Notes on Evolution and Population Genetics

Michael M. Desai[1]

[1] *Lewis-Sigler Institute for Integrative Genomics,*

Princeton University,

Princeton NJ 08544

(Dated: July 1, 2007)

Many scientists (and not a few creationists) are struck at first by the incredible improbability of evolution. How could so many phenomenally unlikely mutations create something as intricate as an eye, or a wing? Biologists can only answer that nature had billions of years, so such things are possible. This answer is wholly unsatisfactory. Can anyone explain why the eye evolved over billions of years, rather than millions? Or why it doesn't take trillions?

These questions are hopelessly broad. Yet even in very simple and well-defined circumstances, surprisingly little is known about what is possible in evolution, over what timescales and in which conditions. In the early 1900s, R. A. Fisher, Sewall Wright, J.B.S. Haldane, and many others studied the basic forces driving evolution [1–7]. They analyzed simple models of general processes, such as genetic drift and the spread of advantageous mutations, in various types of populations. This work forms the basis for decades of subsequent work in population genetics. Yet this quantitative understanding remains limited. We understand the evolution of a single genetic locus, assuming its behavior is independent of everything else in the genome, and the evolution of linked neutral and deleterious variation. We understand how one beneficial mutation affects neighboring unselected loci, and how existing genetic variation responds to selection. This is often sufficient to understand the behavior of strongly selected traits in small populations, including most complex multicellular organisms. But whenever positive selection is common in larger populations, including many viruses and simple unicellular organisms, our understanding is relatively poor. Similarly, when the parameters of the evolutionary models, such as mutation rates and population structures, can themselves evolve, we typically do not understand how evolution works.

In these notes, I give a brief history of the quantitative study of evolution, and describe the limits of our understanding of the evolution of linked variation.

I. THE SINGLE-LOCUS FRAMEWORK OF EVOLUTIONARY THEORY

Two main goals dominate the study of evolution: understanding the diversity and histories of natural populations, and understanding the basic mechanisms by which evolution works. These goals have been closely linked since Darwin first connected observational analysis of natural variation to the basic mechanism of evolution: the generation of variation by

mutations and selection acting on this variation [8]. In these notes, however, we focus primarily on the latter question of how evolution works. We begin in this section by describing the early steps towards a quantitative understanding of evolutionary dynamics at a single locus.

When Darwin first proposed his theory of evolution, there was no understanding of the genetic system. This changed with Mendel's work, and in the early 1900s, Fisher [2], Wright [3, 4], and Haldane [5–7] began to quantify the operation of mutation, selection, and random genetic drift, and hence shed light on the basic mechanisms by which evolution works. Much of their focus was on the evolution of a single genetic locus. That is, they imagined a single gene at which two or more alleles were possible, and studied the dynamics of evolution at this locus under the assumption that it is independent of all other loci in the genome.

A. The Wright-Fisher Model of Evolution at a Single Locus

Fisher and Wright independently introduced the widely-used Wright-Fisher model of stochastic evolutionary dynamics at a single locus [2, 3]. In its simplest form, this model imagines a haploid population of N individuals, each of which can have one of two possible selectively neutral alleles at a particular genetic locus, with no mutation possible. Time is measured in discrete generations, and each subsequent generation is drawn by random sampling, with replacement, from the alleles in the previous generation (Fig. 1). That is, if in generation t there are i alleles of type A and $j = N - i$ alleles of type B, the number of alleles of type A in generation $t + 1$ is i', where the distribution of i' given i is

$$p_{i'|i} = \binom{N}{i'} \left(\frac{i}{N}\right)^{i'} \left[1 - \frac{i}{N}\right]^{N-i'} . \tag{1}$$

In a diploid population, this can become more complex. In this case, there are $2N$ rather than N copies of each allele in the population, since each individual has two. If mating is random, the Hardy-Weinberg law then states that if the frequency of allele A is a, then the probability of an AA diploid individual is just a^2, of an AB diploid is $2a(1-a)$, and of a BB diploid is $(1 - a)^2$ [9, 10]. In other words, the diploids are just random assemblages of the two alleles. This means that in a randomly mating sexual diploid population, the Wright-Fisher model applies exactly as described above, with N replaced by $2N$. In an asexual

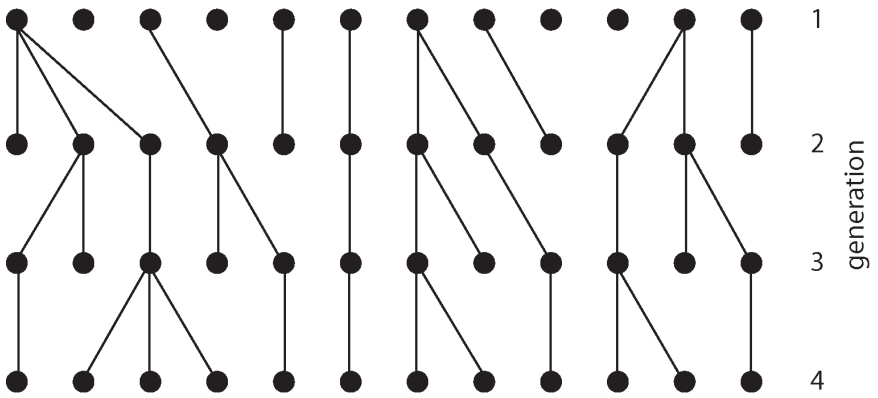

FIG. 1: Schematic of the neutral Wright-Fisher model without mutations. Individuals in generation are drawn by random sampling, with replacement, from the individuals in the previous generation.

diploid population, however, the haploid model is more appropriate. In the purely asexual case, the two sister chromosomes in each diploid individual evolve independently. Thus the dynamics of the allele at a single sister chromosome is given by the Wright-Fisher model with population size N. In these notes we are primarily interested in asexual populations, so we always use the Wright-Fisher model with N genes, which applies to either haploids or asexual diploids.

The basic Wright-Fisher model describes only the process of random genetic drift. Its dynamics can be understood by noting that the frequency of allele A is a Markov process with transition matrix $p_{i'|i}$. Since there is no mutation, eventually either allele A or B must become fixed in the population. Analysis of $p_{i'|i}$ shows that the probability allele A eventually becomes fixed, given that it started at frequency $\frac{i}{N}$, is just $\frac{i}{N}$ [3, 11]. Of particular interest is the case where we start with a single A allele. In this case, the probability that allele A will eventually fix is $\frac{1}{N}$. If this happens, it will take a time on average of order N generations. The A allele is much more likely to go extinct, and if this happens it will occur in a time of order $\ln N$ generations [3, 11, 12].

The effects of selection can be incorporated into the Wright-Fisher model by changing the transition probabilities $p_{i'|i}$. Rather than each generation being created by random sampling of alleles from the previous generation, this sampling is weighted to reflect the selective differences between alleles. For a haploid population, where allele A has fitness

$1+s$ and allele B has fitness 1, the Wright-Fisher model becomes [3, 11]

$$p_{i'|i} = \binom{N}{i'} \eta_i^{i'} \left(1-\eta_i\right)^{N-i'}, \tag{2}$$

where

$$\eta_i = \frac{(1+s)i}{(1+s)i + (N-i)}. \tag{3}$$

We can view this expression as the definition of fitness. It implies that when $|s| \ll 1$ (which we will always assume), allele A will tend to increase or decrease exponentially at rate s when rare. We call s the "selective advantage" (if $s > 0$) or "selective disadvantage" (if $s < 0$) of allele A. When A becomes more common, it increases or decreases exponentially at a rate $1+s$ minus the mean fitness of the population. In a sexual diploid population, the expressions for η_i are complicated by dominance. We do not explore these complications here, except to note than with no dominance the dynamics are similar to the haploid case. In an asexual diploid population, the relevant value for s is simply the fitness advantage or disadvantage of allele A given whatever allele is present at the sister locus.

As with the neutral case, either allele A or B will eventually become fixed in this model. Markov chain analysis or the diffusion methods described in the next section [3, 11, 13] show that if we start with a single individual with allele A, the probability that A fixes is

$$\Pr\left[A \text{ fixes}\right] = \frac{s}{1-e^{-Ns}}. \tag{4}$$

From this expression, we can see that Ns is a key parameter. When $|s| \ll \frac{1}{N}$, the selection pressure is always weak compared to drift and the probability of fixation is approximately $\frac{1}{N}$, the same as in the neutral case. However, when $s \gg \frac{1}{N}$, then the selection pressure is strong enough to be felt and the probability of fixation of A is s (or 0 if $s < 0$).

It is useful to pause to get an intuitive sense for this result. It is straightforward to show from the definition of the Wright-Fisher model that if there are i individuals with type A, the variance in the number of individuals of type A in the next generation is (for $s \ll 1$) of order $\sqrt{i}$. This is a diffusive process, so it takes i generations for i to change substantially due to drift. Over this time, selection has added (or removed) $i^2 s$ individuals. So selection becomes comparable to drift when $i = i^2|s|$, or $i = 1/|s|$. Above this size, the population will behave mostly deterministically (exponential growth or decline depending on the size of s) and below it, it will drift neutrally. Thus if we start from 1 individual, the number of A

alleles drifts approximately neutrally until A either goes extinct or reaches $\frac{1}{|s|}$ individuals. This immediately implies that if $N \ll \frac{1}{|s|}$, allele A will either fix or go extinct while it is effectively neutral. However, if $N \gg \frac{1}{|s|}$, allele A can reach size $\frac{1}{|s|}$ well before fixing. Since it is effectively neutral while less common than this, it has a probability $\frac{1}{1/|s|} = |s|$ of reaching this size. If and when it does reach size $\frac{1}{|s|}$, the selective difference can be felt. If $s > 0$, allele A begins to increase in frequency until it fixes. If $s < 0$, allele A can never grow past a size of order $\frac{1}{s}$. Hence when $|s| \gg \frac{1}{N}$, the fixation probability is s when $s > 0$ and 0 when $s < 0$.

We can also calculate the typical fixation and extinction times from this picture. When $s \ll \frac{1}{N}$, the fixation and extinction times are given approximately by the neutral result. When $s \gg \frac{1}{N}$, if $s < 0$ allele A must eventually go extinct, and this occurs in a time of order the extinction time in a population of size $\frac{1}{|s|}$, which is $\ln\left[\frac{1}{|s|}\right]$. If $s > 0$, if allele A goes extinct it does so in a time on average of order $\ln\left[\frac{1}{s}\right]$. On the other hand, if it fixes, it does so by reaching size $\frac{1}{s}$, which takes an average time given by the neutral fixation time in a population of size $\frac{1}{s}$, which is $\frac{1}{s}$ generations. It then grows roughly deterministically until it fixes in the population, which takes of order another $\frac{2}{s}\ln[Ns]$ generations.

This analysis of the Wright-Fisher model without mutations gives us a sense of the fate of a single mutant allele. We can again generalize this model to incorporate mutations. This can be done by introducing a probability that an individual of type A gives rise to an offspring of type B, and vice versa. We have

$$p_{i'|i} = \binom{N}{i'} \xi_i^{i'} \left(1 - \xi_i\right)^{N-i'}, \tag{5}$$

where

$$\xi_i = (1-u)\eta_i + v(1-\eta_i). \tag{6}$$

That is, each generation is created by random sampling, weighted by selective differences, of alleles from the previous generation, but each sampled A allele mutates to B with probability u each generation, and B mutates to A with probability v. Unlike the previous cases, there is now a steady state distribution of the frequency of the two alleles. This steady state is known as the mutation-selection-drift balance. It is characterized by the leading eigenvector of the Markov transition matrix $p_{i'|i}$ [3]; it can also be found by diffusion methods [14]. The

probability of finding allele A at a frequency $x \equiv \frac{i}{N}$ in this steady state is

$$f(x) = Cx^{Nu-1}(1-x)^{Nv-1}e^{Nsx}, \tag{7}$$

where C is a normalization constant.

For simplicity, we can study this expression in the special case where the mutation rates are symmetrical (i.e. $u = v$). In this case, we have $f(x) = C\left[x(1-x)\right]^{Nu-1}e^{Nsx}$. This distribution is depicted in Fig. 2. We can immediately see that the relevant quantities are Nu and Ns. When $Nu \ll 1$, $f(x)$ is a U-shaped distribution. The population is usually fixed for one or the other allele. Occasionally a mutation occurs and that mutant quickly either fixes, causing the population to become fixed for the other allele, or goes extinct. The amount of time the population spends fixed for A rather than B is determined by $N|s|$. When $N|s| \ll 1$, selection has little impact and the population is fixed for A roughly half the time. When $N|s| \gg 1$, the population is usually fixed for the selectively favored allele. On the other hand, when $Nu \gg 1$, the population is typically polymorphic; $f(x)$ is peaked around a most likely frequency for the two alleles. In this case, when $s \ll u$ the two alleles are maintained at roughly equal frequencies; the favored allele represents a fraction $\frac{1}{2} + \frac{s}{8u}$ of the population. On the other hand, when $s \gg u$, the selectively disfavored allele is maintained at frequency $\frac{u}{s}$.

This large-Nu behavior is known as the mutation-selection balance, because when $N \to \infty$, drift becomes unimportant in the steady state. That is, we get results identical to the $N \to \infty$ case if we simply write deterministic differential equations involving mutation and selection acting on the frequency of the two alleles. It is important to note, however, that even for very large N drift is important if we start from a population fixed for the deleterious allele; it is only unimportant if we are in the steady state or starting from the beneficial allele. For this reason, we will sometimes call this the deleterious mutation-selection balance, since when the population is fixed for less favorable alleles (and hence beneficial mutations are possible) drift will still be important.

When mutations are rare or selection is strong (i.e. when $Nu \ll 1$ or $Nu \gg 1$ and $s \gg u$), the full mutation-selection-drift balance can be related to the simpler drift-and-selection case described above. When $Nu \ll 1$, the population is usually fixed for one allele or the other. Thus each time a new mutation occurs it enters a population fixed

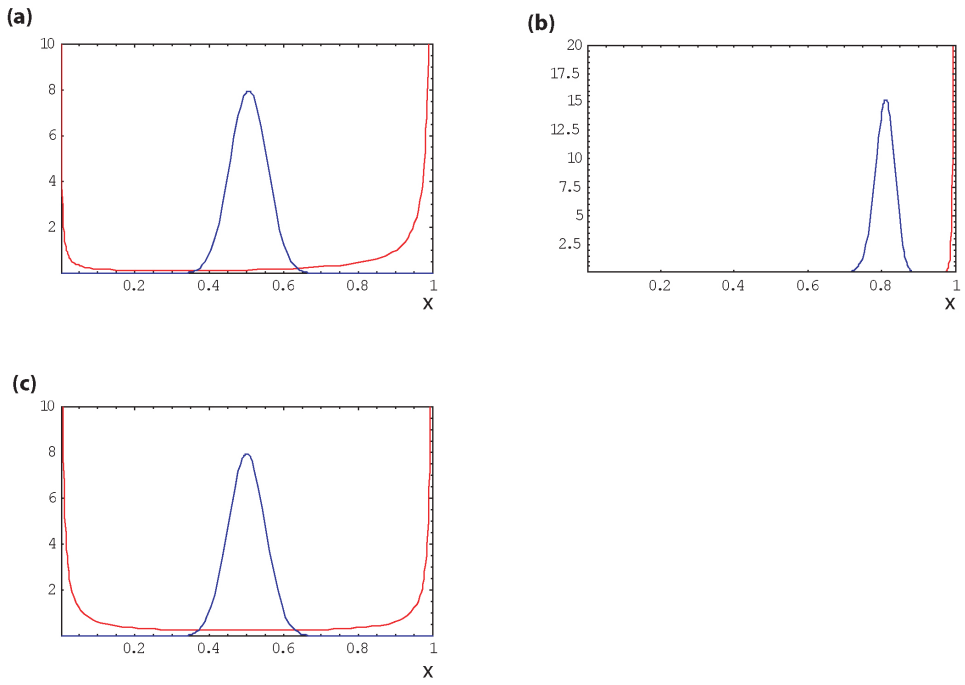

FIG. 2: The steady state mutation-selection-drift balance in the Wright-Fisher model with selection and mutation. In each panel, the small Nu ($Nu = 0.02$) steady state is shown in red, while the large Nu ($Nu = 50$) steady state is shown in blue. **(a)** The case of moderate selection, $Ns = 2$. Note that for $Nu \ll 1$, the population is typically nearly fixed for one allele or the other, with a bias towards being fixed for the more-fit allele. For $Nu \gg 1$, the population is usually polymorphic, with the favored allele only slightly more common because $s \ll u$. **(b)** The case of strong selection, $Ns = 200$. For $Nu \ll 1$ this looks similar to the $Ns = 2$ case, but the bias towards the more-fit allele is stronger. For $Nu \gg 1$, we now have $s \gg u$ so the favored allele is substantially more common. **(c)** The case of weak selection, $Ns = 10^{-3}$. For $Nu \ll 1$ the population is usually nearly fixed for one or the other allele, with both alleles equally likely. For $Nu \gg 1$ the population is typically polymorphic, with each allele present at a frequency roughly $\frac{1}{2}$.

for the other allele, and its fate is given by the drift-and-selection Wright-Fisher model. This provides a complete picture of the dynamics. On the other hand, when $Nu \gg 1$ and $s \gg u$, once the population is near its mutation-selection steady state, new mutations are constantly producing new individuals with the deleterious allele. Since these individuals are rare, they do not affect the mean fitness of the population (i.e. they can be neglected in the denominator of the expression for η_i). Thus the fate of each mutant is roughly independent

of all other mutants, and is given by the drift-and-selection model as if it had occurred in a population fixed for the favored allele. Each of these deleterious alleles will eventually go extinct, but more are constantly being produced. The mutation-selection-drift steady state is composed of many such mutant alleles which arose some time ago and are in the midst of their drift-and-selection transient dynamics which will eventually lead to their extinction.

This sort of heuristic understanding of a complex steady state in terms of transient dynamics of simpler processes is often very useful. Nevertheless, this type of ad-hoc approach is unusual in population genetics, which tends to focus on exact solutions and formal approximations.

B. Other Models of Single-Locus Evolution

The two-allele Wright-Fisher model is only one of many models of the evolution at a single locus. Cannings [15] introduced a generalization of the Wright-Fisher model where each individual can have a more general distribution of offspring number. Depending on the form of this distribution, there are a variety of possible Cannings models (of which the Wright-Fisher model is one) which share certain general properties. Moran [16] introduced a different type of model. In this Moran model, at each discrete timestep, a single individual is chosen to die, and another individual is chosen to reproduce to replace it. The choice of either individual can be weighted by selection, and the reproducing individual can mutate.

It is difficult to say which, if any, of these models is most appropriate for describing real populations. Though they all produce slightly different stochastic dynamics, with slight redefinitions of appropriate parameters they all lead to essentially the same behavior. This makes it difficult to determine from data which model best describes a particular population, but fortunately it also means that our choice of model is not of great importance. Although we may estimate slightly different parameters if we fit different models to the behavior of a given real population, as long as we are consistent the choice of model will not strongly affect any predictions we make.

C. Multiple Alleles

Thus far we have described models in which only two alleles are possible at the locus in question, A and B. This may be directly relevant if, for example, the locus in question is a gene and there are two possible alleles: the "ideal" functional allele A, and all possible non-functional alleles which we refer to collectively as B. Or similarly, the locus in question may be a single nucleotide, and one of its four possible states is best, and is represented by allele A, while the other three are deleterious, and are represented as allele B.

In general, however, we must consider the case when more than two alleles are possible. If the locus in question is a single nucleotide, there can be 4 possible alleles, A, G, C, and T. If the locus is a whole gene, then there may be an arbitrary number of different alleles.

We can extend the Wright-Fisher model described above to account for the situation where there are K possible alleles. Each new generation is still created by sampling of all of the individuals in the previous generation, weighted by selection and changed by mutation. The transition probabilities become more complex, because they now involve the frequencies of all other alleles. If we denote by i_k the number of individuals of allelic type k in generation t, and i'_k the number in generation $t+1$, we have for the case of no selection or mutation [17]

$$p_{i'_1,\ldots i'_K|i_1,\ldots,i_K} = \left[\frac{N!}{\prod_{k=1}^{K} i'_k!}\right] \prod_{k=1}^{K} \left[\frac{i_k}{N}\right]^{i'_k}. \tag{8}$$

The obvious generalizations apply when we add selection or mutation. Other such K-allele generalizations are similarly possible for the Moran model and various Cannings models [17].

These multiple-allele models are much more difficult to analyze than the simple two-allele model. In the two-allele case, there is only one independent variable: the frequency of allele A (with the frequency of B entirely determined by this). In a K-allele model, there are in general $K-1$ independent variables, which naturally makes the analysis much more complex. Unless there is a high degree of symmetry among the alleles, either computational methods or special tricks are typically required to get any understanding when K is large.

D. The Diffusion Approximation

Even the relatively simple two-allele models described above are difficult to analyze exactly. However, for large N we can we can make the approximation that the frequency of allele A, $x \equiv \frac{i}{N}$, is a continuous rather than a discrete variable. Since for large N, x only changes slightly each generation, we can also approximate time as continuous. We define $f(x; y, t)$ to be the probability distribution of x at time t, given that allele A started at frequency y at time 0. Since f is translationally invariant in time, we can write

$$f(x; y, t + \delta t) = \int_0^1 f(x + \delta x; y, t) f(x; x + \delta x, \delta t) d\delta x. \tag{9}$$

For many biologically reasonable forms of f, including the Wright-Fisher model, the mean $a(x)$ and variance $b(x)$ of an instantaneous change in x are of the same order in $\frac{1}{N}$, while all higher moments are of higher order in $\frac{1}{N}$ [12]. This means we can neglect these higher moments for large N, so that on Taylor expanding f in Eq. (9), we find an equation involving diffusion plus drift, where the drift term is scaled by $a(x)$ and the diffusion term is scaled $b(x)$. We have [12, 18]

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[a(x) f(x,t)\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b(x) f(x,t)\right]. \tag{10}$$

This is called the forward Kolmogorov equation, or simply the diffusion equation. The form of $a(x)$ and $b(x)$ depends on the specific model. For the Wright-Fisher model described above,

$$a(x) = sx(1-x) - ux + v(1-x) \tag{11}$$

$$b(x) = \frac{x(1-x)}{N}, \tag{12}$$

where time is measured in generations [14]. The dependence on $x(1-x)$ in the diffusion and selection terms is quite generic; it arises because in order for x to change, we must replace an individual of one type with an individual of the other, and the rate of this event depends on the product of their frequencies.

A variety of quantities, such as the fixation probabilities of mutants and the form of the mutation-selection-drift steady state, are much easier to calculate in diffusion approximations than in the exact Markov chain model. Calculation of the times to fixation or

extinction in models without mutation is also possible. The use of diffusion approximations for these purposes was pioneered by Kimura [12–14, 19]. However, the exact solutions of the diffusion approximations and many of the fixation or extinction time results are complex and opaque. For a number of practical purposes branching process methods of calculating these quantities are preferable.

The diffusion approximation generalizes easily to K-allele models. However, it quickly becomes unwieldy. The diffusion equation for two-allele models is one-dimensional because there is only one independent variable, x. However, for K-allele models, there are $K-1$ independent variables and hence a $K-1$-dimensional diffusion equation. It is typically impractical to use such a multi-dimensional diffusion equation to calculate any quantity of interest.

E. Branching Process Models

An alternative approach to studying the fate of a mutant in a large population is to use a branching process model (Fig. 3). In a discrete-time branching process model, each individual has some assigned probability f_k of having k offspring in the next generation ($k = 0, 1, 2, \ldots$). Each of these offspring then have k offspring with probability f_k, and so on. The key assumption of these models is that the distribution of offspring number of each individual is identical and independent. This is only true if the mutant lineage being followed represents a small fraction of the population. Once the mutant lineage becomes a substantial fraction of the population, the fate of each individual must depend on the others if population size is to be maintained.

Despite this limitation, branching process models are still useful in understanding the stochastic dynamics of mutant lineages while those lineages are rare. In particular, when $Ns \gg 1$, the lineage of a beneficial mutation is only stochastic when that lineage is rare. Thus we can understand the fate of a mutant lineage with a branching process model while it is rare, coupled to a deterministic model when it is common. Fitness is incorporated by choosing values of f_k such that the average number of offspring of each individual is $1 + s$. This is slightly different from the definition of fitness in the Wright-Fisher model, but is equivalent whenever $s \ll 1$ and the mutant lineage is rare. When the mutant lineage

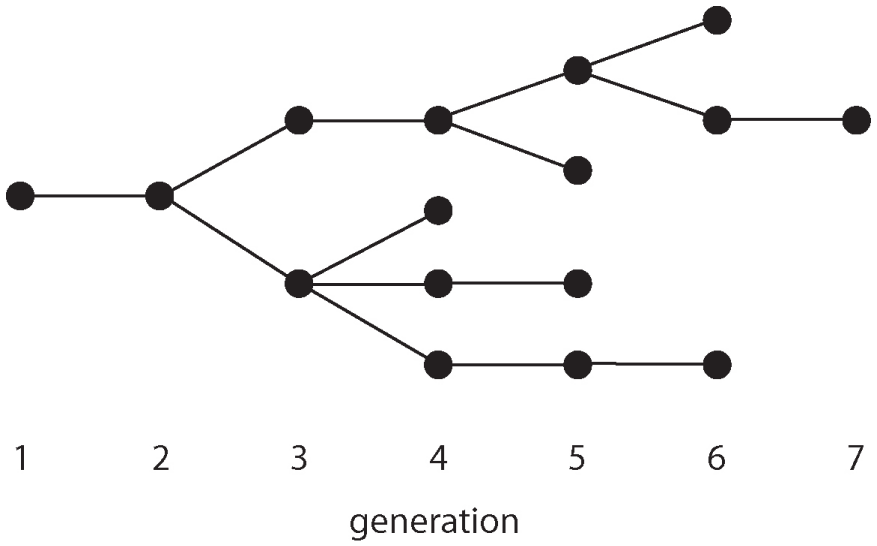

FIG. 3: Following the lineage of an individual with a branching process model. Initially we have a single individual at left. In each generation, each individual has k offspring with probability f_k. In this case, the lineage goes extinct after 7 generations.

becomes common, the branching process model fails because the mutants begin to affect the mean fitness of the population. Provided $Ns \gg 1$, these effects can be handled by a deterministic model, since the mutant lineage will no longer be stochastic by the time they become important.

The advantage of this approach, compared to a diffusion approximation which would be accurate both when a mutant lineage is rare and when it is common, is that many quantities are much easier to calculate in the branching process model. This idea was originally used by Haldane [7] to calculate the fixation probability of a beneficial mutation when $Ns \gg 1$; it is useful in a variety of other contexts.

II. MULTIPLE LOCI

The analysis above provides a complete framework for understanding the evolution of a single genetic locus. Yet any organism is composed of many different loci, which are all evolving at the same time. To understand how evolution works, therefore, we must understand how multiple loci can evolve at once.

In an purely asexual population, all these loci are perfectly linked. They cannot evolve

independently of one another: for example, if a mutation A_1 at locus 1 occurs in an individual with another mutation A_2 at locus 2, then A_1 cannot fix unless A_2 does. Thus in asexual populations, the evolution of many loci may look very different from what we would expect if all the loci evolved independently according to the single-locus results described above. In a sexual population, linkage is imperfect, as recombination mixes up the different loci. Loci that are close together on the genome are tightly linked, and hence must evolve together for long periods, while those that are far apart may be close to independent.

In order to appreciate how evolution works, we have to understand how many linked or partially linked sites can evolve together. This is also a crucial task if we are to use our knowledge of how evolutionary forces work to help understand the diversity and histories of natural populations. In many cases, we will be able to collect data on the frequencies of various alleles at some set of loci from an actual population. We must understand how the evolutionary forces shape the joint distribution of diversity at this possibly linked set of loci if we are to infer the evolutionary histories of the population from the observed patterns of variation.

The importance of this program became apparent in the 1960s, when molecular data on the variation in natural populations became available. At first, this data typically showed only some of the variation at the whole-gene level. More recently, it has become possible to cheaply sequence a gene or set of genes in many individuals within a population, giving us nucleotide-level resolution on the variation [20]. The appropriate definition of a "locus" and an "allele" depends on the data available. For example, a locus could be a gene, with various alleles representing sequences which give a particular spectrum of restriction fragment lengths. Or a locus could be a single nucleotide, with four possible alleles A, T, C, or G. As we describe below, our choice of definition of a locus typically involves a tradeoff between fewer loci with more alleles at each, or more loci with less alleles.

In this section, we describe various ways of modelling the evolution of multiple loci. Our focus is driven by the goal of understanding how linkage affects the basic mechanisms of evolution, but we also keep in mind methods of inferring evolutionary histories from data.

A. Free Recombination

The simplest model of the evolution of many loci is to assume that there is free recombination between all loci. In this case, each locus evolves independently of the others, and the single-locus results of the previous section apply to all loci independently.

In practice, completely free recombination is not necessary for this model to be a good approximation. The importance of linkage depends on how often linked loci "segregate" (i.e. are polymorphic) simultaneously, and the recombination rates compared to the timescales on which they segregate. If mutation rates and population sizes are such that it is very improbable that two loci ever segregate simultaneously, then even with perfect linkage the sites evolve independently and the free recombination results apply. If two loci do segregate simultaneously, linkage can only be ignored if recombination between the loci is fast compared to the timescale on which they segregate.

B. Few-locus Models

When free recombination is not a good approximation, we must consider the effects of linkage between loci. A natural approach to is to be begin with the simplest possible model involving multiple loci: a two-locus model with two alleles possible at each locus. We refer to the two alleles at one locus as A_1 and B_1, and the alleles at the other locus as A_2 and B_2. In such a model of a haploid population (or of a diploid population in Hardy-Weinberg equilibrium) there are four quantities of interest: the frequencies i_1, i_2, i_3, and i_4 of each possible combination of the two alleles at the two different loci (A_1A_2, A_1B_2, B_1A_2, and B_1B_2 respectively). Since the frequencies of these four combinations must add to 1, there are only three independent quantities.

This two-locus situation can be analyzed by the same methods used for the one-locus case. In each subsequent generation, the frequencies of the four alleles are given by a generalization of Eq. (8),

$$p_{i'_1 \ldots i'_4 | i_1 \ldots i_4} = \frac{N!}{\prod_{k=1}^{4} i'_k!} \prod_{k=1}^{4} \phi_k^{i'_k}, \tag{13}$$

where $\phi_k = i_k + \eta_k R\,[i_1 i_4 - i_2 i_3]$, R is the recombination rate, $\eta_1 = \eta_4 = 1$, and $\eta_2 = \eta_3 = -1$ [21–23].

This model can be analyzed with Markov chain methods or by a diffusion approximation [24–26]. In either case, it is useful to choose as independent variables three slightly different quantities: the frequency a_1 of allele A_1 (with the frequency b_1 of allele B_1 given by $b_1 = 1 - a_1$), the frequency a_2 of allele A_2 (with $b_2 = 1 - a_2$), and the "linkage disequilibrium" $D \equiv i_1 - a_1 a_2$. Note that a_1 and a_2 the first moments (i.e. averages) of the allele frequencies, while D is a second moment (related to the correlation in allele frequencies at the two loci). These three quantities completely describe all allele frequencies; they can be related by simple algebra to the frequencies of the four possible combinations of alleles at the two loci.

The analysis of the two-locus case is more complex than the one-locus case because instead of one independent variable, we have three. In the diffusion approximation, for example, a three-dimensional diffusion equation is required to describe the dynamics of the three variables a_1, a_2, and D. This three-dimensional diffusion approximation provides a full picture of the two-locus dynamics, but it is much more complex to analyze than the one-dimensional single-locus case.

We can use a similar analysis to study the evolution of several loci, each with two or more possible alleles. Each increase in the number of loci or alleles requires a higher-dimensional diffusion equation. For example, a two-allele three-locus case requires us to consider seven independent variables: the frequencies a_1, a_2, and a_3, the three possible linkage disequilibria (the frequency of $A_1A_2A_3$ plus $A_1A_2B_3$ minus a_1a_2, etc.), and a third-order correlation in allele frequencies. This latter term describes the three-point correlation between the allele frequencies, above and beyond what would be expected based on the frequencies of the alleles and correlations between any two allele frequencies [27–29]. A four-locus system would also involve fourth moments, and so on. This quickly becomes prohibitively complex to analyze.

However, when linkage is relatively weak, higher-order correlations will typically be negligible, so the frequencies of all possible combinations of alleles at the various loci will be well-described by the average allele frequencies and the correlations between these frequencies (i.e. the linkage disequilibria). In this situation, a many-locus system can be analyzed approximately with a few-dimensional diffusion equation involving allele frequencies and disequilibria only (particularly when there is a high degree of symmetry between loci). This is often used to understand the evolution of a large number of linked selected sites. It is important to remember, however, that this is a moment approximation, and is only valid

for weak linkage. It provides a first-order multilocus correction to the free recombination assumption. It is unclear how common this weak linkage case is in nature [17, 30–32]. In a purely asexual population, or a sexual population with low recombination rates, it will certainly be incorrect.

C. Infinite-Alleles Model

An alternative approach to the study of many loci is to make the extreme tradeoff between multiple alleles and multiple loci. In this approach, we assume that the entire genome (or the part of the genome for which we have data) is a single locus. Different mutations, typically at different sites in the genome, are considered to be different alleles.

In principle, we could use this approach to construct an exact model of the evolution. This would require keeping track of how the different alleles are related to each other: that allele 1 can mutate to allele 2 which can mutate to allele 3, but if instead allele 1 mutates to allele 4, it can then mutate to allele 5 but not allele 3, and so on. If recombination is included, we also have to know which alleles can be created from recombination between which other alleles.

In practice, such an approach is impractical. Instead, we assume that there are an infinite number of possible alleles, and that each mutation creates a new allele. This will be reasonable whenever we are considering small enough population sizes and mutation rates that the same mutation never occurs a second time before the previous such mutation has either fixed or gone extinct. In contrast to the weak-linkage moment-based approach above, we assume that recombination can be neglected, so that linkage is perfect.

In the case where all mutations are neutral, this infinite-alleles situation can be analyzed with the K-allele Wright-Fisher model, Eq. (8), in the limit $K \to \infty$ (since recombination converts alleles between types, at rates depending on allele frequencies, we cannot incorporate recombination into this model [33]). The distribution of allele frequencies clearly has no steady state, since new alleles are constantly arising and old ones are fixing or going extinct. However, there is a steady state distribution of the number of alleles in the population and their frequencies, although the identities of the alleles involved changes with time. This distribution can be calculated using diffusion methods [34–36]. It can also be studied by

calculating directly from the Markov model the probability that a sample of n individuals will have n_1 individuals with one allelic type, n_2 alleles of another allelic type, and so on [37]. This can be used to estimate Nu or to test whether the variation observed in a population is neutral [37, 38].

The infinite-alleles model throws away information on the history of the differences between the different alleles (i.e. by how many mutations two alleles differ), even though this information is present in comparative sequence data. This limits the ability of the infinite alleles model to explain observed patterns of diversity, and limits the statistical power in estimates of parameters from data and in tests of neutrality. It also limits the ability of the infinite alleles model to describe the effects of selection. The model can analyze the case when the population is originally fixed for one allele with a given fitness, and all other alleles have some other fitness (i.e. any initial mutation from the original genotype carries a fitness cost or benefit, but all subsequent mutations are neutral). It can also handle certain aspects of the case where all new alleles are deleterious [17, 39], or the case when new alleles have random fitness [40–42]. However, it cannot easily handle the situation where a variety of different beneficial mutations are available. In this case, those mutations will tend to spread, and the historical relationships of the alleles (i.e. the distribution of the number of mutations each individual has) becomes important.

D. Infinite-Sites Model

The infinite-sites model is an alternative approach to multilocus evolution, which directly analyzes nucleotide-level variation. In this model, each individual nucleotide is considered as a separate locus (a "site"). In its simplest version, all of these loci are perfectly linked, and all mutations are neutral (as in the infinite alleles model, it is difficult to account for the effects of recombination [17]). All new mutations are assumed to occur at previously unmutated sites. Formally, each individual is assumed to have an infinite number of sites, but the total mutation rate is finite, so there is vanishing probability that the same site ever mutates twice. The infinite sites assumption means that each mutation creates a new allele. Thus the infinite sites model is an infinite alleles model. However, the infinite sites model contains additional information on the number of mutations by which two individuals differ.

It can thus explain more aspects of the observed patterns of diversity in natural populations.

This model was originally introduced by Kimura [43, 44] to describe expected neutral patterns of variation in sequence data. It makes predictions about the expected number and frequency distribution of segregating sites in the population and among a sample of two or more individuals [45, 46]. These can be used to estimate parameters such as the population size and mutation rate from observed patterns of variation in samples of sequences from real populations [45, 47], or to test for neutrality [48].

Selection can be included in the infinite sites model in certain cases when the stochastic processes at each site can be treated independently. This makes it possible to analyze the situation when all possible mutations are deleterious. However, it is impossible to use this framework to study the situation where many beneficial mutations are available, except when those mutations are rare enough that they simply cause occasional selective sweeps that are well separated in time. When many beneficial mutations are possible and segregate simultaneously, their fates interact and interfere, and traditional infinite sites analysis does not apply [49].

E. Coalescent Theory

Thus far, we have analyzed population genetics from a forward-time perspective: given a particular state of the variation in a population, we described how mutation, selection, and drift lead to changes in this variation in the future. This provides an understanding of how basic evolutionary forces determine the genetic changes in populations and shape patterns of diversity. An alternative perspective is to note that the genetic variation between individuals depends on how closely they are related. Close relatives will typically be similar, while there will be more variation between distant relations. The variation in a population thus depends on its genealogical history.

Thus rather than study the forward-time evolutionary dynamics, we can understand the genetic diversity in a population by studying its genealogical history. This is called "coalescent theory" [50, 51]. It is an unnatural way to think about how mutation, selection, and drift drive the evolution of a population. It is therefore less useful in understanding how evolution works. However, it is a very natural way to think about how the evolutionary

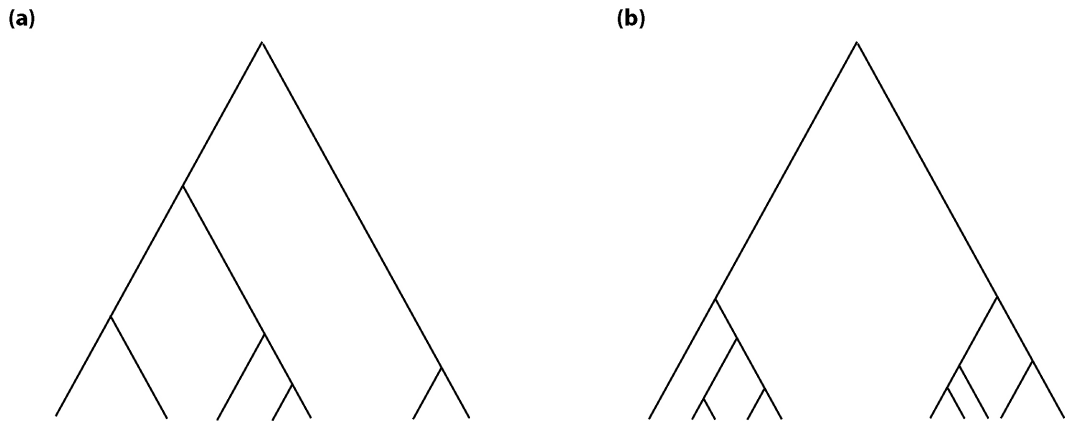

FIG. 4: Two examples of coalescent trees. The sample at the current time is represented by the tips of the tree at the bottom, and the genealogical history is drawn with the most distant past at the top. Each joining of branches represents two individuals with the same parent. **(a)** A typical coalescent tree for the neutral Wright-Fisher model. **(b)** A coalescent tree that might arise if the 5 individuals at left were isolated geographically from the 5 at right. Each individual is closely related to the others in its group, but only distantly related to all the individuals in the other group.

history of a population affects the current patterns of diversity, and hence is extremely useful in making inferences from data.

Coalescent theory was introduced by Kingman in the early 1980s [52–56]. The basic idea is simple: if we have a sample of n individuals in the present, there is some probability that two of these individuals had the same parent in the previous generation. If so, there were only $n-1$ individuals ancestral to our current sample in that previous generation. If not, there were n ancestors in the previous generation. Looking back further, some of these individuals in the previous generation may have had the same parent two generations ago, and so on. We can depict this as a genealogical tree (Fig. 4). The branch tips in this tree represent the individuals in the current sample. Each joining of branches represents two individuals which had the same parent. This is termed a "coalescent" event. Eventually all of the branches come together at the root of the tree, which represents the most recent common ancestor of all the individuals in the current sample.

The structure of the coalescent tree describes the pattern of relatedness among the individuals in the current sample. For example, the tree in Fig. 4b represents a sample with

two different groups of individuals which are closely related to others in their own group but only distantly related to individuals of the other group, perhaps because the two groups have long been geographically isolated.

Within any particular population genetic model, we can calculate the probability distribution of possible coalescent trees. The simplest example of this is the neutral Wright-Fisher model. Because of neutrality, the probability that two individuals in one generation have the same parent in the previous generation is just $\frac{1}{N}$. This completely determines the distribution of coalescent trees. If we have a small sample of n individuals in a large population ($N \gg n$), the time to the most recent coalescence event is exponentially distributed with mean $\frac{2N}{n(n-1)}$, and this coalescence event will occur between any two individuals in the sample with equal probability. The time to the next coalescence event is then again exponentially distributed with mean $\frac{2N}{(n-1)(n-2)}$, and so on, until the most recent common ancestor of the whole sample is reached [50].

When mutations are neutral, they have no effect on the distribution of tree structures. Rather, mutations simply label all descendants of the branch of the tree on which they occur. The mutational process is therefore completely independent of the genealogical process described above. Mutations are randomly distributed on the trees, with the probability of a mutation on any particular branch proportional to the branch length, population size, and mutation rate.

This completes the link between the population genetic model and the expected patterns of sequence variation. A particular model predicts a particular distribution of tree structures. Each such tree structure has a random distribution of mutations on it, which implies some pattern of genetic variation in the sampled individuals. Thus the probability distribution of genetic variation is determined by the probability distribution of tree structures and the expected variation given each possible tree.

Complications to the simple neutral Wright-Fisher model can be incorporated into the framework of coalescent theory, provided they do not affect the independence between the structure of the trees and the mutational labelling process. Geographical structure [57, 58] and variable population size [59] are two particularly important examples. Both of these complications change the distribution of genealogical tree structures (e.g. branches between geographically isolated populations are longer, and all branches are shorter during periods

of reduced population size). Given a particular tree structure, however, mutations are still distributed randomly among the branches.

The effect of recombination is somewhat more complex. Thus far, we have implicitly assumed no recombination by assuming that there is a single genealogical tree which describes the history of the entire genomes of the individuals in our sample. However, two parts of the genome of an individual have the same genealogy only so long as they are linked. When a recombination event occurs between these two parts, their genealogies can differ. This is because if two different individuals recombined in the past to create the genome of one individual in our sample, the parts of the genome that came from these different individuals have different genealogies before the recombination event combined them. This fact can be incorporated into coalescent theory by allowing lineages to diverge whenever a recombination event occurs [50, 55, 60, 61].

In contrast to geographic structure, demographic history, and recombination, the effects of selection cannot be incorporated naturally into the framework of coalescent theory. This is because selection causes the mutational labels to become intertwined with the genealogies. When mutations are selected, we cannot first calculate the distribution of tree structures and then add mutational labels, since the mutations affect the tree structure.

Despite this, two methods have been developed to incorporate selection into coalescent theory in certain limited circumstances. When selection is weak, $|s| \lesssim \frac{1}{N}$, we can draw a coalescent tree, but with lineages that diverge at a rate depending on the selective coefficients, to simulate the possibility that selectively favored individuals have more offspring. This results in an "ancestral selection graph," on which mutations occur randomly. This graph is then "pruned" to eliminate branches which, given the particular mutational pattern, are disfavored by selection, to generate an actual genealogy and mutational pattern [62–64]. An alternative approach is to model selection in an ad-hoc way: we assume that we already know the dynamics of the selected mutants, and incorporate this as an externally imposed constraint on the coalescent structure [65]. For example, we could assume that a strongly selected beneficial mutation occurred some specified number of generations ago, and caused a selective sweep. This then affects the structure of coalescent trees; it is similar to an exponential population expansion from a single individual [66]. This approach clearly cannot tell us anything about the dynamics of the selected mutants, but does tell us how

given dynamics affect the tree structure.

Although coalescent theory cannot in general predict the expected patterns of variation in the presence of many types of selection, it still provides a basis for testing whether or not observed variation is neutral. This is because the neutral coalescent makes definite predictions about the relationship between the frequency distribution of segregating sites, the number of segregating sites, and the distribution of the numbers of polymorphisms between individuals. If the observed relationships between these quantities differs from the neutral infinite sites approximation, then some assumption of this model must be violated. This can be quantified by a variety of statistics, such as Tajima's D and Fu and Li's D* and F* [48], which are expected to have certain values in the neutral coalescent, or by more detailed maximum likelihood techniques [67]. Selection is expected to cause deviations of these parameters in certain directions, so this procedure is often used to test for the presence of selection. Similar tests for selection can be used for the infinite sites and infinite alleles models. However, in all cases caution is required because a variety of other violations of the assumptions, such as spatial structure and variable population size, can also cause similar deviations of these test statistics even if all mutations are neutral.

III. LINKED BENEFICIAL MUTATIONS

In the previous section, we explored a number of techniques for understanding the evolution of a large number of linked loci. However, these methods all have trouble handling selection. While in some cases they can be applied to the case where all mutations are slightly deleterious, none of them provides any insight into the dynamics of many linked beneficial mutations. Although beneficial mutations are presumably rare compared to neutral or deleterious ones, this is a serious shortcoming. The accumulation of beneficial mutations is by definition responsible for the long-term adaptation of any population, so we can hardly claim to understand the basic mechanism of how evolution works if we do not understand how multiple linked beneficial mutations accumulate.

This gap in our understanding also limits our ability to draw inferences about the evolutionary histories of natural populations. In particular, the extensive body of work on detecting the presence of positive selection from sequence data is limited to qualitative and

ad-hoc methods. The statistical tests for selection described above (based on coalescent theory or the infinite-sites or infinite-alleles models) are only able to find deviations in certain statistics from their expectations under neutrality. We expect these deviations to have certain signs depending on the presence of different types of selection, but this is little more than a vague qualitative intuition. We do not know what patterns of diversity *are* expected to look like in the presence of positive selection, only what they are *not* expected to look like in the neutral case. Hence while we can say that a particular data set is inconsistent with neutrality, we have no way of estimating the strength of positive selection if indeed it is present. A similar problem applies to other tests for selection, such as dN/dS, which are not based on a specific population genetics model. These methods use a guess about the biology to posit that certain variation (e.g. synonymous sites) is neutral, and can hence act as a neutral "control" to compare to potentially selected variation (e.g. nonsynonymous sites). These methods can show that a data set is *not* consistent with neutrality, but again we do not know in any detail what such a data set *should* look like if many linked beneficial mutations are actually present.

Recently, we studied this issue of linked beneficial mutations. We studied the dynamics by which a large number of linked beneficial mutations accumulate [49], and tested these predictions experimentally [68]. This work focused on understanding how evolution "works" in the presence of a large number of linked beneficial mutations. We hope that in the future this will provide a basis for forming expectations about the variation in sequence data, and hence improved methods for detecting selection, but we have not yet explicitly addressed these questions.

IV. THE EVOLUTION OF EVOLVABILITY

The discussion of the previous sections makes clear that the strength of the linkage between different loci can have a dramatic effect on the way in which evolution works. The strength of this linkage is determined by whether the population reproduces asexually or sexually and, in the latter case, on recombination rates. Yet the mode of reproduction and the recombination rates are themselves controlled by evolution. This is an example of an important general phenomenon: the basic parameters of an evolutionary model, such

as recombination rates, mutation rates, dominance, and the sign and strength of epistasis, can themselves evolve. This is referred to as the "evolution of evolvability." The study of the evolution of evolvability can be complex, because mutations that change quantities like mutation and recombination rates have important indirect effects. This leads to an indirect selection which acts only at a remove in time.

Much early work on the evolution of evolvability focused on optimization arguments. The mutation rate [69], recombination rates [70, 71], sexual versus asexual reproduction [2, 72, 73], or other evolvability properties were assumed to evolve towards whatever values maximize the mean fitness (or some related measure of optimality) of the population. This approach gave no insight into how, or even if, selection would actually operate to change parameters towards their "optimum" values.

We can improve on these optimization arguments and study the actual effects of selection on evolvability by considering the fate of a modifier locus linked to directly selected loci [74]. For example, we could consider a three-locus system where two loci affect fitness and the third "modifier" locus controls the recombination rate between the other two [75, 76]. We can then take a game theory approach. We assume that the population is fixed for a particular allele at the modifier locus, and ask whether a new allele which increases or decreases the recombination rate can invade. We find the value of the recombination rate into which no other allele which increases or decreases this rate can invade. This is called an "evolutionarily stable" value of the recombination rate [77]. A similar approach can be used to find evolutionarily stable mutation rates [78–81], or to study the evolution of epistasis [82], dominance [83–88], and other evolvability properties [89, 90].

This game theory approach can find the long-term stable value of evolutionary parameters, assuming that these parameters can change freely during evolution. However, it typically does not address the dynamics of the modifier alleles in any detail. In some cases, these dynamics may be important. For example, in many natural bacterial populations mutator alleles which increase the mutation rate by a factor of 10 to 10^4 are seen at percent-level frequencies [91]. These mutator alleles are presumably not stable. Despite this, they are present at much higher frequencies than can be explained by the balance between their production and selection against them [91]. They must therefore be selected for sometimes, and may play an important role in the overall evolution of these populations despite not

being evolutionarily stable [92, 93]

Selection for evolvability can leave signatures in data which tell us about the evolutionary history of a population. For example, finding mutator alleles at high frequencies is a sign that positive selection was important in a population's recent past. We have recently made use of an unusual aspect of the evolution of evolvability to develop the theoretical basis for a new test for selection in sequence data [94]. This test is called "volatility," and is based on the idea that selection for or against changes in proteins will lead organisms to choose more or less evolvable gene sequences for those proteins [95]. Because the genetic code is redundant, each amino acid can be coded for by several different codons. At the most volatile codons, most of the nine possible mutations change the amino acid coded for. At less volatile codons, less mutations cause a chance in the amino acid. If there is selection on the protein to remain unchanged, there will be indirect selection on the gene sequence to choose less volatile codons. On the other hand, if a protein has recently changed, it will tend to have more volatile codons. Hence we can use this indirect selection for the "volatility" of a gene sequence to detect selection pressures that have acted on this sequence [94].

---

[1] R. A. Fisher, Transactions of the Royal Society of Edinburgh **52**, 399 (1918).

[2] R. A. Fisher, *The Genetical Theory of Natural Selection* (Oxford University Press, Oxford, UK, 1930).

[3] S. Wright, Genetics **16**, 97 (1931).

[4] S. Wright, *Evolution and the Genetics of Populations* (University of Chicago Press, Chicago, 1969).

[5] J. B. S. Haldane, Proceedings of the Cambridge Philosophical Society **23**, 363 (1926).

[6] J. B. S. Haldane, Proceedings of the Cambridge Philosophical Society **23**, 607 (1927).

[7] J. B. S. Haldane, Proceedings of the Cambridge Philosophical Society **23**, 838 (1927).

[8] C. Darwin, *On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life* (Murray, London, 1859).

[9] G. H. Hardy, Science **28**, 49 (1908).

[10] W. Weinberg, Jahresh. Wuertt. Ver. vaterl. Natkd. **64**, 369 (1908).

[11] R. A. Fisher, *The Genetical Theory of Natural Selection* (Dover, New York, 1958).

[12] M. Kimura, Proceedings of the National Academy of Sciences **41**, 144 (1955).

[13] M. Kimura, Annals of Mathematical Statistics **28**, 882 (1957).

[14] M. Kimura, Cold Spring Harbor Symposia on Quantitative Biology **20**, 33 (1955).

[15] C. Cannings, Advances in Applied Probability **6**, 260 (1974).

[16] P. A. P. Moran, Proceedings of the Cambridge Philosophical Society **54**, 60 (1958).

[17] W. J. Ewens, *Mathematical Population Genetics: I. Theoretical Introduction* (Springer, New York, NY, 2004).

[18] A. Kolmogorov, Mathematische Annalen **104**, 415 (1931).

[19] M. Kimura, Evolution **9**, 419 (1955).

[20] D. L. Hartl, *A Primer of Population Genetics* (Sinauer Associates, Sunderland, MA, 1988).

[21] S. Karlin and J. L. McGregor, Genetics **58**, 141 (1968).

[22] W. Hill and A. Robertson, Genetical Research **8**, 269 (1966).

[23] W. Hill and A. Robertson, Theoretical and Applied Genetics **38**, 226 (1968).

[24] T. Ohta and M. Kimura, Genetical Research **13**, 47 (1969).

[25] T. Ohta and M. Kimura, Genetics **63**, 229 (1969).

[26] R. A. Littler, Theoretical Population Biology **4**, 259 (1973).

[27] H. Geiringer, Annals of Mathematical Statistics **15**, 15 (1944).

[28] J. H. Bennett, Annals of Eugenics **18**, 311 (1954).

[29] M. Slatkin, Genetics **72**, 157 (1972).

[30] T. Ohta and M. Kimura, Genetical Research **25**, 313 (1975).

[31] I. R. Franklin and R. C. Lewontin, Genetics **65**, 701 (1970).

[32] R. C. Lewontin, *The Genetic Basis of Evolutionary Change* (Columbia University Press, New York, 1974).

[33] C. Strobeck and K. Morgan, Genetics **88**, 829 (1978).

[34] J. F. C. Kingman, Journal of the Royal Statistical Society B **37**, 1 (1975).

[35] J. F. C. Kingman, Theoretical Population Biology **11**, 274 (1977).

[36] G. A. Watterson and H. A. Guess, Theoretical Population Biology **11**, 141 (1977).

[37] W. J. Ewens, Theoretical Population Biology **3**, 87 (1972).

[38] G. A. Watterson, Genetics **85**, 789 (1977).

[39] W. H. Li and M. Nei, Genetics **86**, 901 (1977).

[40] S. N. Ethier and T. G. Kurtz, *The infinitely-many alleles model with selection as a measure-valued diffusion* (1987).

[41] S. N. Ethier and T. G. Kurtz, Stochastic Processes and their Applications **54**, 1 (1994).

[42] P. Joyce, Journal of Applied Probability **32**, 609 (1995).

[43] M. Kimura, Genetics **61**, 893 (1969).

[44] M. Kimura, Theoretical Population Biology **2**, 174 (1971).

[45] G. A. Watterson, Theoretical Population Biology **7**, 256 (1975).

[46] S. Tavare, Theoretical Population Biology **26**, 119 (1984).

[47] W. J. Ewens, Theoretical Population Biology **6**, 143 (1974).

[48] M. Kreitman, Annual Review of Genomics and Human Genetics **1**, 539 (2000).

[49] M. M. Desai and D. S. Fisher, Genetics **in press** (2007).

[50] M. Nordborg, in *Handbook of Statistical Genetics*, edited by D. J. Balding, M. Bishop, and C. Cannings (John Wiley and Sons, New York, 2001), pp. 179–212.

[51] J. Wakeley, *Coalescent Theory, an Introduction* (Roberts and Company, Greenwood Village, CO, 2006 (expected)).

[52] J. F. C. Kingman, Stochastic Processes and their Applications **13**, 235 (1982).

[53] J. F. C. Kingman, in *Exchangeability in Probability and Statistics*, edited by G. Koch and F. Spizzichino (North-Holland, Amsterdam, 1982), pp. 97–112.

[54] J. F. C. Kingman, Journal of Applied Probability **19A**, 27 (1982).

[55] R. R. Hudson, Evolution **37**, 203 (1983).

[56] F. Tajima, Genetics **105**, 437 (1983).

[57] R. R. Hudson, Molecular Ecology **7**, 413 (1998).

[58] H. M. Wilkinson-Herbots, Journal of Mathematical Biology **37**, 535 (1998).

[59] R. C. Griffiths and S. Tavare, Philosophical Transactions of the Royal Society of London B **344**, 403 (1994).

[60] R. C. Griffiths and P. Marjoram, Journal of Computational Biology **3**, 479 (1996).

[61] R. C. Griffiths and P. Marjoram, *An ancestral recombination graph* (1997).

[62] C. Neuhauser and S. M. Krone, Genetics **145**, 519 (1997).

[63] S. M. Krone and C. Neuhauser, Theoretical Population Biology **51**, 210 (1997).

[64] C. Neuhauser, Theoretical Population Biology **56**, 203 (1999).

[65] N. L. Kaplan, T. Darden, and R. R. Hudson, Genetical Research **57**, 83 (1988).

[66] N. L. Kaplan, R. R. Hudson, and C. H. Langley, Genetics **123**, 887 (1989).

[67] M. Stephens, in *Handbook of Stastical Genetics*, edited by D. J. Balding, M. Bishop, and C. Cannings (John Wiley and Sons, New York, 2001), pp. 213–238.

[68] M. M. Desai, D. S. Fisher, and A. W. Murray, Current Biology **17**, 385 (2007).

[69] M. Kimura, Journal of Genetics **57**, 21 (1960).

[70] H. J. Muller, Mutation Research **1**, 2 (1964).

[71] I. Eshel and M. W. Feldman, Theoretical Population Biology **1**, 88 (1970).

[72] J. F. Crow and M. Kimura, American Naturalist **909**, 439 (1965).

[73] J. Maynard-Smith, American Naturalist **102**, 469 (1968).

[74] L. Altenberg, Ph.D. thesis, Stanford University (1984).

[75] M. Nei, Genetics **57**, 625 (1967).

[76] M. W. Feldman, Theoretical Population Biology **3**, 324 (1972).

[77] J. Maynard-Smith, *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, UK, 1982).

[78] E. G. Leigh, American Naturalist **104**, 301 (1970).

[79] U. Liberman and M. W. Feldman, Theoretical Population Biology **30**, 125 (1986).

[80] T. Johnson, Genetics **151**, 1621 (1999).

[81] J. B. Andre and B. Godelle, Genetics **172**, 611 (2006).

[82] U. Liberman and M. W. Feldman, Theoretical Population Biology **67**, 141 (2005).

[83] R. A. Fisher, American Naturalist **62**, 115 (1928).

[84] R. A. Fisher, American Naturalist **62**, 571 (1928).

[85] R. A. Fisher, American Naturalist **64**, 385 (1930).

[86] R. A. Fisher, Biological Reviews **6**, 345 (1931).

[87] R. A. Fisher, American Naturalist **68**, 370 (1934).

[88] S. Wright, American Naturalist **63**, 556 (1929).

[89] S. Karlin and J. L. McGregor, Theoretical Population Biology **5**, 59 (1974).

[90] M. W. Feldman and U. Liberman, Proceedings of the National Academy of Sciences **83**, 4824 (1986).

[91] E. Denamur and I. Matic, Molecular Microbiology **60**, 820 (2006).

[92] M. M. Desai and D. S. Fisher, Genetics **submitted** (2007).

[93] D. A. Thompson, M. M. Desai, and A. W. Murray, Current Biology **16**, 1581 (2006).

[94] J. B. Plotkin, J. Dushoff, M. M. Desai, and H. B. Fraser, Journal of Molecular Evolution **63**, 635 (2006).

[95] J. B. Plotkin, J. Dushoff, and H. B. Fraser, Nature **248**, 942 (2004).