (in this case Ns) from data. However, the quest for exact results can also have its drawbacks. The analysis above provides little insight into *why* the fixation probability is given by Eq. (1.12), and the derivation breaks down under slight generalizations of the model, e.g. if we allow N or s to vary in time.

Fortunately, this level of exactness is often unnecessary. The reason is that our evolutionary models are typically parameterized by numbers that are either very large (like the population size) or very small (like mutation rates or selection coefficients). We have already taken advantage of this fact in passing to the continuum limit in Eq. (1.4). A second consequence of this fact is that, given a random combination of large N and small s, the magnitude of their product is itself very likely to either be very large or very small. Intermediate values of $N|s| \sim \mathcal{O}(1)$ would typically be unlikely unless there is some additional constraint to enforce this fine-tuning. In the absence of such constraints, we should be able to get a *nearly* complete solution to the problem by analyzing Eq. (1.4) in the asymptotic limits where $N|s| \ll 1$ and $N|s| \gg 1$.

Although it might seem like we have only made the problem more complicated (or less accurate) by focusing on these limits, we will see that they encapsulate the essential features of the dynamics in a way that is robust to slight perturbations of the model. In the following sections, we present a heuristic[‡] analysis of Eq. (1.4) in these two limits. Our presentation will be similar in spirit to [70, 91, 106], although the analysis will be slightly different. This heuristic approach will provide a more mechanistic picture of the dynamics in Eq. (1.4), and it will enable us to rederive $p_{\text{fix}}(s, f)$ (and many other quantities) in a more transparent way. Most importantly, the intuition gleaned from this excercise will become useful later in the thesis, when we attempt to extend the model to more complicated scenarios.

Dynamics of a neutral mutation

In the limit that $Ns \to 0$, the selection term in Eq. (1.4) can be neglected, and

$$\frac{\partial f}{\partial t} = \sqrt{\frac{f(1-f)}{N}}\eta(t)\,. \tag{1.19}$$

Since genetic drift is unbiased, $\partial_t \langle f(t) \rangle = 0$, and the average frequency of the mutation is equal to $\langle f(t) \rangle = f$ for all t. At long times, the mutation must either fix or go extinct, so that

$$\lim_{t\to\infty} \langle f(t) \rangle = 1 \cdot p_{\text{fix}} + 0 \cdot (1 - p_{\text{fix}})\,. \tag{1.20}$$

[‡]Although these calculations are presented as heuristics, they can all be made more rigorous using standard techniques from asymptotic analysis [141]. We leave this as an excercise to the reader; a "problem set" describing these calculations is available at https://github.com/benjaminhgood/popgen_notes.

These two relations can be combined to show that $p_{\text{fix}} = f$, in agreement with the $s \to 0$ limit of Eq. (1.12). However, this argument also shows how $\langle f(t) \rangle$ can be a misleading summary of $f(t)$, since it is very far from the *typical* value of $f(t)$ at sufficiently long times. So what does the trajectory of an actual mutation look like? And approximately how much time is required for the mutation to transition to its final state?

To investigate these questions, let us consider a mutation that starts at a small initial frequency $f(t) \ll 1$. In an infinitesimal time δt, the definition of the model implies that $f(t)$ wanders diffusively:

$$\delta f \equiv f(t + \delta t) - f(t) = \sqrt{\frac{f(t)\delta t}{N}} \cdot \text{Normal}(0, 1) \tag{1.21}$$

But how does this extend to finite intervals, Δt, where the finite changes in $f(t)$ can alter the effective diffusion constant in Eq. (1.19)?

We can analyze this case heuristically by employing a crude form of integration. First, we note that Eq. (1.21) should continue to be a good approximation as long as $\Delta f \ll f$, since $f(t)$ will be effectively constant throughout the time interval. On these timescales, the average $\langle f(t) \rangle$ is still a good approximation to the typical value of $f(t)$. However, this will eventually break down for times of order $\Delta t \sim Nf$, when $\Delta f \sim \pm f$. With probability $\frac{1}{2}$, the perturbation will be in the negative direction, and the mutation will likely die out. Otherwise, the perturbation will be positive and the process will start over at an initial frequency of $2f$. Repeating this line of reasoning, we find that after an additional $2Nf$ generations, the mutation will again be nearly extinct or else have grown to frequency $4f$, and so on. Iterating this process d times, we see that the population will reach size $f(t + \Delta t) = f(t)2^d$ with probability $\left(\frac{1}{2}\right)^d$, and it will take a time of order $\Delta t \sim \sum_{k=0}^{d-1} Nf(t)2^k$ to do so. We can summarize this behavior in two equivalent ways depending on whether the time interval Δt or the frequency change Δf is the independent variable:

1. With probability $Nf/(Nf + \Delta t)$ the mutant will survive for Δt generations and grow to size $f + \Delta t/N$.
2. With probability $f/(f + \Delta f)$, the mutant will reach size $f + \Delta f$, and it will take $\Delta t = N\Delta f$ generations to do so.

Note how the displacement scales as a linear function of Δt, rather than the characteristic $\sqrt{\Delta t}$ dependence observed on short timescales.

In order to fix, the mutant must drift to frequencies of order $\mathcal{O}(1)$. The analysis above shows that this will happen with probability

$$p_{\text{fix}} \sim f\,, \tag{1.22}$$

and that it will require a time of order

$$T_{\text{fix}} \sim N \tag{1.23}$$

to do so. Conditioned on eventual success, the linear relationship between Δf and δt implies that a typical mutant trajectory will spend equal amounts of time at each frequency between f and 100% before it fixes. Most of this time is spent traversing intermediate frequencies (e.g., 80% of the fixation time is spent between $f = 10\%$ and $f = 90\%$).

These timescales also provide a way to check the validity of the neutral approximation when $N|s|$ is small but finite. When the selection term is added back to Eq. (1.19), it contributes a change of order $\Delta f_{\text{sel}} \sim N|s|$ over the $\mathcal{O}(N)$ generations that the mutation can drift before it fixes. Since the actual change in $f(t)$ is $\mathcal{O}(1)$, we see that selection will be a negligible correction whenever $N|s| \ll 1$.

Dynamics of a selected mutation

In the opposite limit, where $Ns \to \infty$, we expect that selection will be the dominant evolutionary force. If we neglect genetic drift entirely, Eq. (1.4) reduces to the deterministic dynamics

$$\frac{\partial f}{\partial t} = sf(1-f)\,. \tag{1.24}$$

The solution is ordinary logistic growth,

$$f(t+\Delta t) = \frac{f(t)e^{s\Delta t}}{1 - f(t) + f(t)e^{s\Delta t}}\,, \tag{1.25}$$

which will drive beneficial mutations to $f(t) \to 1$ and deleterious mutations to $f(t) \to 0$. However, we can immediately see that something wrong this with approximation: it predicts that the fixation probability of a beneficial mutant is $p_{\text{fix}}(s, f) \approx 1$, and that it takes the mutation $T_{\text{fix}} = \frac{1}{s}\log\left(\frac{Nf}{1-f}\right) \approx \infty$ generations to fix!

The problem is that genetic drift is not *entirely* negligible in Eq. (1.4), even in the limit that $Ns \to \infty$. To see this, let us again focus on a mutation that starts at an initial frequency $f \ll 1$. In an infinitesimal time δt, selection and drift change the frequency of the mutation by an amount

$$\delta f \sim \underbrace{sf(t)\delta t}_{\text{selection}} + \underbrace{\sqrt{\frac{f(t)\delta t}{N}}\text{Normal}(0,1)}_{\text{genetic drift}}\,, \tag{1.26}$$

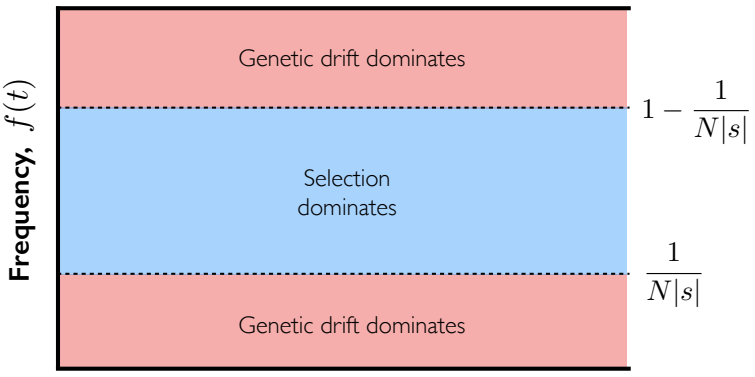

Figure 1.4: A schematic illustration of the three distinct frequency regimes in Eq. (1.4) in the limit that $Ns \gg 1$. Genetic drift is only negligible in the middle region.

As we argued before, this will continue to be a good approximation for finite Δt provided that $\Delta f \ll f$, but it will eventually break down when $\Delta f \sim f$, or

$$f \sim \underbrace{|s| f \Delta t}_{\text{selection}} + \underbrace{\sqrt{\frac{f \Delta t}{N}}}_{\text{genetic drift}} . \tag{1.27}$$

When $N|s|f \gg 1$, we can see that the selection term is indeed the dominant contribution over this time interval, as we anticipated in Eq. (1.24). But for $N|s|f \ll 1$, genetic drift is actually the dominant evolutionary force. This suggests that the proper approximation to Eq. (1.4) for large Ns is the piecewise model,

$$\frac{\partial f}{\partial t} \approx \begin{cases} \sqrt{\frac{f}{N}}\eta(t) & \text{if } f \lesssim 1/N|s|, \\ sf(1-f) & \text{if } f(1-f) \gg 1/N|s|, \\ \sqrt{\frac{(1-f)}{N}}\eta & \text{if } 1-f \ll 1/N|s|, \end{cases} \tag{1.28}$$

which is illustrated in Fig. 1.4. A mutation that is introduced at an "observable" frequency [$f(1-f) \ll 1/N|s|$] will behave nearly deterministically, so that we can estimate s by monitoring its frequency over a

pair of timepoints:

$$s = \frac{1}{\Delta t} \log\left(\frac{1 - f(t)}{f(t)} \frac{f(t + \Delta t)}{1 - f(t + \Delta t)} \right) . \tag{1.29}$$

This is the standard method used to estimate fitness in laboratory evolution experiments. However, a spontaneous mutation will always start at an initial frequency $1/N \ll 1/N|s|$, so we cannot predict its probability of fixation without accounting for the interplay between selection and genetic drift.

When a mutation starts at a frequency $f \ll 1/N|s|$, it will initially evolve according to the neutral dynamics discussed above: with probability $N|s|f$, it will drift to size $1/N|s|$, and it will require a time of order $\Delta t \sim 1/s$ to do so. At this point, natural selection takes over. A beneficial mutation will start to grow logistically, and it will reach size $1 - 1/Ns$ after $\Delta t \sim \frac{1}{s}\log(Ns)$ generations. A deleterious mutation will be pinned near $1/N|s|$, and will typically drift to extinction within $\Delta t \sim 1/|s|$ generations. Similarly, a beneficial mutation near $f(t) \sim 1 - 1/Ns$ will typically drift to fixation within the next $1/s$ generations, while a deleterious mutation near $f(t) \sim 1 - 1/N|s|$ will fix with probability $N|s|(1 - f)$. Combining these facts, we conclude that beneficial mutations will fix with probability

$$p_{\text{fix}}(s, f) \sim \begin{cases} 1 & \text{if } f \gtrsim 1/Ns, \\ Ns & \text{if } f \lesssim 1/Ns, \end{cases} \tag{1.30}$$

while deleterious mutations will fix with probability

$$p_{\text{fix}}(s, f) \sim \begin{cases} 0 & \text{if } 1 - f \gtrsim 1/N|s|, \\ Ns(1 - f) & \text{if } 1 - f \lesssim 1/N|s|. \end{cases} \tag{1.31}$$

One can check that these expressions agree with the exact formula in Eq. (1.12) in the appropriate limits. For the special case of a spotaneous mutation ($f = 1/N$), we recover the simple formula,

$$p_{\text{fix}}(s, 1/N) \sim \begin{cases} s & \text{if } s > 0, \\ 0 & \text{if } s < 0. \end{cases} \tag{1.32}$$

which matches the large Ns limit of Eq. (1.14). A mutation that is twice as beneficial will fix twice as often, while a deleterious mutation will essentially never fix. Conditioned on fixation, a new beneficial mutation

will require

$$T_{\text{fix}} \approx \frac{1}{s} + \frac{2}{s}\log(Ns) + \frac{1}{s} \sim \frac{1}{s}\log(Ns) \tag{1.33}$$

generations to fix. This timescale is dominated by the time that it takes the mutation to travel from one drift-dominated regime ($\sim 1/Ns$) to the other ($\sim 1 - 1/Ns$), which is somewhat similar to the neutral case above. However, unlike neutral mutations, which spend equal amounts of time at these intermediate frequencies, these strongly beneficial mutations spend much of their time traversing "unobservable" frequencies where $1/Ns \ll f(1-f) \ll 1$. E.g., only a fraction $\log(9)/\log(Ns)$ of the fixation time is spent between $f = 10\%$ and $f = 90\%$.

Our heuristic analysis also provides a natural way to identify the limits of the quasi-deterministic approximation in Eq. (1.28). When $1/N|s| = 1 - /N|s| \sim 1/2$, the drift-dominated region of Eq. (1.28) will cover the entire range of frequencies, and the mutation will behave as if it were effectivley neutral. This transition to nearly neutral behavior for $|s| \lesssim 1/N$ is known as the *drift barrier*, since selection is unable to efficiently act on mutations that are weaker than this threshold [267]. This is one example where dynamical considerations (e.g., the strength of genetic drift) can place fundamental limitations on how well evolution can optimize certain aspects of biology [219].

1.4 Genetic diversity: making inferences about the past

So far, we have focused on genetic changes that accumulate over long timescales, when most successful mutations have fixed in the population. The advantage of focusing on these fixed differences is that they can be measured by sequencing a single individual from the population. But as the costs of sequencing have fallen, it is increasingly common to sequence DNA from multiple individuals in the population (an experimental design known as *resequencing*). The genetic differences between these contemporary sequences can then provide information about the *genetic diversity* within the population as well (Fig. 1.5). A major focus of modern population genetics is to use these within-population differences, also known as *polymorphic mutations*, to make inferences about the recent evolutionary history of various natural populations (e.g. humans) [285].

For the single-site model in Eq. (1.13), all information about genetic diversity is contained in the *polymorphic site frequency spectrum* (SFS),

$$P_n(i) = \left\langle \binom{n}{i} f^i (1-f)^{n-i} \right\rangle , \tag{1.34}$$

which represents the probability of observing a mutation in i individuals in a sample of size n. When

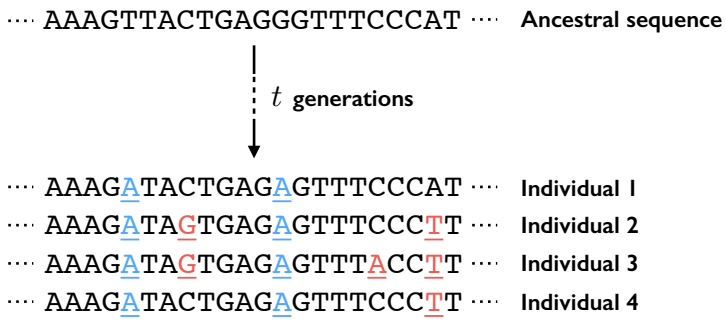

Figure 1.5: Genetic diversity in a sample of contemporary DNA sequences. Fixed mutations are shown in blue and polymorphic mutations are shown in red.

$i = n$, this statistic corresponds to the fixed-differences studied in the previous section; polymorphic mutations require that $1 \leq i \leq n-1$, so we will focus on this case from now on. The simplest version of the SFS, $P_2(1)$, is canonically designated by the variable π and is known as the *pairwise heterozygosity*[§]. In humans, the pairwise heterozygosity is of order $\pi \sim 0.1\%$, and values in other organisms range from $\pi \sim 0.01\%$ to $\pi \sim 10\%$ [202].

If we know the SFS for a sample of size n, it is easy to show that we can recover the SFS for a smaller sample using the "downsampling" formula,

$$P_{n'}(i') = \sum_{i=i'}^{n-n'+i'} \frac{\binom{i}{i'}\binom{n-i}{n'-i'}}{\binom{n}{n'}} P_n(i)\,, \tag{1.35}$$

which, in the case of $n' = 2$ and $i' = 1$, yields an empirical formula for the pairwise heterozygosity,

$$\pi = \frac{2n}{n-1} \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)\left(1 - \frac{i}{n}\right) P_n(i)\,. \tag{1.36}$$

This shows that, at least from a theoretical perspective, all information about genetic diversity is contained in the large-n limit of the SFS,

$$p(f) \equiv \lim_{n\to\infty} P_n(nf)\,, \tag{1.37}$$

[§]This name stems from the fact that, in diploid organisms, π is also the probability that two chromosomes from the same individual will differ at a particular site.

which is related to the finite-sample SFS through the integral relation,

$$P_n(i) = \int \binom{n}{i} f^i (1-f)^{n-i} p(f)\, df \,. \tag{1.38}$$

In the previous section, we calculated moments like $\langle f \rangle$ by assuming that at long times, most of the contribution to the mean was due to fixed mutations. This assumption no longer applies for the higher-order moments in the definition of Eq. (1.34), since they always include at least one factor of $(1-f)$ which zeros out any contribution from fixed mutations. In principle, this means that it is much more difficult to calculate diversity statistics like $p(f)$ than it is to calculate divergence statistics like p_{fix}, since the former will depend on the entire trajectory of the mutation in addition to its ultimate fate.

Fortunately, we can again appeal to a trick that yields a closed-form solution for $p(f)$ with minimal effort. We note that the *shape* of $p(f)$ is unchanged if we replace Eq. (1.13) with an altered model, in which the mutation is immediately restarted at some infinitesimal frequency ϵ whenever it fixes or goes extinct [83]. This modification will only change the overall scale of $p(f)$ by some constant factor. For frequencies larger than ϵ, the altered $p(f)$ will satisfy the same equation as $\phi(x,t)$ in Eq. (1.6), except that by definition the time derivative vanishes for $p(f)$:

$$0 = \frac{\partial^2}{\partial f^2}\left[\frac{f(1-f)p(f)}{2N}\right] - \frac{\partial}{\partial f}\left[sf(1-f)p(f)\right] \,. \tag{1.39}$$

It is a straightforward excercise to solve Eq. (1.39), subject to the condition that $p(1) < \infty$, to obtain

$$p(f) \propto \frac{1-e^{-2Ns(1-f)}}{f(1-f)} \,, \tag{1.40}$$

which fixes $p(f)$ up to an overall constant.

We can gain some intuition about the functional form of Eq. (1.40) by appealing to our heuristic picture above. We saw that successful beneficial mutations spend most of their lives growing logistically from $f \sim 1/Ns$ to $1-f \sim 1/Ns$. In this frequency range, $p(f) \sim 1/f(1-f)$, which simply corresponds to averaging the logistic growth function over the unknown birth-time of the mutation. We can estimate the SFS for a neutral mutation in a similar way: with probability $1/t$, a mutation that arose t generations ago will survive to reach size $f \sim t/N$, which implies that $p(f) \propto 1/f$. Finally, the SFS for a deleterious mutation will be similar to the neutral case when $f \ll 1/N|s|$, but selection prevents the mutation from rising much above this threshold.

To compare Eq. (1.40) with data, we must solve for the unknown normalization constant, which will depend on Ns and $N\mu$. To do so, it is useful to first use the downsampling formula in Eq. (1.36) to rewrite

the normalization in terms of the pairwise heterozygosity,

$$p(f) = \frac{2Ns\pi \left(1 - e^{-2Ns(1-f)}\right)}{2f(1-f)(2Ns - 1 + e^{-2Ns})}\,, \tag{1.41}$$

Then to determine π, we note that we can take an average of Eq. (1.13) to show that

$$\frac{\partial \langle f \rangle}{\partial t} = s\langle f(1-f)\rangle + \mu\,. \tag{1.42}$$

In the section on fixed mutations, we showed that $\partial_t \langle f \rangle = N\mu p_{\text{fix}}(s, 1/N)$. This leads to a simple relation between π and $p_{\text{fix}}(s, 1/N)$,

$$\pi = \frac{2\mu}{s}\left[N p_{\text{fix}}(s, 1/N) - 1\right]\,. \tag{1.43}$$

Using the formula for p_{fix} in Eq. (1.12), we find that

$$\pi = \frac{2\mu}{s}\left[\frac{2Ns - 1 + e^{-2Ns}}{1 - e^{-2Ns}}\right] \sim \begin{cases} 4N\mu & \text{if } Ns \gg 1, \\ 2N\mu & \text{if } N|s| \ll 1, \\ \frac{2\mu}{|s|} & \text{if } s < 0 \text{ and } N|s| \gg 1, \end{cases} \tag{1.44}$$

Again, we can understand the asymptotic behavior of π by appealing to the heuristic calculations above. Mutations arise at rate $N\mu$, and contribute to $\pi \sim \langle f(1-f)\rangle$ before they fix or go extinct. A neutral mutation will reach frequencies where $f(1-f) \sim \mathcal{O}(1)$ with probability $1/N$, and it will drift at these frequencies for $\sim N$ generations. The heterozygosity is therefore given by $\pi \sim N\mu \cdot 1 \cdot \frac{1}{N} \cdot N \sim N\mu$. In comparison, a strongly beneficial mutation will reach $f(1-f) \sim \mathcal{O}(1)$ with a much higher probability ($p_{\text{fix}} \sim s$), but it will only spend $1/s$ generations passing through $f(1-f) \sim \mathcal{O}(1)$, so that $\pi \sim N\mu \cdot 1 \cdot s \cdot \frac{1}{s} \sim N\mu$. This means that beneficial and neutral mutations have similar contributions to pairwise diversity, despite their vastly different rates of genetic divergence. On the other hand, a strongly deleterious mutation can never reach frequencies where $f(1-f) \sim \mathcal{O}(1)$ — the best it do is to reach frequencies of order $f \sim 1/N|s|$, which happens with probability $\sim |s|$. The mutation drifts at these frequencies for $1/|s|$ generations, so that $\pi = N\mu \cdot \frac{1}{N|s|} \cdot |s| \cdot \frac{1}{|s|} \sim \mu/|s|$. As a result, strongly deleterious mutations will contribute little to diversity as well as divergence.

1.5 Neutral theory and the coalescent

The situation becomes particularly simple in the *neutral limit*, when most sites have $|s| \ll 1/N$. This scenario was originally championed by Kimura [177], who saw it as a way to explain the surprisingly high levels of intra-population variability in protein electrophoresis measurements [211, 130]. It was also motivated by the (then) recent discovery of the degenerate genetic code [261], which suggested that many DNA sequence changes could have no discernible effect on the phenotype. In the past half century, the *neutral theory* has grown into the standard "null model" of population genetics, and a rich body of theoretical, statistical, and computational techniques have been developed around this assumption [133].

In the neutral limit, genetic diversity is exclusively governed by drift, and the predictions for π and $P_n(i)$ reduce to

$$\pi = 2N\mu\,, \tag{1.45a}$$

$$P_n(i) \propto \frac{1}{i}\,. \tag{1.45b}$$

These simple relations have been extremely influential in population genetics. Equation (1.45a) suggests that we can infer the population size directly from genome-wide estimates of heterozygosity. In fact, the neutral theory has become so widely ingrained in modern population genetics that the quantity

$$N_e = \frac{\pi}{2\mu} \tag{1.46}$$

is sometimes *defined* to be the "effective" population size [46]. E.g., in humans, the observation that $\pi \sim 10^{-3}$ is often reported as $N_e \sim 10^4$, while *Drosophila melanogaster* has $N_e \sim 10^6$ [46]. In this way, one may even talk about variation in N_e along the genome [120]. This variation can sometimes reflect actual differences in population size (e.g., there are only $\sim 3/4$ as many X chromosomes as there are autosomes in the human population). However, the observed variation cannot always be explained in this way (e.g., the "N_e" for the Y chromosome is ~ 30 times lower than the X chromosome [354]). In these cases, N_e can only be interpreted as an effective parameter that may not represent any physical population of genes.

Equation (1.45b) has been equally influential, since it predicts that the frequencies of genetic polymorphisms should be distributed according to the parameter-free distribution $p(f) \sim 1/f$. This strong prediction, which is readily testable with modern sequence data, has spawned an entire subfield of neutrality tests (e.g., Tajima's D [329]) which aim to reject the neutral null model in a rigorous statistical setting [259].

We derived Eqs. (1.45a) and (1.45b) above by averaging over mutation frequency trajectories. But they can also be obtained using an entirely different picture known as the *coalescent*, which emphasizes the

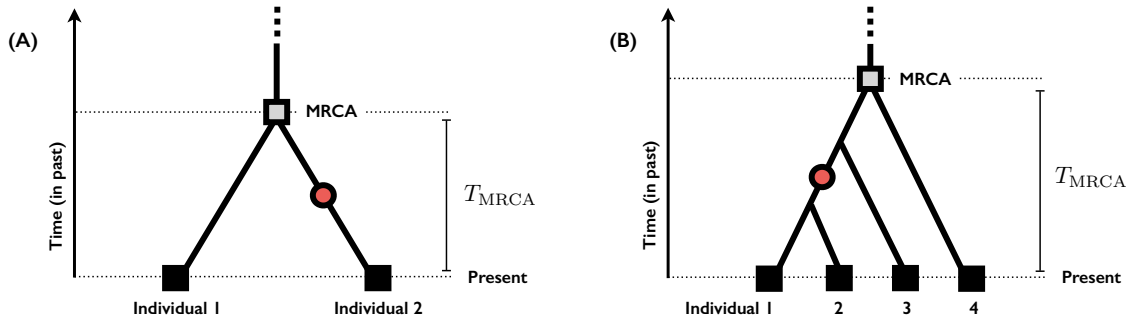

Figure 1.6: (A) Genealogy for a sample of $n = 2$ individuals, which traces the line of descent back in time until the lineages ``coalesce'' into their most recent common ancestor. Mutations that occur before coalescence (red circle) are polymorphic. (B) Genealogy for a sample of $n = 4$ individuals. The red mutation is present in $i = 2$ of the n individuals.

genealogical relationships among the sampled DNA sequences. A random pair of individuals will share a common ancestor T_{MRCA} generations in the past (Fig. 1.6). Neutral mutations will neither influence nor be influenced by the value of T_{MRCA}, and will occur with probability μ per unit time along the branches of the genealogy. In order for a mutation to be polymorphic in the sample, it must occur on one of the ancestral branches before they have merged (or *coalesced*) into their most recent common ancestor (see Fig. 1.6A). The pairwise heterozygosity is therefore given by

$$\pi = 2\mu\langle T_{\mathrm{MRCA}}\rangle \, . \tag{1.47}$$

In a neutral population, two individuals will share a common ancestor in the previous generation with probability $1/N$, so that $\langle T_{\mathrm{MRCA}}\rangle = N$, and $\pi = 2N\mu$ as expected. Predictions for the SFS follow from similar considerations, except that now each of the $\binom{n}{2}$ pairs of lineages can coalesce with probability $1/N$. When two lineages coalesce, they are replaced by their common ancestor, and the process continues until all n lineages have coalesced into a single common ancestor (see Fig. 1.6B).

The advantage of this line of reasoning is that it can be immediately extended to cases where the population size varies in time. In this case, pairs of individuals coalesce with probability $1/N(t)$, where $N(t)$ is the size of the population going backwards in time. For a sample of size $n = 2$, the distribution of T_{MRCA} is given by

$$\Pr[T_{\mathrm{MRCA}} \leq t] = 1 - e^{-\int_0^t \frac{dt'}{N(t')}\, dt'} \, , \tag{1.48}$$

so that

$$\pi = 2\mu \int_0^\infty e^{-\int_0^t \frac{dt'}{N(t')}}\, dt\,. \tag{1.49}$$

Predictions for the site frequency spectrum can be obtained in a similar way, either from more complicated numerical expressions [119] or by directly simulating the coalescent process backwards in time [143]. This is extremely important, since frequency spectra in natural populations rarely agree with Eq. (1.45b). The simplest way to "fix" this (and still preserve the nice properties of the neutral model) is to assume that the population size has varied with time (which is surely true at some level).

A variety of computational methods now exist to infer $N(t)$ from the observed SFS [121, 29]. A related class of methods exploits the fact that sufficiently closely spaced sites will often share the same genealogy, rather than evolving independently. For a pair of individuals, the probability of observing k differences in a block of L such sites is given by

$$\Pr[\Pi = k] = \left\langle \frac{(L\mu T_{\mathrm{MRCA}})^k}{k!} e^{-L\mu T_{\mathrm{MRCA}}} \right\rangle . \tag{1.50}$$

When aggregated across the genome, variation in Π will provide information about the distribution of T_{MRCA}, and hence $N(t)$ [212, 276, 131, 314]. The accuracy of such estimates can be difficult to verify; however, in humans (one of the best studied examples), many qualitative features of the inferred demographic history appear to be consistent with the archaeological record [212].

1.6 Multi-locus models and the importance of linkage

Our discussion thus far has focused on the competition between two variants at a single site in the genome. But in practice, real genomes contain anywhere from several thousand sites (in the smallest viruses) to several billion sites (in higher eukaryotes), so that many more than two variants are possible. What use are our single-site models in this case?

As we alluded to earlier, there are three scenarios where single-site genomes can still be useful as *effective models*, even when we know that they are embedded in a larger genome:

1. Evolution is perfectly neutral. If there are no fitness differences, sites cannot influence each other by definition, and neighboring sites are only relevant when investigating correlations between them. In this case, there is little evolution in the traditional sense: DNA is just another phenotype that is slaved to the genealogy of the population.

2. Evolution is perfectly sequential. Either the population size or the mutation rate is sufficiently small

that no more than two variants are ever present in the population at the same time, and we may safely restrict our attention to the most recently mutated site. On long timescales, the population can be viewed as a single point that wanders through the space of possible genotypes. This is known as the *weak mutation limit* [228].

3. Evolution is perfectly parallel across the genome. Neighboring sites recombine sufficiently frequently that they effectively evolve independently. This is known as *free recombination* or *linkage equilibrium* [254]

If the goal is to investigate evolution (rather than demography), then the most interesting regions of the genome will naturally be those where selection plays an important role. In most of these cases, evolution will neither be perfectly sequential nor perfectly parallel. Instead, entire chromosomes are inherited and selected as a unit. Since neighboring sites are rarely broken up by recombination, blocks of L consecutive sites will often be inherited together for their entire genealogical history [251]. The number of sites in each block can vary dramatically depending on the rate of recombination: in humans, L is estimated to be $\sim 10^4$ [341], while self-fertilizing organisms like nematodes can have much larger blocks ($L \sim 10^6$) [293]. In asexually reproducing microbes and cancer tumors, L can encompass the entire genome.

To model evolution in one of these non-recombining blocks, we must in principle track the frequencies of all 2^L possible¶ genotypes. We can index these genotypes with an L-dimensional vector $\vec{g} \in \{0,1\}$, so that the frequency of each genotype is denoted by $f(\vec{g})$. A natural generalization of the continuum model in Eq. (1.13) in this case is

$$\frac{\partial f(\vec{g})}{\partial t} = \underbrace{\left[X(\vec{g}) - \sum_{\vec{g'}} X(\vec{g'}) f(\vec{g'})\right] f(\vec{g})}_{\text{selection}} + \underbrace{\sum_{|\vec{g'}-\vec{g}|=1} \left[\mu_{\vec{g'}\to\vec{g}} f(\vec{g'}) - \mu_{\vec{g}\to\vec{g'}} f(\vec{g})\right]}_{\text{mutation}} + \underbrace{\sum_{\vec{g'}} \left[\delta_{\vec{g},\vec{g'}} - f(\vec{g})\right] \sqrt{\frac{f(\vec{g'})}{N}} \eta(\vec{g'})}_{\text{genetic drift}}, \tag{1.51}$$

where $\mu_{\vec{g}\to\vec{g'}}$ is the mutation rate from $\vec{g}$ to $\vec{g'}$, and $X(\vec{g})$ is the fitness of genotype $\vec{g}$. Like Eq. (1.13), this model includes a selection term, which acts to amplify genotypes that are more fit than average, a mutation term, which converts individuals from one genotype to another, and a term corresponding to genetic drift,

¶Since there are four letters in the DNA alphabet (A,T,C,G), it might be more correct to say that there are 4^L possible genotypes. However, once we account for insertions and deletions, the true number can be higher still.

whose complicated correlation structure is designed to ensure that $\sum_{\vec{g}} f(\vec{g}, t) = 1$.

In contrast to the $L = 1$ case, the behavior of Eq. (1.51) is still poorly understood. A large body of existing work has focused on the $L = 2$ case, where some exact results can still be obtained [84]. But until quite recently, few results have been available for larger L, except in highly symmetric special cases.

The major difficulty for $L \gg 1$ is that the space of genotypes is incredibly vast. This obviously creates bookkeeping issues for Eq. (1.51), but the actual problems are more fundamental than that. When $NL\mu \gtrsim 1$, evolution will not be perfectly sequential, since multiple genetic variants will arise before any single variant can sweep to fixation. In this case, the population can no longer be approximated by a single pair of mutant and wildtype strains, but will instead spread out into a "cloud" of nearby genotypes that compete with each other for dominance. These "genotype clouds" have traditionally been analyzed by appealing to deterministic equilibrium approximations known as *mutation-selection balance*, in which the long-time behavior of $f(\vec{g}, t)$ is assumed to be close to its expected value in the absence of genetic drift [84, 78, 41]. But this places implicit restrictions on L, which are often violated in practice. In particular, since the number of genotypes grows exponentially with L, we will often have $N \ll 2^L$, even when N itself is large. The vast majority of genotypes will therefore be unoccupied, and the *typical* values of $f(\vec{g}, t)$ will be very different than the expected value at mutation-selection balance. Instead, the typical values of $f(\vec{g}, t)$ will be dominated by the transient dynamics of the "genotype cloud" as it moves through isolated regions of genotype space. In the language of statistical physics, this means that replica symmetry is broken, and that standard mean-field approaches can no longer be applied. To predict the genetic structure of the population in this regime, we must explicitly account for these correlated transient dynamics, in which drift, selection, and mutation will all play an important role.

1.7 Overview of the thesis

One of the major aims of this thesis is to characterize the large-L behavior of Eq. (1.51) in more biologically relevant scenarios, with a particular focus on generating predictions for the genetic divergence and diversity statistics introduced above. Much of this work focuses on a special case of Eq. (1.51), where the fitness can be written as a linear function of the genotype:

$$X(\vec{g}) = \sum_{i=1}^{L} s_i g_i \,. \tag{1.52}$$

In this case, correlations between sites arise purely from the dynamics in Eq. (1.51), and will generally depend on the *distribution of fitness effects* (DFE),

$$U\rho(s) = \mu \sum_{i=1}^{L} \delta(s - s_i) \,. \tag{1.53}$$

The DFE will be a key empirical parameter that enters into our models.

In Chapter 2, we present a general framework for analyzing sequence evolution in Eqs. (1.51) and (1.52), which works by coarse-graining genotype frequencies into a single-dimensional *fitness distribution*,

$$f(X) = \sum_{\vec{g}} \delta[X - X(\vec{g})] f(\vec{g}) \,, \tag{1.54}$$

and analyzing the dynamics of this quantity directly. We investigate some formal properties of the model in the limit that the *net* selective pressures are small. This provides some intuition, but it has limited direct relevance for the rapidly evolving populations we wish to study.

We extend this approach to more realistic parameter regimes in Chapters 3-5. Chapters 3 and 4 focus on genetic divergence in rapidly adapting populations. In particular, we obtain predictions for the rate of accumulation of beneficial "driver" mutations (Chapter 3) and deleterious "passengers" that hitchhike to fixation (Chapter 4). In Chapter 5, we present a method for predicting genetic diversity in more well-adapted populations, where selection operates on a large number of weakly deleterious mutations with a significant net effect. This method works by coarse-graining the fitness distribution in Eq. (1.54), thereby approximating the effect of several weak mutations with a single strongly-selected mutation. We also introduce the concept of a "linkage block" (independently discovered by Neher et al. [253] and Weissman & Hallatschek [345]), which shows how the effectively asexual dynamics in Eq. (1.51) can emerge in a linear genome with a nonzero rate of recombination.

In Chapters 6 and 7, we apply the theoretical models in Chapters 3 and 4 to data from microbial evolution experiments. In Chapter 6, we present a method for inferring the DFE in experimental populations of *S. cerevisiae*, by tracking the frequencies of beneficial mutations in ~ 1000 "engineered" selective sweeps. In Chapter 7, we utilize a similar model-fitting approach to infer evolutionary parameters in a long-term experiment in *E. coli*, based on existing measurements of fitness [355] and mutation accumulation [349]. The unique length of this experiment also allows us to investigate departures from the additive fitness model in Eq. (1.52) due to epistasis. We analyze the empirical support for alternative fitness functions, $X(\vec{g})$, based on the "macroscopic" fitness and mutation trajectories.

In Chapters 8 and 9, we investigate the behavior of these simple models when some of the parameters

are allowed to evolve or change in time. Chapter 8 focuses on a scenario where the *local* DFE,

$$U\rho(s|\vec{g}) = \mu \sum_{i=1}^{L} \delta[s - (1 - 2g_i)s_i] \,, \tag{1.55}$$

evolves in time as the population moves throughout the space of genotypes. We investigate the behavior of this model on long timescales, when adaptation plateaus and the local DFE becomes stable under further evolution. This provides a connection between the rapidly adapting populations in Chapters 3, 4, 6, and 7 and the weak selection regimes in Chapters 2 and 5. Finally, in Chapter 9, we return to the single-site model in Eq. (1.4) and investigate its dynamics when the fitness effects fluctuate in time.

References

[1] Adams, J. & Oeller, P. (1986). Structure of evolving populations of *Saccharomyces cerevisiae*: adaptive changes are frequently associated with sequence alterations involving mobile elements belonging to the ty family. *Proc. Natl. Acad. Sci. USA*, 83, 7124–7127.

[2] Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2015). *Molecular Biology of the Cell*. New York, NY: Garland Science.

[3] Andersen, E. C., Gerke, J. P., Shapiro, J. A., Crissman, J. R., Ghosh, R., Bloom, J. S., Félix, M.-A., & Kruglyak, L. (2012). Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nature Genetics*, 44, 285–290.

[4] Atwood, K., Schneider, L., & Ryan, F. (1951). Periodic selection in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA*, 37, 146.

[5] Audi, G., Bersillon, O., Blachot, J., & Wapstra, A. (2003). The nubase evaluation of nuclear and decay properties. *Nuclear Physics A*, 729(1), 3–128.

[6] Bachtrog, D. & Gordo, I. (2004). Adaptive evolution of asexual populations under muller's ratchet. *Evolution*, 58, 1403–1413.

[7] Bank, C., Hietpas, R. T., Wong, A., Bolon, D. N., & Jensen, J. D. (2014). A bayesian mcmc approach to assess the complete distribution of fitness effects of new mutations: Uncovering the potential for adaptive walks in challenging environments. *Genetics*, 196, 841–852.

[8] Barraclough, T. G., Fontaneto, D., Ricci, C., & Herniou, E. A. (2007). Evidence for inefficient selection against deleterious mutations in cytochrome oxidase i of asexual bdelloid rotifers. *Mol. Biol. Evol.*, 24, 1952–1962.

[9] Barrett, R. D., MacLean, R. C., & Bell, G. (2006). Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas fluorescens* populations. *Biology Letters*, 2, 236–238.

[10] Barrick, J. E., Kauth, M. R., Strelioff, C. C., & Lenski, R. E. (2010). *Escherichia coli* rpob mutants have increased evolvability in proportion to their fitness defects. *Mol Biol Evol*, 27, 1338.

[11] Barrick, J. E. & Lenski, R. E. (2013). Genome dynamics during experimental evolution. *Nature Reviews Genetics*, 14.

[12] Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., & Kim, J. F. (2009). Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461, 1243–1247.

[13] Barroso-Batista, J., Sousa, A., Lourenço, M., Bergman, M.-L., Demengeot, J., Xavier, K. B., & Gordo, I. (2014). The first steps of adaptation of escherichia coli to the gut are dominated by soft sweeps. *arXiv*.

[14] Bartolomé, C. & Charlesworth, B. (2006). Evolution of amino-acid sequences and codon usage on the *Drosophila miranda* neo-sex chromosomes. *Genetics*, 174, 2033–2044.

[15] Barton, N. & Etheridge, A. M. (2004). The effect of selection on genealogies. *Genetics*, 166, 1115–1131.

[16] Barton, N. H. (1995). Linkage and the limits to natural selection. *Genetics*, 140, 821–841.

[17] Barton, N. H. & Etheridge, A. M. (2011). The relation between reproductive value and genetic contribution. *Genetics*, 188, 953–973.

[18] Barton, N. H. & Turelli, M. (1991). Natural and sexual selection at many loci. *Genetics*, 127, 229–255.

[19] Bataillon, T., Zhang, T., & Kassen, R. (2011). Cost of adaptation and fitness effects of beneficial mutations in *Pseudomonas fluorescens*. *Genetics*, 189, 939–949.

[20] Batorsky, R., Kearney, M. F., Palmer, S., Maldarelli, F., Rouzine, I., & Coffin, J. (2011). Estimate of effective recombination rate and average selection coefficient for hiv in chronic infection. *Proc Natl Acad Sci*, 108, 5661–5666.

[21] Begun, D. J. & Aquadro, C. F. (1992). Levels of naturally occuring dna polymorphism correlate with recombination rates in *D. melanogaster*. *Nature*, 356, 519–520.

[22] Begun, D. J., Holloway, A. K., Stevens, K., Hillier, L. W., Poh, Y.-P., & *et al* (2007). Population genomics: Whole-genome analysis of polymorphism and divergence in drosophila simulans. *PLoS Biol*, 5, e310.

[23] Bell, G. (2010). Experimental genomics of fitness in yeast. *Proc Biol Sci*, 277, 1459–1467.

[24] Bender, C. M. & Orszag, S. A. (1978). *Advanced Mathematical Methods for Scientists and Engineers*. New York: McGraw-Hill.

[25] Berg, J. & Lässig, M. (2003). Stochastic evolution of transcription factor binding sites. *Biophysics (Moscow)*, 48(Suppl. 1).

[26] Berg, J., Willmann, S., & "assig, M. L. (2004). Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.*, 4, 42.

[27] Bergland, A. O., Behrman, E. L., O'Brien, K. R., Schmidt, P. S., & Petrov, D. A. (2014). Genomic Evidence of Rapid and Stable Adaptive Oscillations over Seasonal Time Scales in Drosophila. *PLoS Genetics*, 10(11), e1004775.

[28] Betancourt, A. J., Welch, J. J., & Charlesworth, B. (2009). Reduced effectiveness of selection caused by a lack of recombination. *Current Biology*, 19, 655–660.

[29] Bhaskar, A., Wang, Y. X. R., & Song, Y. S. (2015). Efficient inference of population size histories and locus-specific mutation rates from large-sample genomic variation data. *Genome Research*, 25, 268–279.

[30] Birky, Jr., C. W. & Walsh, J. B. (1988). Effects of linkage on rates of molecular evolution. *Proc Natl Acad Sci*, 85, 6414–6418.

[31] Blatt, J. M. & Weisskopf, V. F. (1979). *Theoretical Nuclear Physics*. New York, NY: Springer-Verlag.

[32] Blount, Z. D., Borland, C. Z., & Lenski, R. E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc Natl Acad Sci*, 105, 7899–7906.

[33] Bolthausen, E. & Sznitman, A.-S. (1998). On ruelle's probability cascades and an abstract cavity method. *Communications in Mathematical Physics*, 197, 247–276.

[34] Boyko, A. R., Williamson, S. H., Indap, A. R., Degenhardt, J. D., & R D Hernandez, e. (2008). Assessing the evolutionary impact of amino acid mutations in the human genome. *PLoS Genet*, 4, e1000083.

[35] Brown, L. D., Cai, T. T., & DasGupta, A. (2001). Interval estimation for a binomial proportion. *Stat. Sci.*, 16, 101–117.

[36] Brunet, E., Derrida, B., Mueller, A. H., & Munier, S. (2006). A phenomenological theory giving the full statistics of the position of fluctuating pulled fronts. *Phys Rev E*, 73, 056126.

[37] Brunet, E., Rouzine, I. M., & Wilke, C. O. (2008). The stochastic edge in adaptive evolution. *Genetics*, 179, 603–620.

[38] Burch, C. L. & Chao, L. (2000). Evolvability of an rna virus is determined by its mutational neighbourhood. *Nature*, 406, 625–628.

[39] Burch, C. L., Guyader, S., Samarov, D., & Shen, H. (2007). Experimental estimate of the abundance and effects of nearly neutral mutations in the rna virus φ6. *Genetics*, 176, 467–476.

[40] Bürger, R. (1991). Moments, cumulants, and polygenic dynamics. *J Math Biol*, 30, 199–213.

[41] Bürger, R. (1998). Mathematical properties of mutation-selection models. *Genetica*, 102/103, 279–298.

[42] Bustamante, C. D., Nielsen, R., Sawyer, S. A., Olsen, K. M., Purugganan, M. D., & Hartl, D. L. (2002). The cost of inbreeding in *Arabidopsis*. *Nature*, 416, 531–534.

[43] Bustamante, C. D., Wakeley, J., Sawyer, S., & Hartl, D. L. (2001). Directional selection and the site-frequency spectrum. *Genetics*, 159, 1779–1788.

[44] Capocelli, R. M. & Ricciardi, L. M. (1974). A diffusion model for population growth in random environment. *Theoretical population biology*, 5(1), 28–41.

[45] Charlesworth, B. (1994). The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research*, 63, 213–227.

[46] Charlesworth, B. (2009). Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.*, 10, 195–205.

[47] Charlesworth, B., Morgan, M. T., & Charlesworth, D. (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics*, 134, 1289–1303.

[48] Charlesworth, D., Charlesworth, B., & Morgan, M. T. (1995). The pattern of neutral molecular variation under the background selection model. *Genetics*, 141, 1619–1632.

[49] Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D., & Marx, C. J. (2011). Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*, 332, 1190–1192.

[50] Cohen, E., Kessler, D. A., & Levine, H. (2005). Front propagation up a reaction rate gradient. *Phys Rev E*, 72, 066126.

[51] Colegrave, N. (2002). Sex releases the speed limit on evolution. *Nature*, 420, 664–666.

[52] Comeron, J. M. & Kreitman, M. (2002). Population, evolutionary, and genomic consequences of interference selection. *Genetics*, 161, 389–410.

[53] Consortium, M. G. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420, 520–562.

[54] Consortium, T. . G. P. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature*, 491, 56–65.

[55] Cook, R. D. & Hartl, D. L. (1974). Uncorrelated random environments and their effects on gene frequency. *Evolution*, 28(2), 265–274.

[56] Cook, R. D. & Hartl, D. L. (1975). Stochastic selection in large and small populations. *Theoretical Population Biology*, 7(1), 55–63.

[57] Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., & *et al* (2010). The genetic landscape of a cell. *Science*, 327, 425–431.

[58] Covert, III, A. W., Lenski, R. E., Wilke, C. O., & Ofria, C. (2013). Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *PNAS*, 110, E3171–E3178.

[59] Cowperthwaite, M. C., Bull, J., & Meyers, L. A. (2005). Distributions of beneficial fitness effects in rna. *Genetics*, 170, 1149–1457.

[60] Cutter, A. D. & Payseur, B. A. (2003). Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol Biol Evol*, 20, 665–673.

[61] Cutter, A. D. & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Gen*, 14, 262–274.

[62] de Sousa, J. A. M., Campos, P. R., & Gordo, I. (2013). An abc method for estimating the rate and distribution of effects of beneficial mutations. *Genome Biol. Evol.*, 5, 794–806.

[63] de Visser, J. A. G. & Krug, J. (2014). Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*, 15, 480–490.

[64] de Visser, J. A. G. M. & Rozen, D. E. (2006). Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics*, 172, 2093–2100.

[65] de Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L., & Lenski, R. E. (1999). Diminishing returns from mutation supply rate in asexual populations. *Science*, 283, 404–406.

[66] de Vos, M. G. J., Poelwijk, F. J., Battich, N., Ndika, J. D. T., & Tans, S. J. (2013). Environmental dependence of genetic constraint. *PLoS Genet*, 9, e1003580.

[67] Dempster, E. R. (1955). Maintenance of genetic heterogeneity. *Cold Spring Harb Quant Biol*, 20, 25–31.

[68] Der, R., Epstein, C., & Plotkin, J. B. (2012). Dynamics of neutral and selected alleles when the offspring distribution is skewed. *Genetics*, 191(4), 1331–1344.

[69] Der, R., Epstein, C. L., & Plotkin, J. B. (2011). Generalized population models and the nature of genetic drift. *Theoretical Population Biology*, 80, 80–99.

[70] Desai, M. M. & Fisher, D. S. (2007). Beneficial mutation selection balance and the effect of genetic linkage on positive selection. *Genetics*, 176, 1759–1798.

[71] Desai, M. M. & Fisher, D. S. (2011). The balance between mutators and nonmutators in asexual populations. *Genetics*, 188, 997–1014.

[72] Desai, M. M., Fisher, D. S., & Murray, A. W. (2007). The speed of evolution and the maintenance of variation in asexual populations. *Curr. Biol.*, 17, 385–394.

[73] Desai, M. M., Walczak, A. M., & Fisher, D. S. (2013). Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics*, 193, 565–585.

[74] Draghi, J. & Plotkin, J. B. (2013). Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution*, 67, 3120–3131.

[75] Drlica, K. (2003). The mutant selection window and antimicrobial resistance. *Journal of Antimicrobial Chemotherapy*, 52(1), 11–17.

[76] Dyke, M. V. (1974). Analysis and improvement of perturbation series. *Q J Mechanics Appl Math*, 27, 423–450.

[77] Eames, M. & Kortemme, T. (2012). Cost-benefit tradeoffs in engineered lac operons. *Science*, 336(6083), 911–915.

[78] Eigen, M. & Schuster, P. (1979). *The Hypercycle: A Principle of Natural Self-Organization*. Berlin, Germany: Springer-Verlag.

[79] Elena, S. F., Ekunwe, L., Hajela, N., Oden, S. A., & Lenski, R. E. (1998). Distribution of fitness effects caused by random insertion mutations in escherichia coli. *Genetica*, 102/103, 349–358.

[80] Elena, S. F. & Lenski, R. E. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.*, 4, 457–469.

[81] Etheridge, A., Pfaffelhuber, P., & Wakolbinger, A. (2007). How often does the ratchet click? facts, heuristics, and asymptotics. In P. M. J. Blath & M. Scheutzow (Eds.), *Trends in Stochastic Analysis* (pp. 365–390). Cambridge University Press.

[82] Ethier, S. N. & Kurtz, T. G. (1987). The infinitely-many-alleles model with selection as a measure valued diffusion. *Lecture Notes in Biomathematics*, 70, 72–87.

[83] Ewens, W. J. (1963). The diffusion equation and a pseudo-distribution in genetics. *Journal of the Royal Statistical Society. Series B*, 25, 405–412.

[84] Ewens, W. J. (2004). *Mathematical Population Genetics*. New York: Springer-Verlag, second edition.

[85] Ewens, W. J. (2008). Commentary: On haldane's 'defense of beanbag genetics'. *Int. J. Epidemiol.*, 37, 447–451.

[86] Excoffier, L. & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends Ecol Evol*, 23, 347–351.

[87] Eyre-Walker, A. (2010). Genetic architecture of a complex trait and its implications for fitness and genome-wide association studies. *Proc Natl Acad Sci*, 107, 1752–1756.

[88] Eyre-Walker, A. & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nat Rev Gen*, 8, 610–618.

[89] Falconer, D. S. (1960). *Introduction to Quantitative Genetics*. Edinburgh/London: Oliver & Boyd.

[90] Feller, W. (1951). Diffusion processes in genetics. In *Proc. Second Berkeley Symp. on Math. Statist. and Prob.* (pp. 227–256).: Univ. of Calif. Press.

[91] Fisher, D. S. (2007). Evolutionary dynamics. In M. M. Jean-Philippe Bouchaud & J. Dalibard (Eds.), *Complex Systems*, volume 85 of *Les Houches* (pp. 395–446). Elsevier.

[92] Fisher, D. S. (2013). Asexual evolution waves: fluctuations and universality. *J Stat Mech*, 2013, P01011.

[93] Fisher, R. A. (1930). The distribution of gene ratios for rare mutations. *Proc Roy Soc Edinburgh*, 50, 204–219.

[94] Fogle, C. A., Nagle, J. L., & Desai, M. M. (2008). Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics*, 180, 2163–2173.

[95] Frank, S. A. (2014). Generative models versus underlying symmetries to explain biological pattern. *Journal of Evolutionary Biology*, 27, 1172–1178.

[96] Franklin, I. & Lewontin, R. C. (1970). Is the gene the unit of selection? *Genetics*, 65, 707–734.

[97] Frenkel, E. M., Good, B. H., & Desai, M. M. (2014). The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. *Genetics*, 196, 1217–1226.

[98] Gabriel, W., Lynch, M., & Burger, R. (1993). Muller's ratchet and mutational meltdowns. *Evolution*, 47, 1744–1757.

[99] Gardiner, C. (1985). *Handbook of Stochastic Methods*. New York: Springer.

[100] Gerland, U. & Hwa, T. (2009). Evolutionary selection between alternative modes of gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 106(22), 8841–8846.

[101] Gerrish, P. & Lenski, R. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica*, 127, 127–144.

[102] Gessler, D. (1995). The constraints of finite size in asexual populations and the rate of the ratchet. *Genet. Res.*, 73, 119–131.

[103] Gifford, D. R. & MacLean, R. C. (2013). Evolutionary reversals of antibiotic resistance in experimental populations of *Pseudomonas aeruginosa*. *Evolution*, 67, 2973–2981.

[104] Gillespie, J. (1983). A simple stochastic gene substitution model. *Theoretical Population Biology*, 23, 202–215.

[105] Gillespie, J. (1984). Molecular evolution over the mutational landscape. *Evolution*, 38, 1116–1129.

[106] Gillespie, J. (1991). *The causes of molecular evolution*. New York: Oxford University Press.

[107] Gillespie, J. H. (1973). Natural selection with varying selection coefficients–a haploid model. *Genetical Research*, 21(2), 115–120.

[108] Gillespie, J. H. (2000). Genetic drift in an infinite population: The pseudohitchhiking model. *Genetics*, 155, 909–919.

[109] Gillespie, J. H. & Guess, H. A. (1978). The effects of environmental autocorrelations on the progress of selection in a random environment. *American Naturalist*, 112, 897–909.

[110] Gillespie, J. H. & Langley, C. H. (1974). A general model to account for enzyme variation in natural populations. *Genetics*, 76(4), 837–848.

[111] Good, B. H. & Desai, M. M. (2013). Fluctuations in fitness distributions and the effects of weak linked selection on sequence evolution. *Theor Pop Biol*, 85, 86–102.

[112] Good, B. H. & Desai, M. M. (2014). Deleterious passengers in adapting populations. *Genetics*, 198, 1183–1208.

[113] Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O., & Desai, M. M. (2012). Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl. Acad. Sci.*, 109, 4950–4955.

[114] Good, B. H., Walczak, A. M., Neher, R. A., & Desai, M. M. (2014). Genetic diversity in the interference selection limit. *PLoS Genetics*, 10, e1004222.

[115] Gordo, I. & Charlesworth, B. (2000). On the speed of muller's ratchet. *Genetics*, 156, 2137–2140.

[116] Gordo, I., Navarro, A., & Charlesworth, B. (2002). Muller's ratchet and the pattern of variation at a neutral locus. *Genetics*, 161, 835–848.

[117] Gould, S. J. (1989). *Wonderful life: the Burgess Shale and the nature of history*. New York: W W Norton.

[118] Goyal, S., Balick, D. J., Jerison, E. R., Neher, R. A., Shraiman, B. I., & Desai, M. M. (2012). Dynamic mutation-selection balance as an evolutionary attractor. *Genetics*, 191, 1309–1319.

[119] Griffiths, R. C., S, & Tavaré (1998). The age of a mutation in a general coalescent tree. *Stochastic Models*, 14, 273–295.

[120] Grossmann, T. I., Woolfit, M., & Eyre-Walker, A. (2011). Quantifying the variation in the effective population size within a genome. *Genetics*, 189, 1389.

[121] Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional snp data. *PLoS Genetics*, 5, e1000695.

[122] Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & *et al* (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3, e189.

[123] Hahn, M. W. (2008). Toward a selection theory of molecular evolution. *Evolution*, 62, 255–265.

[124] Haigh, J. (1978). The accumulation of deleterious genes in a population. *Theor Pop Biol*, 14, 251–267.

[125] Haldane, J. (1927). The mathematical theory of natural and artificial selection, part v: selection and mutation. *Proc. Camb. Philos. Soc.*, 23, 828–844.

[126] Haldane, J. & Jayakar, S. D. (1963). Polymorphism due to selection of varying direction. *Journal of Genetics*.

[127] Hallatschek, O. (2011). The noisy edge of traveling waves. *Proc. Natl. Acad. Sci. USA*, 108, 1783–1787.

[128] Hallatschek, O. & Korolev, K. S. (2009). Fisher waves in the strong noise limit. *Phys Rev Lett*, 103, 108103.

[129] Hallatschek, O. & Nelson, D. R. (2008). Gene surfing in expanding populations. *Theor Popul Biol*, 73, 158–170.

[130] Harris, H. (1966). Enzyme polymorphisms in man. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 164, 298–310.

[131] Harris, K. & Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. *PLoS Genet*, 9, e1003521.

[132] Hartfield, M. & Otto, S. P. (2011). Recombination and hitchhiking of deleterious alleles. *Evolution*, 65, 2421–2434.

[133] Hartl, D. & Clark, A. (2007). *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates.

[134] Hartl, D. L. & Cook, R. D. (1973). Balanced polymorphisms of quasineutral alleles. *Theoretical population biology*, 4(2), 163–172.

[135] Hartl, D. L. & Cook, R. D. (1974). Autocorrelated random environments and their effects on gene frequency. *Evolution*, 28(2), 275–280.

[136] Hedrick, P. W. (1974). Genetic Variation in a Heterogeneous Environment. I. Temporal Heterogeneity and the Absolute Dominance Model. *Genetics*, 78(2), 757–770.

[137] Hegreness, M., Shoresh, N., Hartl, D., & Kishony, R. (2006). An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science*, 311, 1615–1617.

[138] Helling, R., Kinney, T., & Adams, J. (1981). The maintenance of plasmid-containing organisms in populations of *Escherichia coli*. *J. Gen. Microbiol.*, 123, 129–141.

[139] Higgs, P. & Woodcock, G. (1995). The accumulation of mutations in asexual populations and the structure of genealogical trees in the presence of selection. *J Math Biol*, 33, 677–702.

[140] Hill, W. G. & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genet. Res.*, 8, 269–294.

[141] Hinch, E. J. (1991). *Perturbation Methods*. New York: Cambridge University Press.

[142] Houchmandzadeh, B. & Vallade, M. (2010). Alternative to the diffusion equation in population genetics. *Phys Rev E*, 82, 051913.

[143] Hudson, R. R. (2002). Generating samples under a wright-fisher neutral model. *Bioinformatics*, 18, 337–338.

[144] Hudson, R. R. & Kaplan, N. L. (1988). The coalescent process in models with selection and recombination. *Genetics*, 120, 831–840.

[145] Hudson, R. R. & Kaplan, N. L. (1994). Gene trees with background selection. In B. Golding (Ed.), *Non-Neutral Evolution: Theories and Molecular Data* (pp. 140–153). London: Chapman & Hall.

[146] Hudson, R. R. & Kaplan, N. L. (1995). Deleterious background selection with recombination. *Genetics*, 141, 1605–1617.

[147] Illingworth, C. & Mustonen, V. (2012). A method to infer positive selection from marker dynamics in an asexual population. *Bioinformatics*, 28, 831–837.

[148] Imhof, M. & Schlötterer, C. (2001). Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. *Proc. Natl. Acad. Sci. USA*, 98, 1113–1117.

[149] Jain, K. (2008). Loss of least-loaded class in asexual populations due to drift and epistasis. *Genetics*, 179, 2125–2134.

[150] Jansen, M. L., Diderich, J. A., Mashego, M., Hassane, A., de Winde, J. H., & *et al* (2005). Prolonged selection in aerobic, glucose-limited chemostat cultures of *Saccharomyces cerevisiae* causes a partial loss of glycolytic capacity. *Microbiology*, 151, 1657–1669.

[151] Jasnos, L. & Korona, R. (2007). Epistatic buffering of fitness loss in yeast double deletion strains. *Nature Genetics*, 39, 550–554.

[152] Jensen, L. (1973). Random selective advantages of genes and their probabilities of fixation. *Genetics Research*, 21, 215–219.

[153] Jensen, L. & Pollak, E. (1969). Random selective advantages of a gene in a finite population. *Journal of Applied Probability*, 6(1), 19–37.

[154] Jiang, X., Xu, Z., Li, J., Shi, Y., Wu, W., & Tao, S. (2011). The influence of deleterious mutations on adaptation in asexual populations. *PLoS ONE*, 6, e27757.

[155] Johnson, T. (1999). The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *Proc. R. Soc. Lond. Ser. B*, 266, 2389–2397.

[156] Johnson, T. & Barton, N. H. (2002). The effect of deleterious alleles on adaptation in asexual populations. *Genetics*, 162, 395–411.

[157] Jones, E., Oliphant, T., Peterson, P., et al. (2001–). SciPy: Open source scientific tools for Python.

[158] Joseph, S. B. & Hall, D. W. (2004). Spontaneous mutations in diploid *Saccharomyces cerevisiae* more beneficial than expected. *Genetics*, 168, 1817–1825.

[159] Joyce, P., Rokyta, D., Beisel, C., & Orr, H. (2008). A general extreme value theory model for the adaptation of dna sequences under strong selection and weak mutation. *Genetics*, 180, 1627–1643.

[160] Judson, H. F. (1996). *The Eighth Day of Creation: Makers of the Revolution in Biology*. New York: Cold Spring Harbor Laboratory Press.

[161] Kaiser, V. B. & Charlesworth, B. (2008). The effects of deleterious mutations on evolution in non-recombining genomes. *Trends in Genetics*, 25, 9–12.

[162] Kao, K. C. & Sherlock, G. (2008). Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nature Genetics*, 40, 1499–1504.

[163] Kaplan, N., Darden, T., & Hudson, R. (1988). The coalescent process in models with selection. *Genetics*, 1988, 819–829.

[164] Karlin, S. & Lieberman, U. (1974). Random temporal variation in selection intensities: case of large population size. *Theoretical population biology*, 6(3), 355–382.

[165] Kassen, R. & Bataillon, T. (2006). Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nature Genetics*, 38, 484–488.

[166] Kauffman, S. & Levin, S. (1987). Towards a general theory of adaptive walks on rugged landscapes. *J Theor Biol*, 128, 11–45.

[167] Kauffman, S. & Weinberger, E. D. (1989). The nk model of rugged fitness landscape and its application to maturation of the immune response. *J Theor Biol*, 141, 211–245.

[168] Kawecki, T. J., Lenski, R. E., Ebert, D., Holis, B., Olivieri, I., & Whitlock, M. C. (2012). Experimental evolution. *Trends in Ecology and Evolution*, 27, 547–560.

[169] Keightley, P. D. & Eyre-Walker, A. (2007). Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics*, 177, 2251–2261.

[170] Keightley, P. D. & Eyre-Walker, A. (2010). What can we learn about the distribution of fitness effects of new mutations from dna sequence data? *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 365, 1187–1193.

[171] Kendall, D. G. (1948). On the generalized "birth-and-death" process. *The annals of mathematical statistics*, 19(1), 1–15.

[172] Khan, A. I., Din, D. M., Schneider, D., Lenski, R. E., & Cooper, T. F. (2011). Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332, 1193–1196.

[173] Kim, Y. & Orr, H. (2005). Adaptation in sexuals vs. asexuals: Clonal interference and the fisher-muller model. *Genetics*, 171, 1377–1386.

[174] Kimura, M. (1954). Process leading to quasi-fixation of genes in natural populations due to random fluctuation of selection intensities. *Genetics*, 39(3), 280.

[175] Kimura, M. (1955). Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb Symp Quant Biol*, 20, 33–53.

[176] Kimura, M. (1962). On the Probability of Fixation of Mutant Genes in a Population . *Genetics*, 47(6), 713–719.

[177] Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217, 624–626.

[178] Kimura, M. & Maruyama, T. (1966). The mutational load with epistatic gene interations in fitness. *Genetics*, 54, 1303–1312.

[179] King, J. L. & Jukes, T. H. (1969). Non-darwinian evolution. *Science*, 164, 788–798.

[180] Kingman, J. F. C. (1976). Coherent random walks arising in some genetical models. *Proc R Soc Lond Ser A*, 351, 19–31.

[181] Kingman, J. F. C. (1978). A simple model for the balance between selection and mutation. *J Appl Prob*, 15, 1–12.

[182] Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes Appl*, 13, 235–248.

[183] Kingman, J. F. C. (2000). Origins of the coalescent: 1974-1982. *Genetics*, 156, 1461–1463.

[184] Kinney, J. B., Murugana, A., Callan, Jr, C. G., & Cox, E. C. (2010). Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc Natl Acad Sci*, 107, 9158–9163.

[185] Kong, A., Frigge, M. L., Masson, G., Besenbacher, S., & *et al* (2012). Rate of *de novo* mutations and the importance of father's age to disease risk. *Nature*, 488, 471–475.

[186] Kong, A., Thorleifsson, G., Gudbjartsson, D. F., Masson, G., & *et al* (2010). Fine-scale recombination rate differences between sexes, populations and individuals. *Nature*, 467, 1099–1103.

[187] Korolev, K. S., Avlund, M., Hallatschek, O., & Nelson, D. R. (2010). Genetic demixing and evolution in linear stepping stone models. *Rev Mod Phys*, 82, 1691–1718.

[188] Kosheleva, K. & Desai, M. M. (2013). The dynamics of genetic draft in rapidly adapting populations. *Genetics*, 195, 1007–1025.

[189] Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase locus of drosophila melanogaster. *Nature*, 304, 412–417.

[190] Kryazhimskiy, S., Tkačik, G., & Plotkin, J. B. (2009). The dynamics of adaptation on correlated fitness landscapes. *Proc Natl Acad Sci USA*, 106, 18638–18643.

[191] Kryazhimskiy, S. K., Rice, D. P., & Desai, M. M. (2012). Population subdivision and adaptation in asexual populations of saccharomyces cerevisiae. *Evolution*, 66, 1931–1941.

[192] Kryazhimskiy, S. K., Rice, D. P., Jerison, E. R., & Desai, M. M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344, 1519–1522.

[193] Kussell, E. & Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, 309(5743), 2075–2078.

[194] Kvitek, D. J. & Sherlock, G. (2013). Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genetics*, 9, e1003972.

[195] Lahti, D. C., Johnson, N. A., Ajie, B. C., Otto, S. P., Hendry, A. P., Blumstein, D. T., Coss, R. G., Donohue, K., & Foster, S. A. (2009). Relaxed selection in the wild. *Trends in Ecology & Evolution*, 24(9), 487–496.

[196] Lamb, T. D., Collin, S. P., & Pugh, Jr, E. N. (2007). Evolution of the vertebrate eye: opsins, photoreceptors, retina and eye cup. *Nat Rev Neurosci*, 8, 960–976.

[197] Lang, G. I., Botstein, D., & Desai, M. M. (2011). Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics*, 188, 647–661.

[198] Lang, G. I. & Murray, A. W. (2008). Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. *Genetics*, 178, 67–82.

[199] Lang, G. I., Rice, D. P., Hickman, M. J., Sodergren, E., Weinstock, G. M., Botstein, D., & Desai, M. M. (2013). Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500, 571–574.

[200] Lee, H., Popodi, E., Tanga, H., & Foster, P. L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci*, 109, E2774–E2783.

[201] Lee, M.-C. & Marx, C. J. (2013). Synchronous waves of failed soft sweeps in the laboratory: remarkably rampant clonal interference of alleles at a single locus. *Genetics*, 193, 943–952.

[202] Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., & *et al* (2012). Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biol*, 10, e1001388.

[203] Leibler, S. & Kussell, E. (2010). Individual histories and selection in heterogeneous populations. *Proceedings of the National Academy of Sciences*.

[204] Leiby, N. & Marx, C. J. (2014). Metabolic erosion primarily through mutation accumulation, and not tradeoffs, drives limited evolution of substrate specificity in *Escherichia coli*. *PLoS biology*, 12, e1001789.

[205] Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-term experimental evolution in escherichia coli. i. adaptation and divergence during 2,000 generations. *The American Naturalist*, 138, 1315–1341.

[206] Lenski, R. E. & Travisano, M. (1994). Dynamics of adaptation and diversification: A 10,000-generation experiment with bacterial populations. *Proc Natl Acad Sci USA*, 91, 6808–6814.

[207] Levikson, B. & Karlin, S. (1975). Random temporal variation in selection intensities acting on infinite diploid populations: diffusion method analysis. *Theoretical population biology*, 8(3), 292–300.

[208] Levins, R. (1969). The effect of random variations of different types on population growth. *Proceedings of the National Academy of Sciences*, 62(4), 1061–1065.

[209] Lewin, R. (1983). How did vertebrates take to the air? *Science*, 221, 38–39.

[210] Lewontin, R. C. & Cohen, D. (1969). On population growth in a randomly varying environment. *Proceedings of the National Academy of Sciences*, 62(4), 1056–1060.

[211] Lewontin, R. C. & Hubby, J. L. (1966). A molecular approach to the study of genic heterozygosity in natural populations. ii. amount of variation and degree of heterozygosity in natural populations of drosophila pseudoobscura. *Genetics*, 54, 595–609.

[212] Li, H. & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, 475, 493–496.

[213] Lieberman, T. D., Flett, K. B., Yelin, I., Martin, T. R., McAdam, A. J., Priebe, G. P., & Kishony, R. (2014). Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nature Genetics*, 46, 82–87.

[214] Loewe, L. & Charlesworth, B. (2006). Inferring the distribution of mutational effects on fitness in drosophila. *Biol Lett*, 2, 426–430.

[215] Lohmueller, K. E., Albrechtsen, A., Li, Y., Kim, S. Y., Korneliussen, T., et al. (2011). Natural selection affects multiple aspects of genetic variation at putatively neutral sites across the human genome. *PLoS Genetics*, 7, e1002326.

[216] Lohmueller, K. E., Indap, A. R., Schmidt, S., Boyko, A. R., Hernandez, R. D., & *et al* (2008). Proportionally more deleterious genetic variation in european than in african populations. *Nature*, 451, 994–997.

[217] Łuksza, M. & Lässig, M. (2014). A predictive fitness model for influenza. *Nature*, 507, 57–61.

[218] Lynch, M. (2011). The lower bound to the evolution of mutation rates. *Genome Biol. Evol.*, 3, 1107–1118.

[219] Lynch, M. (2012). Evolutionary layering and the limits to cellular perfection. *Proc Natl Acad Sci USA*, 109, 18851–18856.

[220] Lynch, M., Sung, W., Morris, K., Coffey, N., Landry, C. R., & *et al* (2008). A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc Natl Acad Sci*, 105, 9272–9277.

[221] MacLean, R. & Buckling, A. (2009). The distribution of fitness effects of beneficial mutations in *Pseudomonas aeruginosa*. *PLoS Genet.*, 5, e1000406.

[222] MacLean, R. C., Perron, G. G., & Gardner, A. (2010). Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics*, 186, 1345–1354.

[223] Maharjan, R. P., Liu, B., Li, Y., Reeves, P. R., Wang, L., & Ferenci, T. (2013). Mutation accumulation and fitness in mutator subpopulations of escherichia coli. *Biol. Lett.*, 9, 20120961.

[224] Masel, J., King, O. D., & Maughan, H. (2007). The loss of adaptive plasticity during long periods of environmental stasis. *The American Naturalist*, 169(1), 38–46.

[225] Matic, I., Radman, M., Taddei, F., Picard, B., Doit, C., Bingen, E., Denamur, E., & Elion, J. (1997). Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science*, 277, 1833–1834.

[226] Mayr, E. (1959). Where are we? *Cold Spring Harbor Symp Quant Biol.*, 24, 1–14.

[227] McCandlish, D., Epstein, C., & Plotkin, J. B. (2014). The inevitability of unconditionally deleterious substitutions during adaptation. *Evolution*, 68, 1351–1365.

[228] McCandlish, D. M. & Stolzfus, A. (2014). Modeling evolution using the probability of fixation: history and implications. *Quarterly Review of Biology*, 89, 225–252.

[229] McCandlish, M., Otwinowski, J., & Plotkin, J. B. (2015). Detecting epistasis from an ensemble of adapting populations. *Evolution*, 69, 2359–2380.

[230] McDonald, M., Cooper, T., Beaumont, H., & Rainey, P. (2011). The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. *Biol. Lett.*, 7, 98–100.

[231] McDonald, M. J., Hsieh, Y. Y., Yu, Y. H., Chang, S. L., & Leu, J. Y. (2012). The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. *Current Biology*, 22, 1235–1240.

[232] McFarland, C., Mirny, L., & Korolev, K. S. (2014). A tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *arXiv*.

[233] McFarland, C. D., Korolev, K. S., Kryukovc, G. V., Sunyaeva, S. R., & Mirny, L. A. (2013). Impact of deleterious passenger mutations on cancer progression. *PNAS*, 110, 2910–2915.

[234] McVean, G. A. T. & Charlesworth, B. (2000). The effects of hill-robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics*, 155, 929–944.

[235] McVicker, G., Gordon, D., Davis, C., & Green, P. (2009). Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet*, 5, e1000471.

[236] Messer, P. W. & Petrov, D. A. (2013). Frequent adaptation and the mcdonald-kreitman test. *Proc. Natl. Acad. Sci.*, 110, 8615–8620.

[237] Miller, C. R., Joyce, P., & Wichman, H. A. (2011). Mutational effects and population dynamics during viral adaptation challenge current models. *Genetics*, 187, 185–202.

[238] Miralles, R., Gerrish, P. J., Moya, A., & Elena, S. F. (1999). Clonal interference and the evolution of rna viruses. *Science*, 285, 1745–1747.

[239] Moran, P. A. P. (1975). Wandering distributions and the electrophoretic profile. *Theor Pop Biol*, 8, 318–330.

[240] Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutat Res*, 1, 2–9.

[241] Munroe, R. (2010). Physicists. https://xkcd.com/793/. Accessed: 2015-04-24.

[242] Mustonen, V. & Lässig, M. (2007). Adaptations to fluctuating selection in Drosophila. *Proceedings of the National Academy of Sciences*, 104(7), 2277–2282.

[243] Mustonen, V. & Lässig, M. (2008). Molecular evolution under fitness fluctuations. *Phys Rev Lett*, 100, 108101.

[244] Mustonen, V. & Lässig, M. (2009). From fitness landscapes to seascapes: non-equilibrium dynamics of selection and adaptation. *Trends Genet*, 25, 111–119.

[245] Mustonen, V. & Lässig, M. (2010). Fitness flux and ubiquity of adaptive evolution. *Proc Natl Acad Sci*, 107, 4248–4253.

[246] Nagel, A. C., Joyce, P., Wichman, H. A., & Miller, C. R. (2012). Stickbreaking: A novel fitness landscape model that harbors epistasis and is consistent with commonly observed patterns of adaptive evolution. *Genetics*, 190, 655–667.

[247] Nagylaki, T. (1993). The evolution of multilocus systems under weak selection. *Genetics*, 134, 627–647.

[248] Neher, R. & Shraiman, B. I. (2011a). Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics*, 188, 975–976.

[249] Neher, R. & Shraiman, B. I. (2012). Fluctuations of fitness distributions and the rate of muller's ratchet. *Genetics*, 191, 1283–1293.

[250] Neher, R., Shraiman, B. I., & Fisher, D. S. (2010). Rate of adaptation in large sexual populations. *Genetics*, 184, 467–481.

[251] Neher, R. A. (2013). Genetic draft, selective interference, and population genetics of rapid adaptation. *Annual Review of Ecology, Evolution, and Systematics*, 44, 195–215.

[252] Neher, R. A. & Hallatschek, O. (2013). Genealogies in rapidly adapting populations. *Proc Nat Acad Sci*, 110, 437–442.

[253] Neher, R. A., Kessinger, T. A., & Shraiman, B. I. (2013). Coalescence and genetic diversity in sexual populations under selection. *Proc Nat Acad Sci*, 110, 15836–15841.

[254] Neher, R. A. & Shraiman, B. I. (2011b). Statistical genetics and evolution of quantitative traits. *Rev Mod Phys*, 83, 1283–1300.

[255] Nelson, M. R., Wegmann, D., Ehm, M. G., Kessner, D., & *et al* (2012). An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science*, 337, 100–104.

[256] Neuhauser, C. & Krone, S. M. (1997). The genealogy of samples in models with selection. *Genetics*, 145, 519–534.

[257] Nicolaisen, L. E. & Desai, M. M. (2012). Distortions in genealogies due to purifying selection. *Molecular Biology and Evolution*, 29, 3589–3600.

[258] Nicolaisen, L. E. & Desai, M. M. (2013). Distortions in genealogies due to purifying selection and recombination. *Genetics*, 195, 221–230.

[259] Nielsen, R. (2001). Statistical tests of selective neutrality in the age of genomics. *Heredity*, 86, 641–647.

[260] Nik-Zainal, S., Loo, P. V., Wedge, D. C., Alexandrov, L. B., Greenman, C. D., Lau, K. W., Raine, K., Jones, D., Marshall, J., Ramakrishna, M., Shlien, A., Cooke, S. L., Hinton, J., Menzies, A., Stebbings, L. A., Leroy, C., Jia, M., Rance, R., Mudie, L. J., Gamble, S. J., Stephens, P. J., McLaren, S., Tarpey, P. S., Papaemmanuil, E., Davies, H. R., Varela, I., McBride, D. J., Bignell, G. R., Leung, K., Butler, A. P., Teague, J. W., Martin, S., Joönsson, G., Mariani, O., Boyault, S., Miron, P., Fatima, A., Langerød, A., Aparicio, S. A., Tutt, A., Sieuwerts, A. M., Borg, A., Thomas, G., Salomon, A. V., Richardson, A. L., Børresen-Dale, A.-L., Futreal, P. A., Stratton, M. R., Campbell, P. J., & of the International Cancer Genome Consortium, B. C. W. G. (2012). The life history of 21 breat cancers. *Cell*, 149, 994–1007.

[261] Nirenberg, M., Leder, P., Bernfield, M., Brimacombe, R., Trupin, J., Rottman, F., & O'Neal, C. (1965). Rna codewords and protein synthesis, vii. on the general nature of the rna code. *Proc Natl Acad Sci USA*, 53, 1161–1168.

[262] Nordborg, M., Charlesworth, B., & Charlesworth, D. (1996). The effect of recombination on background selection. *Genet Res*, 67, 159–174.

[263] Novick, A. & Szilard, L. (1950). Experiments with the chemostat on spontaneous mutations of bacteria. *Proc. Natl. Acad. Sci. USA*, 36, 708.

[264] O'Fallon, B. (2013). Purifying selection causes widespread distortions of genealogical structure on the human x chromosome. *Genetics*, 194, 485–492.

[265] O'Fallon, B. D., Seger, J., & Adler, F. R. (2010). A continuous-state coalescent and the impact of weak selection on the structure of genealogies. *Mol Biol Evol*, 27, 1162–1172.

[266] Ohta, T. (1972). Fixation probability of a mutant influenced by random fluctuation of selection intensity. *Genetical research*, 19(1), 33–38.

[267] Ohta, T. (1973). Slightly deleterious mutant substitutions in evolution. *Nature*, 246, 96–98.

[268] Ohta, T. (1992). The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst*, 23, 263–286.

[269] Ohta, T. & Kimura, M. (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet Res*, 22, 201–204.

[270] Orr, H. (2002). The population genetics of adaptation: the adaptation of dna sequences. *Evolution*, 56, 1317–1330.

[271] Orr, H. (2003). The distribution of fitness effects among beneficial mutations. *Genetics*, 163, 1519–1526.

[272] Orr, H. A. (2000). The rate of adaptation in asexuals. *Genetics*, 155, 961–968.

[273] Orr, H. A. (2010). The population genetics of beneficial mutations. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 365, 1195–1201.

[274] Ostrow, S. L., Barshir, R., DeGregori, J., Yeger-Lotem, E., & Hershberg, R. (2014). Cancer evolution is associated with pervasive positive selection on globally expressed genes. *PLoS Genet*, 10, e1004239.

[275] P D Keightley, S. P. O. (2006). Interference among deleterious mutations favours sex and recombination in finite populations. *Nature*, 443, 89–92.

[276] Palamara, P. F., Lencz, T., Darvasi, A., & Pe'er, I. (2012). Length distributions of identity by descent reveal fine-scale demographic history. *Am J Hum Genet*, 91, 809–822.

[277] Paquin, C. & Adams, J. (1983). Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature*, 302, 495–500.

[278] Park, S. & Krug, J. (2007). Clonal interference in large populations. *Proc. Natl. Acad. Sci. USA*, 104, 18135–18140.

[279] Park, S., Simon, D., & Krug, J. (2010). The speed of evolution in large asexual populations. *J. Stat. Phys.*, 138, 381–410.

[280] Peck, J. R. (1994). A ruby in the rubbish. *Genetics*, 137, 597–606.

[281] Pepin, K. M. & Wichman, H. A. (2008). Experimental evolution and genome sequencing reveal variation in levels of clonal interference in large populations of bacteriopage $\varphi x174$. *BMC Evol. Biol.*, 8, 85–98.

[282] Perfeito, L., Fernandes, L., Mota, C., & Gordo, I. (2007). Adaptive mutations in bacteria: High rate and small effects. *Science*, 317, 813.

[283] Perfeito, L., Sousa, A., Bataillon, T., & Gordo, I. (2014). Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution*, 68, 150–162.

[284] Pinkel, D. (2007). Analytical description of mutational effects in competing asexual populations. *Genetics*, 177, 2135–2149.

[285] Pool, J. E., Hellmann, I., Jensen, J. D., & Nielson, R. (2010). Population genetic inference from geomic sequence variation. *Genome Res*, 20, 291–300.

[286] Provine, W. B. (2001). *The origins of theoretical population genetics*. Chicago, IL: Chicago University Press.

[287] Przeworski, M., Charlesworth, B., & Wall, J. D. (1999). Genealogies and weak purifying selection. *Mol Biol Evol*, 16, 246–252.

[288] Pybus, O. G., Rambaut, A., Belshaw, R., Freckleton, R. P., Drummond, A. J., , & Holmes, E. C. (2007). Phylogenetic evidence for deleterious mutation load in rna viruses and its contribution to viral evolution. *Mol Biol Evol*, 24, 845–852.

[289] Qian, W., Ma, D., Xiao, C., Wang, Z., & Zhang, J. (2012). The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Reports*, 2, 1399–1410.

[290] Quandt, E. M., deatherage, D. E., Ellington, A. D., Georgiou, G., & Barrick, J. E. (2014). Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proc Natl Acad Sci*, 111, 2217–2222.

[291] Rambaut, A., Pybus, O. G., Nelson, M. I., Viboud, C., Taubenberger, J. K., & Holmes, E. C. (2008). The genomic and epidemiological dynamics of human influenza a virus. *Nature*, 453, 617–620.

[292] Robertson, A. (1961). Inbreeding in artificial selection programmes. *Genet Res*, 2, 189–194.

[293] Rödelsperger, C., Neher, R. A., Weller, A. M., Eberhardt, G., Witte, H., Mayer, W. E., Dieterich, & Sommer, R. (2014). Characterization of genetic diversity in the nematode *Pristionchus Pacificus* from population-scale resequencing data. *Genetics*, 196, 1153–1165.

[294] Rodríguez-Verdugo, A., Carrillo-Cisneros, D., González-González, A., Gaut, B. S., & Bennett, A. F. (2014). Different tradeoffs result from alternate genetic adaptations to a common environment. *Proceedings of the National Academy of Sciences*, 111(33), 12121–12126.

[295] Rokyta, D. R., Beisel, C. J., Joyce, P., Ferris, M. T., Burch, C. L., & Wichman, H. A. (2008). Beneficial fitness effects are not exponential for two viruses. *J. Mol. Evol.*, 67, 368–376.

[296] Rouzine, I., Brunet, E., & Wilke, C. (2008). The traveling-wave approach to asexual evolution: Muller's ratchet and the speed of adaptation. *Theor Pop Bio*, 73, 24–46.

[297] Rouzine, I. M. & Coffin, J. M. (2005). Evolution of hiv under selection and weak recombination. *Genetics*, 170, 7–18.

[298] Rouzine, I. M. & Coffin, J. M. (2010). Many-site adaptation in the presence of infrequent recombination. *Theor Pop Bio*, 77, 189–204.

[299] Rouzine, I. M., Wakeley, J., & Coffin, J. M. (2003). The solitary wave of asexual evolution. *Proc. Natl. Acad. Sci. USA*, 100, 587–592.

[300] Rozen, D. E., de Visser, J. A., & Gerrish, P. J. (2002). Fitness effects of beneficial mutations in microbial populations. *Curr Biol*, 12, 1040–1045.

[301] Rozen, D. E. & Lenski, R. E. (2000). Long-term experimental evolution in escherichia coli. viii. dynamics of a balanced polymorphism. *American Naturalist*, 155, 24–35.

[302] Sanjuán, R., Moya, A., & Elena, S. F. (2004). The distribution of fitness effects caused by single-nucleotide substitutions in an rna virus. *Proc. Natl. Acad. Sci. USA*, 101, 8396–8401.

[303] Santiago, E. & Caballero, A. (1998). Effective size and polymorphism of linked neutral loci in populations under directional selection. *Genetics*, 149, 2105–2117.

[304] Sargsyan, O. & Wakeley, J. (2008). A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organsisms. *Theoretical Popululation Biology*, 74, 104–114.

[305] Sawyer, S. A. & Hartl, D. L. (1992). Population genetics of polymorphism and divergence. *Genetics*, 132, 1161–1176.

[306] Schiffels, S., Szöllösi, G., Mustonen, V., & Lässig, M. (2011). Emergent neutrality in adaptive asexual evolution. *Genetics*, 189, 1361–1375.

[307] Schloissnig, S., Arumugam, M., Sunagawa, S., Mitreva, M., Tap, J., Zhu, A., Waller, A., Mende, D. R., Kultima, J. R., Martin, J., Kota, K., Sunyaev, S. R., Weinstock, G. M., & Bork, P. (2013). Genomic variation landscape of the human gut microbiome. *Nature*, 493, 45–50.

[308] Schoustra, S. E., Bataillon, T., Gifford, D. R., & Kassen, R. (2009). The properties of adaptive walks in evolving populations of fungus. *PLoS Biol*, 7, e1000250.

[309] Seger, J., Smith, W. A., Perry, J. J., Hunn, J., Kaliszewska, Z. A., et al. (2010). Gene geneologies strongly distorted by weakly interfering mutations in constant environments. *Genetics*, 184, 529–545.

[310] Segrè, D., Deluna, A., Church, G., & Kishony, R. (2005). Modular epistasis in yeast metabolism. *Nat Genet*, 37, 77–83.

[311] Sella, G. & Hirsh, A. E. (2005). The application of statistical physics to evolutionary biology. *Proc. Natl. Acad. Sci. USA*, 102, 9541–9546.

[312] Sella, G., Petrov, D. A., Przeworski, M., & Andolfatto, P. (2009). Pervasive natural selection in the *Drosophila* genome? *PLoS Genetics*, 5, e1000495.

[313] Sequencing, T. C. & Consortium, A. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 437, 69–87.

[314] Sheehan, S., Harris, K., & Song, Y. S. (2013). Estimating variable effective population sizes from multiple genomes: A sequentially markov conditional sampling distribution approach. *Genetics*, 194, 647–662.

[315] Sheldahl, L. A., Weinreich, D. M., & Rand, D. M. (2003). Recombination, dominance and selection on amino acid polymorphism in the drosophila genome: Contrasting patterns on the x and fourth chromosomes. *Genetics*, 165, 1195–1208.

[316] Sibani, P., Brandt, M., & Alstrøm, P. (1998). Evolution and extinction dynamics in rugged fitness landscapes. *Intl J Mod Phys*, 12, 361–391.

[317] Silander, O. K., Tenaillon, O., & Chao, L. (2007). Understanding the evolutionary fate of finite populations: The dynamics of mutational effects. *PLoS Biol*, 5, e94.

[318] Slatkin, M. (1972). On treating the chromosome as the unit of selection. *Genetics*, 72, 157–168.

[319] Sliwa, P. & Korona, R. (2005). Loss of dispensable genes is not adaptive in yeast. *Proc Natl Acad Sci USA*, 102, 17670–17674.

[320] Smith, J. M. & Haigh, J. (1974). Hitch-hiking effect of a favorable gene. *Genet Res*, 23, 23–35.

[321] Sniegowski, P. D. & Gerrish, P. J. (2010). Beneficial mutations and the dynamics of adaptation in asexual populations. *Phil Trans R Soc B*, 365, 1255–1263.

[322] Söderberg, R. J. & Berg, O. G. (2007). Mutational interference and the progression of muller's ratchet when mutations have a broad range of deleterious fitness effects. *Genetics*, 177, 971–986.

[323] Song, Y. S. & Steinrücken, M. A. (2012). A simple method for finding explicit analytic transition densities of diffusion processes with general diploid selection. *Genetics*, 190, 1117–1129.

[324] St Onge, R. P., Mani, R., Oh, J., Proctor, M., & *et al* (2007). Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nature Genetics*, 39, 199 – 206.

[325] Stephan, W., Chao, L., & Smale, J. G. (1993). The advance of muller's ratchet in a haploid asexual population: approximate solutions based on diffusion theory. *Genet Res*, 61, 225–231.

[326] Strelkowa, N. & Lässig, M. (2012). Clonal interference in the evolution of influenza. *Genetics*, 192, 671–682.

[327] Subramanian, S. (2012). The abundance of deleterious polymorphisms in humans. *Genetics*, 190, 1579–1583.

[328] Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G., & Lynch, M. (2012). Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl. Acad. Sci.*, 109, 18488–18492.

[329] Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics*, 123, 585–595.

[330] Takahata, N., Ishii, K., & Matsuda, H. (1975). Effect of temporal fluctuation of selection coefficient on gene frequency in a population. *Proceedings of the National Academy of Sciences*, 72(11), 4541–4545.

[331] Takahata, N. & Kimura, M. (1979). Genetic variability maintained in a finite population under mutation and autocorrelated random fluctuation of selection intensity. *Proceedings of the National Academy of Sciences*, 76(11), 5813–5817.

[332] Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R. L., McDonald, P., Bennett, A. F., Long, A. D., & Gaut, B. S. (2012). The molecular diversity of adaptive convergence. *Science*, 335, 457–461.

[333] Tsimring, L., Levine, H., & Kessler, D. (1996). Rna virus evolution via a fitness-space model. *Phys. Rev. Lett.*, 90, 088103.

[334] Turelli, M. & Barton, N. H. (1990). Dynamics of polygeneic characters under selection. *Theor Pop Biol*, 38, 1–57.

[335] Turner, P. E. & Elena, S. F. (2000). Cost of host radiation in an RNA virus. *Genetics*, 156(4), 1465–1470.

[336] Uecker, H. & Hermisson, J. (2011). On the fixation process of a beneficial mutation in a variable environment. *Genetics*, 188(4), 915–930.

[337] Wahl, L. M., Gerrish, P. J., & Saika-Voivod, I. (2002). Evaluating the impact of population bottlenecks in experimental evolution. *Genetics*, 162, 961–971.

[338] Waite, A. J. & Shou, W. (2012). Adaptation to a new environment allows cooperators to purge cheaters stochastically. *Proc. Natl. Acad. Sci. USA*, 109, 19079–19086.

[339] Wakeley, J. (2009). *Coalescent Theory, an Introduction*. Greenwood Village, CO: Roberts and Company.

[340] Walczak, A. M., Nicolaisen, L. E., Plotkin, J. B., & Desai, M. M. (2012). The structure of genealogies in the presence of purifying selection: a fitness-class coalescent. *Genetics*, 190, 753–779.

[341] Wall, J. D. & Pritchard, J. K. (2003). Haplotype blocks and linkage disequilibrium in the human genome. *Nat Rev Genet*, 4, 587–597.

[342] Wang, W., Thornton, K., Berry, A., & Long, M. (2002). Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science*, 295, 134–137.

[343] Waxman, D. & Loewe, L. (2010). A stochastic model for a single click of muller's ratchet. *Journal of Theoretical Biology*, 264, 1120–1132.

[344] Weinreich, D. M., Delaney, N. F., DePristo, M. A., & Hartl, D. L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, 312, 111–114.

[345] Weissman, D. & Hallatschek, O. (2014). The rate of adaptation in large sexual populations with linear chromosomes. *Genetics*, 196, 1167–1183.

[346] Weissman, D. B. & Barton, N. H. (2012). Limits to the rate of adaptive substitution in sexual populations. *PLoS Genetics*, 8, e1002740.

[347] Wenger, J. W., Piotrowski, J., Nagarajan, S., Chiotti, K., Sherlock, G., & *et al* (2011). Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS Genet*, 7, e1002202.

[348] Wichman, H. A., Badgett, M. R., Scott, L. A., Boulianne, C. M., & Bull, J. J. (1999). Different trajectories of parallel evolution during viral adaptation. *Science*, 285, 422–424.

[349] Wielgoss, S., Barrick, J. E., Tenaillon, O., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R. E., Schneider, D., & Andrews, B. J. (2011). Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with escherichia coli. *G3 (Bethesda)*, 1, 183–186.

[350] Wielgoss, S., Barrick, J. E., Tenaillon, O., Wiser, M. J., Dittmar, J., Cruveiller, S., Chane-Woon-Ming, B., Médigue, C., Lenski, R. E., & Schneider, D. (2013). Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci*, 110, 222–227.

[351] Wilke, C. O. (2004). The speed of adaptation in large asexual populations. *Genetics*, 167, 2045–2053.

[352] Williamson, S. & Orive, M. E. (2002). The genealogy of a sequence subject to purifying selection at multiple sites. *Mol Biol Evol*, 19, 1376–1384.

[353] Williamson, S. H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R., & Bustamante, C. D. (2005). Simultaneous inference of selection and population growth from patterns of variation in the human genome. *Proc Nat Acad Sci*, 102, 7882–7887.

[354] Wilson Sayres, M. A., Lohmueller, K. E., & Nielsen, R. (2014). Natural selection reduced diversity on human y chromosomes. *PLoS Genet*, 10, e1004064.

[355] Wiser, M. J., Ribeck, N., & Lenski, R. E. (2013). Long-term dynamics of adaptation in asexual populations. *Science*, 342, 1364–1367.

[356] Wloch, D. M., Szafraniec, K., Borts, R. H., & Korona, R. (2001). Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics*, 159, 441–452.

[357] Woodcock, G. & Higgs, P. G. (1996). Population evolution on a multiplicative single-peak fitness landscape. *J Theor Biol*, 179, 61–73.

[358] Woods, R. J., Barrick, J. E., Cooper, T. F., Shrestha, U., Kauth, M. R., & Lenski, R. E. (2011). Second-order selection for evolvability in a large *Escherichia coli* population. *Science*, 331, 1433–1436.

[359] Wright, S. (1931). Evolution in mendelian populations. *Genetics*, 16, 97–159.

[360] Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings of the VI International Congress of Genetics* (pp. 356–366).

[361] Wright, S. (1948). On the roles of directed and random changes in gene frequency in the genetics of populations. *Evolution*, (pp. 279–294).

[362] Wylie, C. S. & Shakhnovich, E. I. (2011). A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl. Acad. Sci. USA*, 108, 9916–9921.

[363] Yona, A. H., Manor, Y. S., Herbst, R. H., Romano, G. H., Mitchell, A., Kupiec, M., Y, P., & Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci*, 109, 21010–21015.

[364] Zanini, F., Brodin, J., Thebo, L., Lanz, C., Bratt, G., Albert, J., & Neher, R. A. (2015). Population genomics of intrapatient hiv-1 evolution. *eLife*, 4, e11282.

[365] Zanini, F. & Neher, R. A. (2012). Ffpopsim: An efficient forward simulation package for the evolution of large populations. *Bioinformatics*, 28, 3332–3333.

[366] Zeng, K. (2012). A coalescent model of background selection with recombination, demography and variation in selection coefficients. *Heredity*, 110, 363–371.

[367] Zeng, K. & Charlesworth, B. (2011). The joint effects of background selection and genetic recombination on local gene genealogies. *Genetics*, 189, 251–266.

[368] Zeyl, C. & DeVisser, J. A. G. M. (2001). Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics*, 157, 53–61.

[369] Zhang, W., Sehgal, V., Dinh, D. M., Azevedo, R. B., Cooper, T. F., & *et al* (2012). Estimation of the rate and effect of new beneficial mutations in asexual populations. *Theor. Popul. Biol.*, 81, 168–178.

[370] Zhu, Y., Siegal, M. L., Hall, D. W., & Petrov, D. A. (2014). Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci*, 111, E2310–E2318.