Steady states in rapidly evolving populations

by

Matthew Melissa

B.S. Engineering Physics
University of California, Berkeley, 2014

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Master of Science
in Physics
at the Massachusetts Institute of Technology

September 2017

Signature of Author: Signature redacted

Department of Physics
August 18, 2017

Certified by: Signature redacted

Mehran Kardar
Francis Friedman Professor of Physics
Thesis Supervisor

Accepted by: Signature redacted

Nergis Mavalvala
Curtis and Kathleen Marble Professor of Astrophysics
Associate Department Head of Physics

Steady states in rapidly evolving populations

by Matthew Melissa

Submitted to the Department of Physics
on August 18, 2017 in partial fulfillment of the
requirements for the degree of Master of Science in
Physics

Abstract

Populations are subject to mutations conferring beneficial effects, as well as mutations conferring deleterious effects. Even if deleterious mutations occur much more frequently than beneficial mutations, the contribution of deleterious mutations to the overall rate of change of the population-wide mean fitness may be limited, due to the efficient action of selection. However, in particularly rapidly evolving populations, the stochastic accumulation of deleterious mutations may negate a significant fraction of the fitness increments provided by beneficial mutations, or even result in an overall decrease in fitness over time. Here we obtain a constraint on beneficial and deleterious mutation rates and selection pressures in order for positive adaptation to counterbalance fitness decline via Muller's ratchet. The steady state that separates parameter regimes of positive adaptation and negative adaptation is characterized by appealing to the exponential dependence of fixation probabilities on fitness effect sizes. We consider a range of fitness-mediated epistatic interactions and the corresponding implications regarding the existence, location and stability of long-term evolutionary fixed points.

Thesis Supervisor: Mehran Kardar
Title: Francis Friedman Professor of Physics

)

)

)

)

Acknowledgments

There are several individuals who have really helped shape my experience at MIT and Harvard these last few years, and have helped me grow both academically and personally. First, I would like to thank all of the members —past and present—of the Desai Lab, who have been enormously helpful in bouncing ideas around, and have really made my experience in the group an enjoyable one. In particular, I'd like to acknowledge Ben Good for initial work on the generalized infinitesimal limit and for useful feedback on this project, and I'd also like to thank Ivana Cvijovic and Ramya Purkanti for providing helpful comments on this thesis. And of course, I'd like to thank my advisor, Michael Desai, for introducing me to (what was to me) an entirely new way of approaching scientific problems.

At MIT, a few individuals stand out as particularly deserving of thanks. I'd like to Nikta Fahkri for helping me grow more confident in teaching. I'd like to thank all of the staff in Academic Programs—in particular Cathy Modica and Sydney Miller—for helping make everything run smoothly. And I'd like to thank Mehran Kardar for his guidance in developing this thesis.

Last but definitely not least, I would like to thank my friends and family, who have been supportive and encouraging at every step so far along my journey.

Contents

Abstract 3

Acknowledgments 4

List of Figures 9

1 Introduction 13
1.1 The Model 16

2 Evolvability phase diagrams 19

3 Evolutionary equations 29
3.1 Evolutionary equations 29
3.2 Infinitesimal limit 31
3.3 Generalized infinitesimal limit 37
3.4 Evolvability phase diagrams in the generalized infinitesimal limit 40

4 Distributions of fitness effects 45
4.1 Evolvability phase diagrams in the space of scaled fitness effects 45
4.2 Evolvability phase diagrams in the space of unscaled fitness effects . . . 50

5 Fitness-mediated epistasis 53

6 Discussion 57

A Simulation Methods 61

Bibliography 63

List of Figures

2.1 Cross sections of an "evolvability phase diagram" in the space of scaled beneficial and scaled deleterious fitness effects, with the color of each point denoting the average rate of adaptation as measured in a Wright-Fisher simulation, with $NU = 10^4$ (see Appendix A for more details on simulation methods). *Unscaled* fitness effects s_b and s_d were specified, and for each point, the pairwise heterozygosity π was measured. (This yields the pairwise coalescence timescale $T_2 = \pi/2U_n$ and thus the corresponding scaled fitness effects.) Unscaled fitness effects were chosen to make up a log-uniformly spaced grid in the space of unscaled fitness effects (see Figure 3.2 for this grid); the non-uniform density of points in the space of scaled fitness effects reflects variation in T_2 as a function of s_b and s_d. The shape of the sampled region in the space of scaled fitness effects is a consequence of the fact that T_2 decays monotonically with s_b, but is a non-monotonic function of s_d. The dashed line is the theoretical prediction for the fixed-point curve given by Equation 2.8. 22

2.2 (Left) Points delineate fixed-point curves as measured from simulations, over a range of U_b/U values with $NU = 10^4$. Solid lines denote the corresponding theoretical predictions for the same set of U_b/U values. (Right) Ratio of fixed-point scaled beneficial effect as measured from simulation to the corresponding theoretically-predicted fixed-point scaled beneficial effect, over a range of scaled deleterious effects and U_b/U values, with $NU = 10^4$. 23

2.3 Scaling of γ_d^* (top) and γ_b^* (bottom), as measured from simulations. The dashed line in the plot of γ_b^* is the theoretical prediction. Error bars for γ_b^* encompass the range of γ_b values that could be an extreme value, assuming that true T_2 values fall within one standard error of their sample mean. Error bars for γ_d^* correspond to the range of γ_d values that correspond to a γ_b within the aforementioned range of γ_b^* values, with the inherent correlation between γ_b and γ_d taken into account. Variation in the rate of adaptation from the sample mean rate is assumed to be negligible compared to variation in the pairwise heterozygosity. (The population fitness and pairwise heterozygosity are sampled approximately once every T_{sweep} generations, a total of 192 times, where T_{sweep} denotes the time until one of the founding individuals sweeps through the population.) 25

2.4 Rate of adaptation as a function of U_b/U_d and $\gamma_d e^{-\gamma_d}/\gamma_b e^{\gamma_b}$. As in Figure 2.1, unscaled fitness effects and mutation rates U_b and U_d were specified; the corresponding scaled fitness effects and the rate of adaptation were then obtained from simulations. (See Figure 3.2 for the log-uniformly spaced grid of unscaled fitness effects, which was specified identically for all values of U_b/U_d.) Results from simulations with $NU = 10^2$, $NU = 10^3$, and $NU = 10^4$ are included. The dashed line is the theoretical prediction for points lying on the $v = 0$ boundary surface, given by Equation 2.10. 26

3.1 Predicted T_2/N vs. simulated T_2/N, over a grid of s_b and s_d values with $NU = 10^4$. Beneficial fitness effects ranged from $s_b = 1/N$ to $s_b = 0.5$; deleterious fitness effects ranged from $s_d = 1/N$ to $s_d = 3$. For some combinations of s_b and s_d, the numerical solution obtained is inconsistent with the initial assumption that both $s_d < b$ and $s_b < b$; these points are omitted. Color denotes the corresponding simulated value of γ_d, which is found to be an reasonably good predictor of substantial deviations of simulated values from predicted values. 40

3.2 Cross sections of an "evolvability phase diagram" in the space of unscaled beneficial and unscaled deleterious fitness effects, with the color of each point denoting the average rate of adaptation as measured in a Wright-Fisher simulation, with $NU = 10^4$. The dashed line is the theoretical prediction for the fixed-point curve given by the numerical solution outlined in Section 3.3. 42

3.3 (Left) Predicted fixed-point ϵ vs. simulated fixed-point ϵ, over a grid of s_b and s_d values at $NU = 10^2$, $NU = 10^3$, and $NU = 10^4$. For each combination of s_b and s_d, the simulated ϵ is obtained by interpolating between the largest simulated ϵ with a negative rate and the lowest simulated ϵ with a positive rate. The corresponding prediction for the fixed-point ϵ is obtained using the numerical solution outlined in the previous section. For some combinations of s_b and s_d, simulations over the entire specified range of ϵ values all resulted in $v > 0$ or all resulted in $v < 0$; these choices of s_b and s_d are omitted from the figure. (Right) ϵ_c as a function of s_b for a range of values of NU. For a particular s_b, ϵ_c is defined as the minimum ϵ above which $v > 0$ regardless of s_d. Curves denote predicted values (using numerical solution) and points denote simulation results. 44

4.1 Cross sections of an evolvability phase diagram, in the space of scaled fitness effects, for gamma-distributed beneficial and deleterious mutations, with $\alpha_b = \alpha_d = 1$ (exponential distributions). The solid lines are the fixed-point curve predictions that account for the finite shapes of the distributions of fitness effects, given by equating the right-hand side Equation 4.2 to zero; the dashed lines are the theoretical predictions given by Equation 2.8 for a single scaled beneficial effect size and a single scaled deleterious effect size. The population-wide mutation rate was specified as $NU = 10^4$. 46

4.2 Scaling of γ_b^* and γ_d^* as a function of U_b/U. Both beneficial and deleterious effect sizes were drawn from gamma distributions, with various choices of the shape parameters α_b and α_d. The population-wide mutation rate was specified as $NU = 10^4$. Points are obtained from simulations; solid curves denote theoretical predictions given by Equations 4.3 and 4.4. 47

4.3 Scaling of s_b^* and s_d^* as a function of U_b/U. Both beneficial and deleterious effect sizes were drawn from gamma distributions, with various choices of the shape parameters α_b and α_d. The population-wide mutation rate was specified as $NU = 10^4$. Points are obtained from simulations; solid curves denote theoretical predictions obtained using the numerical solution outlined above. 51

Chapter 1

Introduction

ADAPTATION to a constant environment tends to slow down over time. This has been observed in numerous microbial evolution experiments, in which competitive fitness (for example, relative to an ancestral genotype) increases at a declining rate [7]. Despite a declining *rate* of adaptation, fitness appears to have continued increasing throughout 60,000 generations of evolution of *Escherichia coli* in the Long-Term Evolution Experiment (LTEE) [49]. Fitting curves to the fitness trajectory supports a power-law trajectory, in which fitness increases indefinitely, over a fitness trajectory that asymptotes to a constant, steady-state fitness. In contrast, another study, in which small populations of bacteriophage were evolved experimentally, suggests that under certain conditions, fitness changes may not persist indefinitely [43]. Instead, over long evolutionary timescales, populations of sufficiently high and low initial fitness may converge to a steady state of intermediate fitness, where adaptation stops. It remains unclear both (i) which microscopic mechanisms are fundamentally responsible for the observed trend of declining adaptability and (ii) which factors determine whether a population will tend to an evolutionary steady state.

Mutations drive adaptation—therefore, the phenomenon of declining adaptability should be able to be explained by changes in the rate at which mutations occur and the effects they confer. One mechanism is intuitively appealing: a finite number of beneficial mutations may be available, so that as a population adapts, beneficial mutations might occur less frequently. Compounding this effect, different mutations may overlap in their functional effects, so that as adaptation proceeds, an increasing number of mutations that were originally beneficial might become redundant [27, 45]. Equivalently, individuals at a lower fitness are more likely to possess a larger number of deleterious mutations; back and compensatory mutations occur more frequently in these individuals, increasing the supply of beneficial mutations at lower fitness [3].

Recent empirical studies have demonstrated that in many cases, the rate of adaptation declines even while beneficial mutations continue accumulating at a nearly constant rate [27, 49, 51]. This suggests that a reduced beneficial mutation rate may not be the primary cause of declining adaptability; instead, this phenomenon may be better explained by a tendency for beneficial mutations to confer weaker effects, on average, as adaptation proceeds. Support for this explanation is found in several empirical studies in which, after introducing beneficial mutations onto several different genetic back-

grounds, beneficial fitness effects are found to be negatively correlated with background fitnesses [5, 30, 46].

Both are explanations that fundamentally invoke *epistasis*—interactions in which the fitness effects of new mutations are dependent on the background genotypes they land on. More precisely, both assume a pattern of *fitness-mediated* epistasis, in which the statistics of epistatic interactions can largely be summarized by a certain dependence of the population-genetic parameters on fitness [14]. We will refer to the phenomenon in which the beneficial mutation rate decreases with fitness as *running out of beneficial mutations* and that in which the typical beneficial effect size decreases with fitness as *diminishing-returns epistasis for beneficial mutations* . Other models of diminishing-returns epistasis are certainly plausible. For instance, the typical selective effects of deleterious mutations might vary systematically with fitness.

Previous theoretical work has focused on running out of beneficial mutations, and established important theoretical consequences for the long-term fitness trajectory of a population [20]. In this work, one of the simplest possible models is considered: the beneficial mutation rate declines with increasing fitness, but at the same time, the deleterious mutation rate increases, such that the total mutation rate remains constant. (As the supply of potential beneficial mutations "runs out", potential beneficial mutations are replaced by potential deleterious back-mutations of the same magnitude but opposite sign.) Consequently, the population will approach an evolutionarily stable fixed point at which the rate of fitness increase due to the accumulation of beneficial mutations is, on average, negated by the stochastic fixation of deleterious mutations (Muller's ratchet). The fixed point is known to be stable by a simple argument: if the fitness is perturbed above (or below) the fixed point, a smaller (or larger) fraction of mutations will then be beneficial, so that the population-wide fitness will decrease (or increase) back to the fixed point. The existence of this stable fixed point may explain the plateaus observed in empirically obtained fitness trajectories. This work assumes that mutational fitness effects are constant over the course of evolution, and that fitness effects of beneficial mutations are of the same magnitude as the fitness effects of deleterious mutations.

Given that mounting empirical evidence highlights the role of changes to the typical selective effects in bringing about declining adaptability, a natural question is whether the qualitative behavior predicted in [20] is implied by more general models of fitness-mediated epistasis. As an example, one might consider a model in which the rates of beneficial and deleterious mutations do not change appreciably over the course of evolution, and effect sizes of deleterious mutations remain roughly constant, with diminishing-returns epistasis for beneficial mutations the primary source of declining adaptability. The rate of adaptation increases monotonically with the size of beneficial effects, just as it increases monotonically with the rate of beneficial mutations, enabling a clear parallel to be drawn to the model studied in in [20]. The same heuristic arguments justify the prediction of the existence of an evolutionarily stable fixed point. Quantitative theory would be necessary to establish the critical beneficial effect size

at the fixed point, just as theory is employed in [20] to identify the critical beneficial mutation rate at the fixed point.

A fundamental asymmetry exists between beneficial and deleterious mutations. Stronger beneficial mutations are more likely to fix within a population, compared to weaker beneficial mutations. In contrast, stronger deleterious mutations are *less* likely to fix within a population, compared to weaker deleterious mutations. Consequently, the rate of adaptation does not necessarily vary monotonically with the strength of deleterious mutations. Therefore, in considering the implications of an alternative model of fitness-mediated epistasis, in which deleterious effect sizes change but beneficial effect sizes remain constant, the heuristic arguments of [20] do not necessarily apply. As a result, theory is necessary to predict not only the location, but also the potential existence and stability of any evolutionary fixed points.

To obtain the implications of more general models of fitness-mediated epistasis, in which fitness effects and/or mutation rates of both beneficial mutations and deleterious mutations depend on fitness, an understanding of how the rate of adaptation depends on these parameters is of central importance. Using this information, and given a model of how the population-genetic parameters vary with fitness, the long-term flow of the population through parameter space can, at least in principle, be integrated up, yielding predictions regarding the flow to (or possibly away from) fixed points. In fact, given knowledge of how only the *sign* of the rate of adaptation depends on the various population-genetic parameters, one can still extract the *path* of the population through parameter space, at the expense of information regarding the temporal dynamics through parameter space. Unfortunately, despite much recent attention, analytical expressions for the rate of adaptation (or its sign) have not been obtained, except in special cases [10, 13, 16, 41].

The rate of adaptation is one observable of the evolutionary process; diversity statistics such as the pairwise heterozygosity π comprise a somewhat disparate class of observables [48]. A related quantity of interest is the pairwise coalescence timescale T_2—the average number of generations since two randomly chosen individuals share a common ancestor [26]. Knowledge of this timescale provides significant predictive power regarding which mutations are likely to survive genetic drift and/or selection [25, 39]. Selection brings about significant changes to allele frequencies on a timescale that is inversely related to fitness differences; if this timescale is larger than the coalescence timescale, diversity is purged by genetic drift before selection is able to operate [13]. The inverse of the coalescence timescale therefore sets the fitness effect threshold above which selection dominates drift: a longer coalescence timescale is associated with a greater probability that a given beneficial mutation will survive genetic drift and a lower probability that a given deleterious mutation will stochastically fix.

In this work, we exploit the relationship between the coalescence timescale, fitness effect sizes and fixation probabilities to predict the rate of adaptation over a broad range of parameters, in which the effects of clonal interference and hitchhiking may play a significant role in the dynamics. We focus particular attention on the regime

in which the strength of selection is weak enough relative to mutation that Muller's ratchet clicks often; this allows us to identify a $v = 0$ "boundary surface" and construct an "evolvability phase diagram" that distinguishes regions of parameter space in which positive adaptation dominates from those in which negative adaptation dominates. From this we can immediately generate the long-term evolutionary implications of diminishing-returns epistasis, running out of beneficial mutations, and various other epistatic models. We find, in contrast to the model studied in [20], that not all fitness-mediated epistatic models give rise to an evolutionary fixed point, and that the precise details of a given epistatic model can determine the existence, location, and stability of evolutionary fixed points.

■ 1.1 The Model

We model an asexual haploid population of N individuals. Each generation, an individual's offspring number is drawn from a Poisson distribution with an expected value proportional to $e^{X-\bar{X}(t)}$, where X is the individual's fitness and $\bar{X}(t)$ is the mean fitness of the population at time t. An individual's fitness is determined by the mutations it has acquired; we'll assume that individuals are subject to both beneficial and deleterious mutations occurring at a total rate U (per generation, per genome), and denote by U_b and U_d the rates of beneficial and deleterious mutations, respectively. Our model does not account for neutral mutations, since these mutations do not impact the rate of adaptation or the pairwise coalescence timescale. The fraction of mutations that are beneficial is denoted by $\epsilon \equiv U_b/U$.

We will assume that an individual's fitness X is given by the sum of the fitness effects it has acquired. In much of the analysis that follows, we'll assume that all beneficial mutations confer the same fitness effect $s_b > 0$, and that all deleterious mutations confer the same fitness effect $-s_d < 0$. Later on we'll generalize much of our analysis to account for *distributions of fitness effects* (DFE), so that the fitness effect of a particular mutation is drawn from a given distribution $\rho(s)$. The distributions of beneficial and deleterious fitness effects need not be identical; we will partition the overall DFE into a distribution of beneficial fitness effects $\rho_b(s)$ and a distribution of deleterious fitness effects $\rho_d(s)$, with $U\rho(s) \equiv U_b\rho_b(s) + U_d\rho_d(s)$. The quantities N and $U\rho(s)$ fully determine the average rate of adaptation and the statistics of genetic diversity; together, these specify a point in population-genetic parameter space.

We will limit our analysis to epistatic interactions in which $\rho(s)$ depends on genotype, but only via the genotype's fitness (fitness-mediated epistasis). Additionally, we will assume that the fitness scale on which the DFE varies is much greater than the scale of fitness differences within the population, so that all individuals in the population share the same DFE at any given point in time, and the DFE does not change appreciably over the course of fixation of a single mutation. We refer to this as a *quasistatic* approximation; invoking this approximation enables a mapping to an epistasis-free model, simplifying the analysis considerably. In the discussion, we will

consider the validity of this approximation and briefly consider epistatic interactions in which the DFE varies more rapidly with fitness.

Chapter 2

Evolvability phase diagrams

GIVEN the assumptions stated in Section 1.1, a specific model of fitness-mediated epistasis will constrain the population to lie on a particular curve in population-genetic parameter space. Since we're assuming that $U\rho(s)$ depends only on the (mean) fitness of the population, the direction of the population's flow along this curve is determined by the sign of the rate of adaptation. This motivates us to construct an "evolvability phase diagram" that describes, for a given mutation rate and distribution of fitness effects, whether the corresponding rate of adaptation is positive or negative. That is, we would like to divide the parameter space into a region in which the rate of adaptation is dominated by Muller's ratchet, and a region in which the rate of adaptation is dominated by the accumulation of beneficial mutations. In a later section, we will use this information to generate the long-term implications of various models of fitness-mediated epistasis on the approach to an evolutionary steady state.

Mutations with an effect on fitness between s and $s + ds$ enter the population at a total rate $NU\rho(s)ds$. Due to genetic drift, the vast majority of lineages carrying these new mutations—even those with sizable beneficial effects—will go extinct before reaching an appreciable frequency in the population. If a mutational lineage survives genetic drift and selection, and expands to include all individuals in a population, that mutation is said to have *fixed*. We'll denote by $p_{\text{fix}}(s)$ the probability that a mutation with effect size s eventually fixes, so that the total fixation rate of mutations with effect size between s and $s + ds$ is given by $NU\rho(s)p_{\text{fix}}(s)ds$. Upon fixation, the mean fitness of the population increases by an amount s; therefore, summing over all available effect sizes, the overall average rate of adaptation is given by

$$v = NU \int \rho(s)p_{\text{fix}}(s)sds \tag{2.1}$$

Our problem is essentially reduced to predicting the fixation probabilities $p_{\text{fix}}(s)$ for all relevant fitness effects s, given an arbitrary set of population-genetic parameters. A classic result in population genetics is that

$$p_{\text{fix}}(s) = \frac{2s}{1 - e^{-2Ns}} \tag{2.2}$$

provided that each mutation either fixes or goes extinct independently [11, 50]. Assum-

ing that beneficial and deleterious mutations each consist of a single effect size,

$$v = 2N\left(\frac{U_b s_b^2}{1 - e^{-2Ns_b}} - \frac{U_d s_d^2}{e^{2Ns_d} - 1}\right) \tag{2.3}$$

The (hyper-)surface of parameter space on which $v = 0$ divides the space into two regions: one in which Muller's ratchet dominates, and the other in which positive adaptation dominates. The equation for this $v = 0$ surface is

$$\frac{s_b^2}{1 - e^{-2Ns_b}} = \frac{U_d}{U_b}\frac{s_d^2}{e^{2Ns_d} - 1} \tag{2.4}$$

Regions of parameter-space with larger s_b or larger U_b values than this surface correspond to positive adaptation; regions with larger U_d values than the surface correspond to the ratchet-dominated region. Perturbing s_d by a positive amount from the surface can result in either adaptation or fitness decline, depending on the values of N, U_b, U_d, and s_b, since, as a function of s_d, the right-hand side of Equation 2.4 is non-monotonic. In absence of beneficial mutations, this non-monotonicity can be illustrated by two limiting cases. When $Ns_d \ll 1$, deleterious mutations accumulate nearly neutrally, with a probability of roughly $1/N$ of fixing. An increase in the magnitude of the deleterious effect therefore has a negligible impact on the fixation probability, and so corresponds to a more negative rate of adaptation. In the opposite limit, deleterious mutations are purged efficiently by selection, with a fixation probability that decays exponentially with Ns_d. When fixed, deleterious mutations of larger effect size s_d will impart a larger fitness cost, but this additional cost is not large enough to compensate for the exponentially reduced probability of fixing: stronger deleterious mutations correspond to a less negative rate of adaptation.

Unfortunately, Equation 2.2 has limited applicability to the identification of the $v = 0$ boundary surface. Muller's ratchet is clearly of importance on the $v = 0$ boundary surface, and prediction of the rate of Muller's ratchet over a broad range of parameters has been a topic of ongoing research [9, 18, 37]. Even in the simple case involving no beneficial mutations and deleterious mutations of a single effect size, deviations from Equation 2.2 are apparent: unless the deleterious mutation rate is much less than the deleterious effect size, deleterious load on the population will significantly reduce the size of the most-fit class, enabling stochastic fixation of deleterious mutations to occur much more frequently than would be predicted by Equation 2.2 [21]. This will likely be the case if Muller's ratchet is able to counteract even a slow rate of accumulation of beneficial mutations.

Even in the absence of deleterious mutations, significant deviations from the prediction given by Equation 2.2 can occur. Multiple beneficial mutations, of potentially different effect sizes, may segregate in a population at the same—this phenomenon is known as *clonal interference* and has been observed in several recent empirical studies [1, 23, 28, 29, 34]. Much of the adaptive potential of the population is wasted: a lineage carrying a beneficial mutation may survive genetic drift and grow to an appreciable frequency, only to be later outcompeted by another more-fit lineage [8, 41]. In a

similar vein, deleterious mutations can "hitchhike" along with beneficial mutations to fixation—this effect both increases the likelihood with which deleterious mutations fix, and decreases the likelihood with which beneficial mutations fix [44].

In general, the probability that a particular mutation fixes depends not only on the fitness effect it confers, but on the entire distribution of fitness effects $\rho(s)$ available to the population (as well as the population size N and the mutation rate U). Previous work has argued that, if the mutation rate is much larger than typical fitness effects, then

$$p_{\text{fix}}(s) \approx \frac{1}{N} e^{T_2 s/2} \tag{2.5}$$

—that is, the effects of clonal interference, deleterious load and hitchhiking affect fixation probabilities only via their influence on the coalescence timescale T_2 [40]. A simple heuristic can provide some intuition on this formula: on average, two randomly chosen individuals share a most recent common ancestor a time T_2 generations in the past. Consequently, roughly T_2 generations *after* a mutation occurs, it is likely to either be close to fixation or close to extinction. The mutation makes a significant contribution to fitness variance in the population only over the course of the pairwise coalescence timescale—after T_2 generations, each of the N extant individuals are equally likely to fix. Assuming that the mutation's lineage grows exponentially at rate s over the pairwise coalescence timescale then yields $p_{\text{fix}}(s) \approx \frac{1}{N} e^{T_2 s}$. This heuristic and Equation 2.5 differ by an $\mathcal{O}(1)$ factor of T_2 in the exponent, which is quantitatively significant, but is consistent with the degree of inaccuracy expected in (i) estimating the timescale on which mutational fates are determined as T_2 and in (ii) assuming exponential growth of the mutational lineage with a *constant* rate throughout the duration of this timescale. In Chapter 3, we will more formally justify the applicability of this formula and detail more precisely its regimes of validity; here we will assume its validity, examine the resulting implications for the $v = 0$ boundary surface and compare our predictions to results from simulations.

Inserting (2.5) into (2.1) yields an expression for $T_2 v$:

$$T_2 v = U_b T_2 s_b e^{T_2 s_b/2} - U_d T_2 s_d e^{-T_2 s_d/2} \tag{2.6}$$

We'll define scaled fitness effects $\gamma_b \equiv \frac{1}{2} T_2 s_b$ and $\gamma_b \equiv \frac{1}{2} T_2 s_d$, and a scaled rate of adaptation $\nu \equiv \frac{1}{2} T_2 v$. The scaled rate of adaptation ν can then be identified as proportional to the fitness increase per pairwise coalescence timescale, and can be written in terms of γ_b, γ_d, U_b, and U_d:

$$\nu = U_b \gamma_b e^{\gamma_b} - U_d \gamma_d e^{-\gamma_d} \tag{2.7}$$

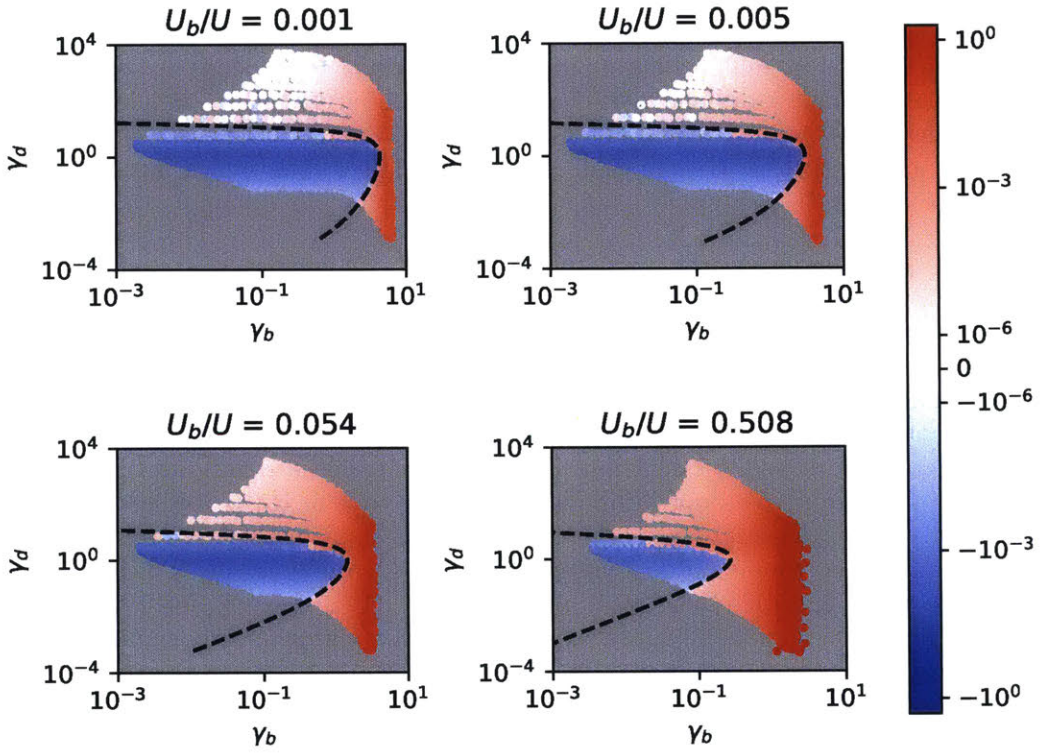

Figure 2.1: Cross sections of an "evolvability phase diagram" in the space of scaled beneficial and scaled deleterious fitness effects, with the color of each point denoting the average rate of adaptation as measured in a Wright-Fisher simulation, with $NU = 10^4$ (see Appendix A for more details on simulation methods). *Unscaled* fitness effects s_b and s_d were specified, and for each point, the pairwise heterozygosity π was measured. (This yields the pairwise coalescence timescale $T_2 = \pi/2U_n$ and thus the corresponding scaled fitness effects.) Unscaled fitness effects were chosen to make up a log-uniformly spaced grid in the space of unscaled fitness effects (see Figure 3.2 for this grid); the non-uniform density of points in the space of scaled fitness effects reflects variation in T_2 as a function of s_b and s_d. The shape of the sampled region in the space of scaled fitness effects is a consequence of the fact that T_2 decays monotonically with s_b, but is a non-monotonic function of s_d. The dashed line is the theoretical prediction for the fixed-point curve given by Equation 2.8.

From this we can construct an "evolvability phase diagram", in which regions of parameter space corresponding to positive adaptation and regions corresponding to negative adaptation can be separated by the surface on which $\nu = 0$. In Figure 2.1,

cross sections of this phase diagram are visualized in two dimensions, as a plot of γ_b vs. γ_d, for various relative frequencies of beneficial mutations to total mutations. Our equation for the phase boundary curve can be written with γ_b an explicit, single-valued function of γ_d:

$$\gamma_b = W\left(\frac{U_d}{U_b}\gamma_d e^{-\gamma_d}\right) \tag{2.8}$$

where W is the Lambert W-function defined by $W(z)e^{W(z)} \equiv z$. In Figure 2.2, predicted phase boundary curves are plotted along with boundary curves extracted from simulation results.

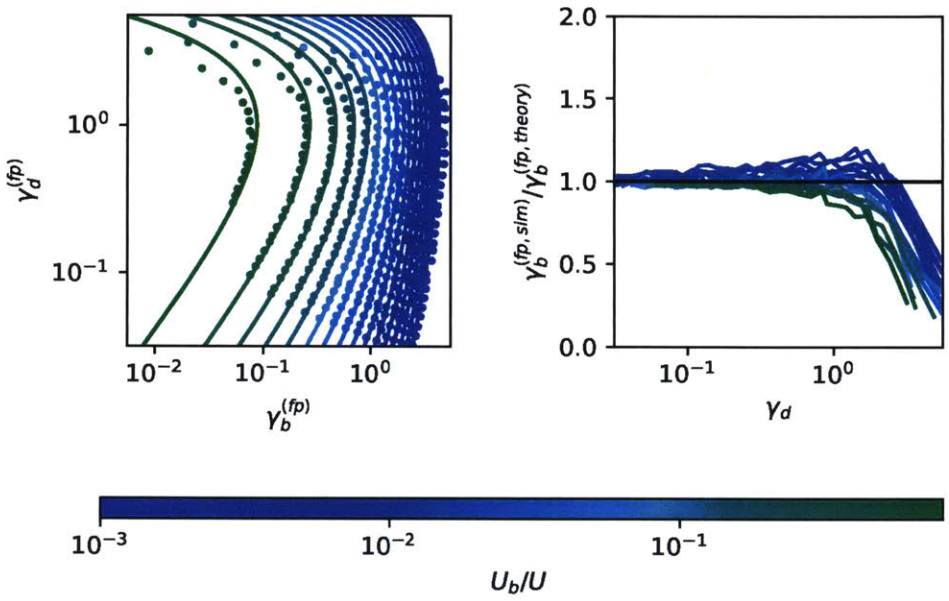

Figure 2.2: (Left) Points delineate fixed-point curves as measured from simulations, over a range of U_b/U values with $NU = 10^4$. Solid lines denote the corresponding theoretical predictions for the same set of U_b/U values. (Right) Ratio of fixed-point scaled beneficial effect as measured from simulation to the corresponding theoretically-predicted fixed-point scaled beneficial effect, over a range of scaled deleterious effects and U_b/U values, with $NU = 10^4$.

Each boundary curve attains exactly one maximum scaled beneficial effect γ_b^*, at

$$\begin{aligned} \gamma_b^* &= W\left(\frac{U_d}{eU_b}\right) \\ \gamma_d^* &= 1 \end{aligned} \tag{2.9}$$

In Figure 2.3, the location of this inflection point as measured in simulations is compared to the prediction in Equation 2.9, over a range of values of NU and U_b/U. For $\gamma_b > \gamma_b^*$, positive adaptation is predicted to occur regardless of the value of γ_d, and for $\gamma_b < \gamma_b^*$, an intermediate range of γ_d values is predicted to result in negative adaptation. The value $\gamma_d^* = 1$ can be interpreted as the scaled deleterious fitness effect that requires the largest γ_b to counteract Muller's ratchet. The fact that $\gamma_d^* = \mathcal{O}(1)$ has a simple intuitive explanation: these mutations are typically purged by selection on approximately the same timescale on which common ancestry is determined [13]. Significantly more deleterious mutations are typically driven to extinction by selection; significantly less deleterious mutations accumulate nearly-neutrally, but confer a smaller effect on the population-wide fitness. Note that this phase boundary curve has no dependence on N or U: these quantities influence the sign of the rate of adaptation only through their effect on the coalescence timescale (which generally increases with N). The boundary *does* depend on the relative frequency of deleterious mutations compared to beneficial mutations, with an enhanced relative frequency of deleterious mutations shifting boundary curves toward larger scaled beneficial fitness effects.

Note that even when $\nu \neq 0$, $\gamma_d = 1$ is the optimally deleterious scaled fitness effect in the sense that it extremizes ν for arbitrary fixed γ_b, U_b and U_d. It does *not* necessarily extremize v; estimating the value of γ_d that extremizes v can be done using the results of Chapter 3.

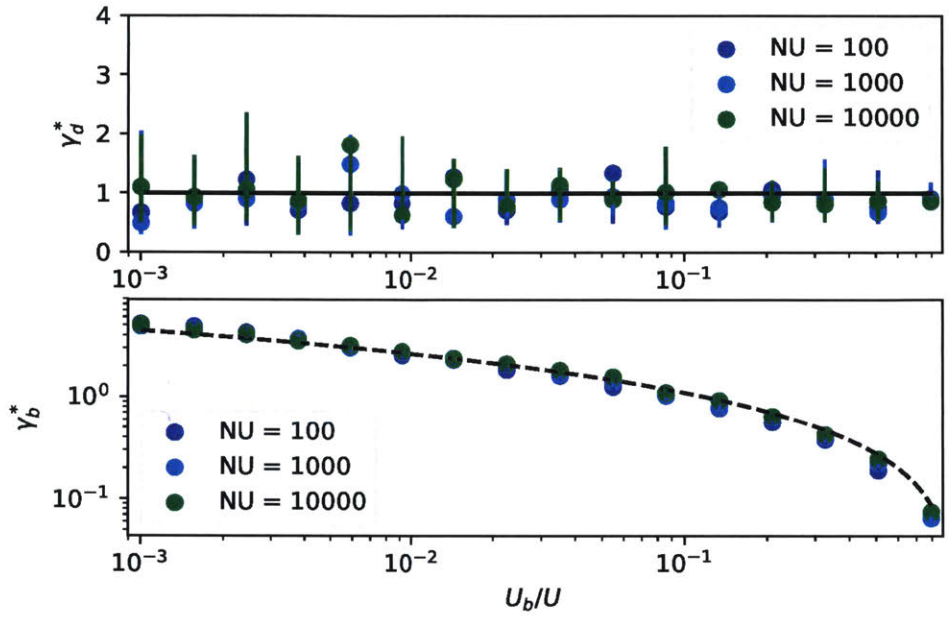

Figure 2.3: Scaling of γ_d^* (top) and γ_b^* (bottom), as measured from simulations. The dashed line in the plot of γ_b^* is the theoretical prediction. Error bars for γ_b^* encompass the range of γ_b values that could be an extreme value, assuming that true T_2 values fall within one standard error of their sample mean. Error bars for γ_d^* correspond to the range of γ_d values that correspond to a γ_b within the aforementioned range of γ_b^* values, with the inherent correlation between γ_b and γ_d taken into account. Variation in the rate of adaptation from the sample mean rate is assumed to be negligible compared to variation in the pairwise heterozygosity. (The population fitness and pairwise heterozygosity are sampled approximately once every T_{sweep} generations, a total of 192 times, where T_{sweep} denotes the time until one of the founding individuals sweeps through the population.)

Alternatively, our evolvability phase diagram can be visualized by fixing either γ_b, γ_d or the ratio $\frac{\gamma_d}{\gamma_b} = \frac{s_d}{s_b}$, and varying the relative frequencies of beneficial and deleterious mutations. Manipulating Equation 2.8, we can solve for the relative frequency of beneficial to deleterious mutations at the fixed point:

$$\frac{U_b}{U_d} = \frac{\gamma_d e^{-\gamma_d}}{\gamma_b e^{\gamma_b}} \tag{2.10}$$

From Equation 2.10, we can see that the fixed-point value of U_b/U_d is a decreasing function of γ_b, but a non-monotonic function of γ_d. For fixed γ_b, U_b/U_d is maximized

with respect to γ_d when $\gamma_d = 1$, and takes on a maximum value

$$\left(\frac{U_b}{U_d}\right)^* = \frac{1}{\gamma_b e^{\gamma_b + 1}} \tag{2.11}$$

If $U_b/U_d > \left(\frac{U_b}{U_d}\right)^*$, then the population will adapt, no matter the (scaled) strength of deleterious mutations. Simulation results corresponding to several specified values of U_b/U are collected into Figure 2.4 to test the validity of Equation 2.10.

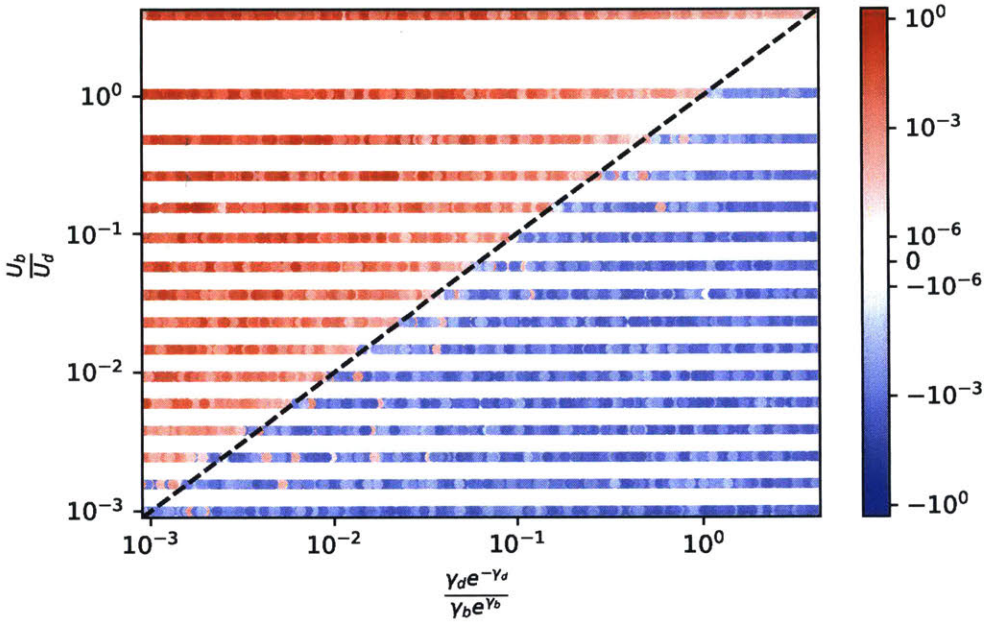

Figure 2.4: Rate of adaptation as a function of U_b/U_d and $\gamma_d e^{-\gamma_d}/\gamma_b e^{\gamma_b}$. As in Figure 2.1, unscaled fitness effects and mutation rates U_b and U_d were specified; the corresponding scaled fitness effects and the rate of adaptation were then obtained from simulations. (See Figure 3.2 for the log-uniformly spaced grid of unscaled fitness effects, which was specified identically for all values of U_b/U_d.) Results from simulations with $NU = 10^2$, $NU = 10^3$, and $NU = 10^4$ are included. The dashed line is the theoretical prediction for points lying on the $v = 0$ boundary surface, given by Equation 2.10.

The above analysis considered the evolvability phase diagram with γ_b and γ_d independent variables. We have defined γ_b and γ_d as fitness effects scaled by their corresponding coalescence rates, which depend on s_b and s_d in a potentially complicated way. We will argue in Chapter 6 that the distribution of scaled fitness effects may be more readily observable than the distribution of unscaled fitness effects, given DNA sequencing data from natural populations. However, it may be desirable to obtain the

evolvability phase diagram with unscaled fitness effects s_b and s_d as the independent variables. We will turn to this objective in the next section, using the full traveling-wave formalism of [10, 16] to solve for the $v = 0$ boundary surface.

Chapter 3

Evolutionary equations

3.1 Evolutionary equations

THERE are two key aspects of rapidly evolving populations that several recent theoretical studies have explicitly accounted for [10, 16, 22, 47]. Individuals in a population comprise a range of fitnesses, so that a *distribution of fitnesses* $f(X,t)$ can be introduced, describing the probability density for a randomly sampled individual to have an absolute fitness X at time t. Additionally, the probability that a mutation eventually fixes within a population depends on the genetic background it lands on: this can be modelled by considering the *non-extinction probability* $w(X,t)$ of a lineage founded by an individual with absolute fitness X at time t. Attention can then be focused on how, in absence of epistasis, the population-genetic parameters N, U, and $\rho(s)$ determine the time-evolution of $f(X,t)$ and $w(X,t)$ [16]. This analysis can also be applied to populations subject to fitness-mediated epistasis, provided that the assumptions stated in Section 1.1, which enable a mapping to an epistasis-free model, are met.

The dynamics of $f(X,t)$ are described by

$$\frac{\partial f(X,t)}{\partial t} = \left[X - \bar{X}(t)\right] f(X,t) + U \int \rho(s) \left[f(X-s,t) - f(X,t)\right] ds \qquad (3.1)$$

The terms of this equation have a clear interpretation as the rate of change to $f(X,t)$ as a result of selection, and as a result of mutation (both to and from a fitness X). (The mean fitness of the population at time t is denoted by $\bar{X}(t)$.) After an initial transient period, $f(X,t)$ takes on a Gaussian profile that increases (or decreases) in fitness at rate v. Upon transforming to the frame of the relative fitness $x = X - \bar{X}$, this equation becomes

$$-v\partial_x f(x) = xf(x) + U \int \rho(s) \left[f(x-s) - f(x)\right] ds \qquad (3.2)$$

A similar equation can be derived for the non-extinction probability $w(X,t)$:

$$-\frac{\partial w(X,t)}{\partial t} = \left[X - \bar{X}(t)\right] w(X,t) + U \int \rho(s) \left[w(X+s,t) - w(X,t)\right] ds - \left[1 + X - \bar{X}(t)\right] w^2(X,t) \qquad (3.3)$$

which takes on the form

$$v\partial_x w(x) = xw(x) + U\int \rho(s)\left[w(x+s) - w(x)\right] ds - \frac{w(x)^2}{2} \tag{3.4}$$

after transforming to the frame of the relative fitness [16]. In this approach, genetic drift is the source of the w^2 term; a noise term could be included in the equation for f but is neglected in obtaining the mean-field fitness distribution. In principle, these "evolutionary equations" can be solved together with the normalization conditions

$$\int f(x)dx = 1 \tag{3.5}$$

and

$$N\int f(x)w(x)dx = 1 \tag{3.6}$$

The first equation ensures that the fitness distribution is properly normalized; the second equation ensures that, at any point in time, exactly one extant individual's lineage will eventually fix.) Note the distinction between $w(x)$, which denotes the fixation probability of an individual with relative fitness x, and $p_{\text{fix}}(s)$, which denotes the fixation probability of an individual which just acquired a mutation of effect size s. The quantity $p_{\text{fix}}(s)$ can be related to $f(x)$ and $w(x)$ by averaging over initial relative fitnesses:

$$p_{\text{fix}}(s) = \int dx f(x)w(x+s) \tag{3.7}$$

Thus, provided that we can solve for $f(x)$ and $w(x)$, we can obtain v as

$$v = NU\int dx\int ds f(x)\rho(s)sw(x+s) \tag{3.8}$$

These equations have been solved in special cases, but a solution for general $U\rho(s)$ has not been found [10, 13, 16, 36]. There are two main challenges in solving these equations: the equation for $w(x)$ is *nonlinear*, and both equations are *nonlocal*. Earlier work has shown that the nonlinearity in Equation 3.4 can be neglected for x below an "interference threshold" x_c, since in this region, the fixation probability $w(x)$ is small enough so that $w(x)^2$ is negligible compared to $xw(x)$ [16, 35]. For $x > x_c$, the nonlinearity in the equation for $w(x)$ is not negligible, and $w(x)$ is determined by the balance between nonlinear and selection terms, so that $w(x) \approx 2x$: lineages *fix* provided that they *establish*, and they establish with probabilities given by neglecting clonal interference. A boundary layer around x_c exists where the behavior is more complex, but to a good approximation it is possible to simply match both the values and derivatives of the two solutions at x_c. The resulting $w(x)$ drops off rapidly below x_c, as a result of interference: lineages can go extinct even if they survive genetic drift.

The second and more central problem in analyzing the equations for $f(x)$ and $w(x)$ lies in the fact that the mutation term is nonlocal, even in the simple case involving

mutations of a single effect size. Earlier work has simplified the analysis of this nonlocality by focusing on two special cases, which represent opposite limits of the strength of selection. In the first case (the strong-selection limit), the amount of fitness variation within the population is assumed to be small compared to two quantities: the typical background fitness of a successful driver, and the fitness effect of a typical driver [13]. (Assuming that all beneficial mutations confer the same effect s_b on fitness, these conditions necessitate that $Ns_b \gg 1$ and $s_b \gg U_b$.) Provided that these conditions are met, another dominant balance approximation—neglecting the mutation term in 3.4—can be made to solve for $w(x)$ below the interference threshold x_c. In this approach, the quantity $U\rho(s)$ is used only in determining the interference threshold x_c (as well another lower cutoff x_{min}, below which $w(x)$ is assumed to vanish identically to zero). This approach has been successfully applied in modeling populations subject only to beneficial mutations [10, 16], as well as to populations subject to both beneficial and deleterious mutations [13, 15], provided that these populations maintain a steady *increase* in fitness over time.

Alternatively, the other approach to treating the nonlocality of Equations 3.2 and 3.4 is justified if mutations occur very frequently, but confer very small fitness effects [6, 13, 17, 22, 35, 36, 47]. More specifically, $\rho(s)$ is assumed to enable a well-behaved Taylor approximation of $f(x-s)$ in and $w(x+s)$ in Eq. (3.2). This approximation resembles a completely linked version of the infinitesimal model from quantitative genetics, in which variation in fitness is determined by an infinite number of infinitesimally weakly selected loci [4]. We will now review the infinitesimal limit, and then present a *generalized infinitesimal limit* which also treats the nonlocality in (3.2) and (3.4) with a Taylor approximation, but which continues to give valid results for substantially larger fitness effects.

■ 3.2 Infinitesimal limit

The infinitesimal limit handles the special case in which adaptation is the result of a large number of very weakly selected mutations. In the infinitesimal limit, we Taylor approximate $f(x-s) \approx f(x) - s\partial_x f(x) + \frac{s^2}{2}\partial_x^2 f(x)$, so that (3.2) reduces to

$$-v\partial_x f(x) = xf(x) + U\int \rho(s)\left[-s\partial_x f(x) + \frac{s^2}{2}\partial_x^2 f(x)\right] ds \tag{3.9}$$

and therefore

$$-\sigma^2\partial_x f(x) = xf(x) + \frac{U}{2}\left\langle s^2\right\rangle \partial_x^2 f(x) \tag{3.10}$$

where we made use of Fisher's fundamental theorem $v = \sigma^2 + U\left\langle s\right\rangle$, with σ^2 the variance of the fitness distribution. Let's try a solution

$$f(x) \propto e^{-ax}\psi(x) \tag{3.11}$$

with a and $\psi(x)$ to be determined. Defining $D \equiv \frac{U}{2}\left\langle s^2\right\rangle$,

$$D\left[a^2e^{-ax}\psi - 2ae^{-ax}\psi' + e^{-ax}\psi''\right] + \sigma^2\left[e^{-ax}\psi' - ae^{-ax}\psi\right] + xe^{-ax}\psi = 0 \tag{3.12}$$

We'll take $\sigma^2 = 2aD$, which simplifies the above to the Airy equation

$$D\psi'' + (x - a^2 D)\psi = 0 \tag{3.13}$$

whose solution is

$$\psi(x) = \mathrm{Ai}\left[\frac{a^2 D - x}{D^{1/3}}\right] \tag{3.14}$$

and therefore

$$f(x) \propto e^{-\frac{\sigma^2 x}{2D}} \mathrm{Ai}\left[\frac{\sigma^4}{4D^{4/3}} - \frac{x}{D^{1/3}}\right] \tag{3.15}$$

Applying the same Taylor approximation to our equation for w, we get

$$\sigma^2 \partial_x w(x) = xw(x) + D\partial_x^2 w(x) - \frac{w(x)^2}{2} \tag{3.16}$$

This equation for w cannot be solved using the modified Airy equation, since it contains a nonlinear w^2 term. We'll treat this nonlinearity using a dominant balance approximation. For $x < x_c$ (with x_c to be determined), $w(x) \ll x$ and the nonlinear term can be neglected. The resulting equation is the same as our equation for f, but with σ^2 replaced by $-\sigma^2$. It follows that below the interference threshold x_c, $w(x) = e^{ax}\phi(x)$, with $\phi(x) \propto \psi(x)$. For $x > x_c$, the dominant balance will be between $xw(x)$ and $\frac{w(x)^2}{2}$, yielding

$$w(x) \approx \begin{cases} C_1 e^{\frac{\sigma^2 x}{2D}} \mathrm{Ai}\left[\frac{\sigma^4}{4D^{4/3}} - \frac{x}{D^{1/3}}\right] & x < x_c \\ 2x & x > x_c \end{cases} \tag{3.17}$$

Above the interference threshold x_c, the fitness distribution $f(x)$ is assumed to vanish identically to zero, so that

$$f(x) \approx \begin{cases} C_2 e^{-\frac{\sigma^2 x}{2D}} \mathrm{Ai}\left[\frac{\sigma^4}{4D^{4/3}} - \frac{x}{D^{1/3}}\right] & x < x_c \\ 0 & x > x_c \end{cases} \tag{3.18}$$

The constants C_1 and x_c can be determined by enforcing the constraint that $w(x)$ and $w'(x)$ are both continuous at $x = x_c$. For convenience let's define $b \equiv D^{1/3}$ and $c \equiv \frac{\sigma^4}{4D}$. The first constraint is easy to enforce and yields

$$C_1 = 2x_c e^{-ax_c} \mathrm{Ai}^{-1}\left[\frac{c - x_c}{b}\right] \tag{3.19}$$

Equating the derivatives we get

$$1 = ax_c - \frac{x_c}{b} \frac{\mathrm{Ai}'\left[\frac{c-x_c}{b}\right]}{\mathrm{Ai}\left[\frac{c-x_c}{b}\right]} \tag{3.20}$$

We'll assume that $ax_c \gg 1$ and $ab >> 1$, and check the self-consistency of these assumption later on. It follows that the two terms on the right-hand side must approximately

cancel. Since $\mathrm{Ai}'(z) < 1$ for all $z > z_0$, the denominator of the second term on the right-hand side must approach zero. This motivates us to define $\frac{c-x_c}{b} = z_0 + \delta$, with $z_0 = -2.33$ satisfying $\mathrm{Ai}(z_0) = 0$, and expand to lowest order in δ:

$$ab \approx \frac{\mathrm{Ai}'[z_0+\delta]}{\mathrm{Ai}[z_0+\delta]} \approx \frac{\mathrm{Ai}'(z_0)}{\mathrm{Ai}(z_0)+\delta\mathrm{Ai}'(z_0)} = \delta^{-1} \tag{3.21}$$

and

$$x_c \approx c - bz_0 - 1/a \approx c = \frac{\sigma^4}{4D} \tag{3.22}$$

Finally, we can solve for σ^2, from which the rate of adaptation v can be obtained. To do so we'll use our normalization conditions

$$N = \frac{\int f(x)dx}{\int f(x)w(x)dx} \tag{3.23}$$

$$= \frac{\int_{-\infty}^{x_c} e^{-ax}\mathrm{Ai}\left[\frac{c-x}{b}\right]dx}{2x_c e^{-ax_c}\mathrm{Ai}^{-1}\left[\frac{c-x_c}{b}\right]\int_{-\infty}^{x_c}\mathrm{Ai}^2\left[\frac{c-x}{b}\right]dx} \tag{3.24}$$

Let's transform variables to $z \equiv z_0 + \frac{x_c-x}{b}$, so that

$$\frac{c-x_c}{b} = z_0 + \frac{1}{ab} \tag{3.25}$$

and

$$\frac{c-x}{b} = \frac{c-x_c}{b} + \frac{x_c-x}{b} \tag{3.26}$$

$$= \left(z_0 + \frac{1}{ab}\right) + (z - z_0) = z + \frac{1}{ab} \tag{3.27}$$

Then we have

$$N = \frac{\int f(x)dx}{\int f(x)w(x)dx} = \frac{e^{-abz_0}\mathrm{Ai}\left[z_0+\frac{1}{ab}\right]\int_{z_0}^{\infty} e^{abz}\mathrm{Ai}\left[z+\frac{1}{ab}\right]dz}{2x_c\int_{z_0}^{\infty}\mathrm{Ai}^2\left[z+\frac{1}{ab}\right]dz} \tag{3.28}$$

We want to evaluate this to lowest order in $(ab)^{-1}$. Using the result $\int \mathrm{Ai}^2(x)dx = x\mathrm{Ai}^2(x) - \mathrm{Ai}'(x)^2$, the denominator evaluates to

$$2x_c\left[\mathrm{Ai}'\left(z_0+\frac{1}{ab}\right)^2 - \left(z_0+\frac{1}{ab}\right)\mathrm{Ai}^2\left(z_0+\frac{1}{ab}\right)\right] \approx 2x_c\mathrm{Ai}'(z_0)^2 \tag{3.29}$$

so

$$N = \frac{e^{-abz_0}}{2abx_c\mathrm{Ai}'(z_0)}\int_{z_0}^{\infty} e^{abz}\mathrm{Ai}\left[z+\frac{1}{ab}\right]dz \approx \frac{e^{-abz_0}}{2abx_c\mathrm{Ai}'(z_0)}\int_{-\infty}^{\infty} e^{abz}\mathrm{Ai}(z)dz \tag{3.30}$$

(For $z < z_0$, the Airy function is oscillatory, and so we can neglect the integral of the exponentially damped Airy function from $-\infty$ to z_0.) Using $\int_{-\infty}^{\infty} e^{pt}\mathrm{Ai}(t)dt = e^{p^3/3}$ we have

$$N = \frac{e^{-abz_0+(ab)^3/3}}{2abx_c\mathrm{Ai}'(z_0)} \tag{3.31}$$

from which it follows that

$$\log(Nc) \approx \log(Nx_c) \approx \frac{(ab)^3}{3} + ab|z_0| - \log(ab) \approx \frac{(ab)^3}{3} \tag{3.32}$$

We can now solve for σ^2 in terms of underlying parameters:

$$\sigma^2 = 2aD = 2D\left(\frac{3\log Nc}{b^3}\right)^{1/3} = \left(24D^2\log Nc\right)^{1/3} \tag{3.33}$$

with

$$Nc = \frac{N\sigma^4}{4D} = \frac{N}{4D}\left(24D^2\log Nc\right)^{2/3} = 3^{2/3}ND^{1/3}\log^{2/3} Nc \tag{3.34}$$

so that

$$\sigma^2 \approx \left[24D^2\log\left(2ND^{1/3}\right)\right]^{1/3} \tag{3.35}$$

where we used $3^{2/3} \approx 2.1$ and neglected the $\log\log Nc$ contribution. It follows that

$$v = \sigma^2 + U\langle s\rangle = \sigma^2 \approx \left[24D^2\log\left(2ND^{1/3}\right)\right]^{1/3} + U\langle s\rangle \tag{3.36}$$

We can now check that the approximations we made are self-consistent. Apart from the initial Taylor approximation of $f(x-s)$ and $w(x+s)$, we assumed that $a^{-1} \ll b \ll c$. From our definitions of b and c, we can see that

$$\frac{c}{b} = \frac{\sigma^4}{4D^{4/3}} \approx 2\log^{2/3}\left(2ND^{1/3}\right) \tag{3.37}$$

and

$$ab = \frac{\sigma^2}{2D^{2/3}} \approx 1.4\log^{1/3}\left(2ND^{1/3}\right) \tag{3.38}$$

so these approximations are self-consistent provided that $ND^{1/3} \gg 1$. Additionally, we can check the self-consistency of our Taylor approximation. Roughly speaking, a Taylor approximation of $f(x-s)$ and $w(x+s)$ should be valid if the typical fitness effect s is smaller than the scale of variation of $f(x)$ and $w(x)$. Both of these functions vary on a scale of $1/a$, as can be seen by their exponential prefactors (they also vary on a scale of $b > 1/a$). So we expect that the Taylor approximation should be valid provided that

$$as = \frac{\sigma^2 s}{2D} \approx 1.8\log^{1/3}\left(2ND^{1/3}\right)\left(\frac{s}{U}\right) \ll 1 \tag{3.39}$$

Given the above constraint on $ND^{1/3}$, the validity of this approximation requires that $s \ll U$. Together, all of our assumptions are met in the limit that $s \to 0$ and $U \to \infty$, with the product $D = \frac{Us^2}{2}$ held fixed and $ND^{1/3} \gg 1$. In this case, adaptation proceeds by the accumulation of a large number of mutations with infinitesimal effect sizes (for this reason it is called the infinitesimal limit).

The constants a, b, and c defined above each have interesting evolutionary interpretations. Equation (3.9) suggests that we can interpret D as a diffusion constant. Under this interpretation, as a lineage acquires a large number of small-effect mutations, it "diffuses" through fitness space. After a time t, mutational diffusion will give rise typical mean-squared-displacement in fitness space $\langle x^2 \rangle \approx Dt$. Selection acts to purge this fitness variance on a timescale inversely related to the typical fitness differences in the population, $\tau_s \approx \frac{1}{\sqrt{\langle x^2 \rangle}}$. At steady-state, the fitness scale x_d on which mutational diffusion is important is then given by

$$x_d^2 \approx D\frac{1}{x_d} \implies x_d = D^{1/3} \equiv b \tag{3.40}$$

so we can identify b as the fitness scale corresponding to mutational diffusion.

Furthermore, in [36] Neher and Hallatschek calculate the pairwise coalescence timescale T_2 in the infinitesimal limit, obtaining the result $T_2 = \frac{\sigma^2}{D}$. This has a simple intuitive explanation: if two individuals are sampled from a population, their fitnesses are likely close to the mean of the fitness distribution. Because of the large number of individuals with fitness near the mean of the distribution, there is a low probability of a coalescence event near this mean fitness value; rather, the most recent common ancestor of the two sampled individuals likely had a relative fitness above the interference threshold x_c. The time T since their ancestors had this relative fitness is given by the fitness difference divided by the rate of fitness increase (without the mutation contribution $U\langle s \rangle$)

$$T = \frac{x_c}{\sigma^2} = \frac{\sigma^2}{4D} \tag{3.41}$$

More precisely, Neher and Hallatschek show that it takes roughly $2T$ generations for the mean fitness to catch up with the high fitness nose, after which coalescence occurs at rate $\frac{1}{2T}$ (with coalescence times after the initial delay time of $2T$ generations distributed exponentially). Using the definition $\sigma^2 = 2aD$, we can identify

$$a = \frac{T_2}{2} \tag{3.42}$$

To the extent that Equation 3.42 holds, we can justify the applicability of Equation 2.5 in the infinitesimal limit. First, let's consider a deleterious effect size, with $s < 0$. Fixation probabilities are then given by

$$p_{\text{fix}}(s) = \int f(x)w(x+s)dx = \int_{-\infty}^{x_c} e^{-ax}\psi(x)e^{a(x+s)}\phi(x+s)dx \tag{3.43}$$

and given the above results for $\phi(x)$ and $\psi(x)$,

$$\int_{-\infty}^{x_c} \psi(x)\phi(x+s)dx = C_1C_2 \int_{-\infty}^{x_c} \mathrm{Ai}\left[\frac{c-x}{b}\right] \mathrm{Ai}\left[\frac{c-(x+s)}{b}\right] dx \tag{3.44}$$

We'll quote an exact indefinite integral from [19]:

$$\begin{aligned} &\int \mathrm{Ai}\left[\alpha(u+\beta_1)\right] \mathrm{Ai}\left[\alpha(u+\beta_2)\right] du \\ &= \left[\alpha^2(\beta_1-\beta_2)\right]^{-1} \left(\mathrm{Ai}'\left[\alpha(u+\beta_1)\right] \mathrm{Ai}\left[\alpha(u+\beta_2)\right] - \mathrm{Ai}\left[\alpha(u+\beta_1)\right] \mathrm{Ai}'\left[\alpha(u+\beta_2)\right]\right) \end{aligned} \tag{3.45}$$

In applying this result, we'll identify $\alpha = -1/b$, $\beta_1 = -c$ and $\beta_2 = s - c$.

$$\begin{aligned} &\int_{-\infty}^{x_c} \mathrm{Ai}\left[\frac{c-x}{b}\right] \mathrm{Ai}\left[\frac{c-(x+s)}{b}\right] dx \\ &= \frac{b^2}{s}\left(\mathrm{Ai}\left[\frac{c-x_c}{b}\right] \mathrm{Ai}'\left[\frac{c-(x_c+s)}{b}\right] - \mathrm{Ai}'\left[\frac{c-x_c}{b}\right] \mathrm{Ai}\left[\frac{c-(x_c+s)}{b}\right]\right) \end{aligned} \tag{3.46}$$

Using the result $x_c \approx c - bz_0 - \frac{1}{a}$, this can be written

$$\begin{aligned} &\frac{p_{\mathrm{fix}}(s)}{C_1C_2e^{as}} \\ &= \frac{b^2}{s}\left(\mathrm{Ai}\left[z_0+\frac{1}{ab}\right] \mathrm{Ai}'\left[z_0+\frac{1}{ab}-\frac{s}{b}\right] - \mathrm{Ai}'\left[z_0+\frac{1}{ab}\right] \mathrm{Ai}\left[z_0+\frac{1}{ab}-\frac{s}{b}\right]\right) \end{aligned} \tag{3.47}$$

$$\approx \frac{b^2}{s}\mathrm{Ai}'(z_0)^2\left(\frac{1}{ab} - \frac{1-as}{ab}\right) = b\mathrm{Ai}'(z_0)^2 \tag{3.48}$$

where we've used $s/b \ll 1$ and $ab \gg 1$. In particular, for $s = 0$, we have

$$\frac{1}{N} = p_{\mathrm{fix}}(0) = C_1C_2b\mathrm{Ai}'(z_0)^2 \tag{3.49}$$

It follows that Equation 2.5 is satisfied in the infinitesimal limit, for deleterious mutations with $s < 0$. Justifying Equation 2.5 when $s > 0$ is a bit more complicated because Equation 3.43 includes a contribution from the region $x_c - s < x < x_c$, where $w(x+s)$ is linear. However, in the infinitesimal limit, this region is particularly narrow (relevant fitness effects are much weaker than the interference threshold), and this extra contribution can be neglected, so Equation 2.5 is also valid when $s > 0$. We emphasize that Equation 2.5 has been justified in the infinitesimal limit, in which $T_2 \approx \sigma^2/D$, but that Equation 2.5 could, at least in principle, still hold in a parameter regime in which T_2 takes on a different functional form. Investigating the applicability of Equation 2.5 to regions of parameter space outside the infinitesimal limit is one motivation for developing the *generalized infinitesimal limit*, which is discussed in the next section.

■ 3.3 Generalized infinitesimal limit

In the derivation of the infinitesimal limit, we assumed that the scale of variation of both $f(x)$ and $w(x)$ was much larger than the scale of fitness effects in the support of $\rho(s)$, enabling Taylor approximations of $f(x-s)$ and $w(x+s)$ inside (3.2) and (3.4). We later found that both $f(x)$ and $w(x)$ contained exponential prefactors that varied on a scale of $1/a \propto 1/T_2$. This suggests that the Taylor approximation used in the infinitesimal limit may become invalid when fitness effects with $T_2 s \approx 1$ are included in the support of $\rho(s)$. Earlier we found that the inflection point of the boundary surface of our evolvability phase diagram is located at $T_2 s_d = 2$; we'd therefore like to modify our approximation to better handle fitness effects of this magnitude.

Motivated by the infinitesimal limit, let's define

$$f(x) = e^{-ax}\psi(x), \; w(x) = e^{ax}\phi(x) \tag{3.50}$$

In the infinitesimal limit, $a \equiv \frac{\sigma^2}{2D}$. This definition was chosen so that, after Taylor approximating $f(x)$ and $w(x)$ inside of (3.2) and (3.4), $\phi(x)$ and $\psi(x)$ would obey the Airy equation. Here, let's instead define a by

$$U\int \rho(s)se^{as}ds \equiv v = NU\int \rho(s)sp_{\text{fix}}(s)ds \tag{3.51}$$

Written in terms of $\psi(x)$ and $\phi(x)$, (3.2) and (3.4) take the form

$$-v\partial_x\psi(x) = (x-va)\psi(x) + U\int \rho(s)\left[e^{as}\psi(x-s) - \psi(x)\right]ds \tag{3.52}$$

$$v\partial_x\phi(x) = (x-va)\phi(x) + U\int \rho(s)\left[e^{as}\phi(x+s) - \phi(x)\right]ds \tag{3.53}$$

Defining

$$M_p = \frac{U}{p!}\int \rho(s)s^p e^{as}ds \tag{3.54}$$

and Taylor approximating $\psi(x-s)$ and $\phi(x+s)$ to 2nd order, we have

$$0 = \left[x - va - U + M_0\right]\psi(x) + M_2\partial_x^2\psi(x) \tag{3.55}$$

$$0 = \left[x - va - U + M_0\right]\phi(x) + M_2\partial_x^2\phi(x) \tag{3.56}$$

where the $\partial_x\psi(x)$ and $\partial_x\phi(x)$ terms vanish because of a cancellation of v and M_1. Evidently $\phi(x)$ and $\psi(x)$ differ only by a constant factor, so let's focus on $\phi(x)$. Defining

$$D_{gm} \equiv b^3 \equiv M_2 = \frac{U}{2}\int \rho(s)s^2e^{as}ds \tag{3.57}$$

$$c \equiv va + U - M_0 = U\int \rho(s)\left[1 + (as-1)e^{as}\right]ds \tag{3.58}$$

we have

$$0 = (x - c)\phi(x) + b^3\partial_x^2\phi(x) \tag{3.59}$$

The solution to this Airy equation is then

$$\phi(x) \propto \mathrm{Ai}\left[\frac{c-x}{b}\right] \tag{3.60}$$

Now, the scale of variation of both $\phi(x)$ and $\psi(x)$ is seen to be b, making our Taylor approximation valid so long as the relevant fitness effects are smaller than b. In summary we find that

$$w(x) \approx \begin{cases} C_1 e^{ax}\mathrm{Ai}\left[\frac{c-x}{b}\right] & x < x_c \\ 2x & x > x_c \end{cases} \tag{3.61}$$

$$f(x) \approx \begin{cases} C_2 e^{-ax}\mathrm{Ai}\left[\frac{c-x}{b}\right] & x < x_c \\ 0 & x > x_c \end{cases} \tag{3.62}$$

Our equations for f and w are identical to what was obtained in the infinitesimal limit, so we can import those results from above, with the exception that in the infinitesimal limit, $a \equiv \frac{\sigma^2}{2D}$, and in our generalized infinitesimal limit, a is defined by $U\int \rho(s)se^{as}ds \equiv v$. In particular, we find

$$C_1 = 2x_c e^{-ax_c}\mathrm{Ai}^{-1}\left[\frac{c-x_c}{b}\right] \tag{3.63}$$

$$x_c \approx c - bz_0 - 1/a \approx c \tag{3.64}$$

$$\log(Nc) \approx \log(Nx_c) \approx \frac{(ab)^3}{3} + ab|z_0| - \log(ab) \approx \frac{(ab)^3}{3} \tag{3.65}$$

and, as in the infinitesimal limit, we assumed $ab \gg 1$ and $b \ll c$. Using our expressions for b and c, we can rewrite 3.65 as follows:

$$a = \frac{6\log\left(NU\int \rho(s)\left[1 + (as-1)e^{as}\right]ds\right)}{U\int \rho(s)(as)^2 e^{as}ds} \tag{3.66}$$

Assuming that N and $U\rho(s)$ are both known population-genetic parameters, Equation 3.66 can be solved numerically for a, which can be used to extract the rate of adaptation v from Equation 3.51. As discussed in Section 3.4, analytical results for fixed-point curves can also be obtained from Equation 3.66. A more accurate solution can be obtained by using Equations 3.20 and 3.24, along with the definitions for b and c in the generalized infinitesimal limit, to solve numerically for a. The definition (3.51) is chosen to exploit the exponential dependence of fixation probabilities on $T_2 s$ (see equation (2.5)) in ensuring that the associated $\phi(x)$ and $\psi(x)$ vary on a scale larger than relevant fitness effects. In fact, we can see that consistency of the above definition

with (2.5) is achieved if a is identified as $T_2/2$ and the function $\phi(x)$ varies on a scale larger than relevant fitness effects:

$$p_{\text{fix}}(s) = \int f(x)w(x+s)dx = \frac{e^{as}}{N}\frac{\int \psi(x)\phi(x+s)dx}{\int \psi(x)\phi(x)dx} \tag{3.67}$$

This suggests that a Taylor approximation of $\psi(x-s)$ and $\phi(x+s)$ inside of the relevant integral equations may be valid. In Figure 3.1, we compare our theoretical prediction for T_2 (which is just $2a$, with a obtained by the numerical solution described above) with T_2 obtained from simulations, over a broad range of beneficial and deleterious fitness effects.

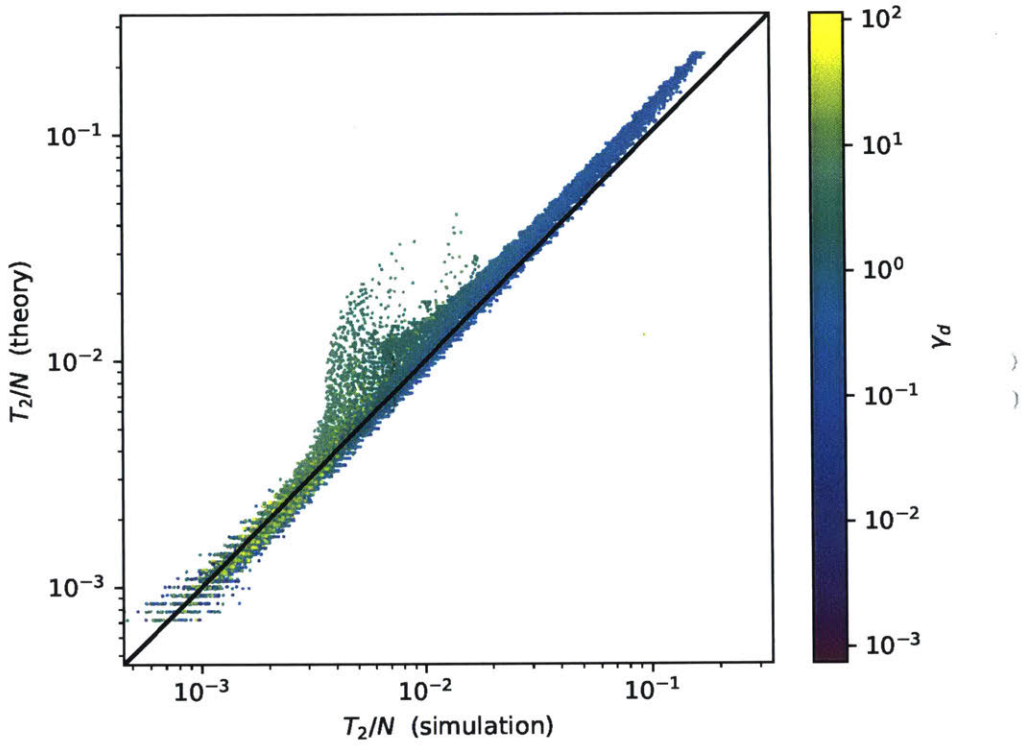

Figure 3.1: Predicted T_2/N vs. simulated T_2/N, over a grid of s_b and s_d values with $NU = 10^4$. Beneficial fitness effects ranged from $s_b = 1/N$ to $s_b = 0.5$; deleterious fitness effects ranged from $s_d = 1/N$ to $s_d = 3$. For some combinations of s_b and s_d, the numerical solution obtained is inconsistent with the initial assumption that both $s_d < b$ and $s_b < b$; these points are omitted. Color denotes the corresponding simulated value of γ_d, which is found to be an reasonably good predictor of substantial deviations of simulated values from predicted values.

■ 3.4 Evolvability phase diagrams in the generalized infinitesimal limit

We will now use the solution to the evolutionary equations in the generalized infinitesimal limit to obtain the $v = 0$ boundary surface in the parameter space with unscaled fitness effects. For now we will confine our attention to populations subject to beneficial mutations of a single effect s_b and deleterious mutations of a single effect $-s_d$; later we will extend this analysis to distributions of fitness effects. From our definition of a in

(3.51) we know that on the $v = 0$ surface

$$as_b = \frac{s_b}{s_b + s_d} \log\left(\frac{U_d s_d}{U_b s_b}\right) \tag{3.68}$$

Along with (3.65) we can solve for s_b and s_d in terms of U_b, U_d and $\delta \equiv s_d/s_b$:

$$s_b = \frac{U_b}{6(1+\delta)^2 \log Nc} \log^3\left(\frac{\delta U_d}{U_b}\right)\left(\frac{\delta U_d}{U_b}\right)^{1/(1+\delta)} \tag{3.69}$$

$$s_d = \frac{\delta U_b}{6(1+\delta)^2 \log Nc} \log^3\left(\frac{\delta U_d}{U_b}\right)\left(\frac{\delta U_d}{U_b}\right)^{1/(1+\delta)} \tag{3.70}$$

with c also a function of U_b, U_d and δ:

$$c = U - U_b\left(\frac{1+\delta}{\delta}\right)\left(\frac{\delta U_d}{U_b}\right)^{\frac{1}{1+\delta}} \tag{3.71}$$

Thus, for any particular N, U_b, and U_d, by varying δ we can generate the boundary curve of the evolvability phase diagram in s_d vs. s_b space. For better quantitative agreement with simulations, the numerical solution described in the previous section can be used (and can be simplified using Equation 3.68). For small beneficial fitness effects and large deleterious fitness effects, we find that the generalized infinitesimal limit breaks down. (The generalized infinitesimal limit may be more applicable at equally large beneficial fitness effects and small deleterious fitness effects because *scaled* fitness effects are smaller in this region of parameter-space.) We find that, in all cases considered, our numerically-obtained fixed-point curves attained a maximum deleterious effect, and that s_d exceeded the corresponding parameter b at this extremum, confirming that the generalized infinitesimal limit is not applicable beyond this point. To address this, we "cut off" our fixed-point curves at this extremum and extended the curves linearly to the s_d axis. In Figure 3.2 we plot cross sections of our phase diagram in the space of unscaled fitness effects, along with our theoretical prediction for the fixed-point curve.

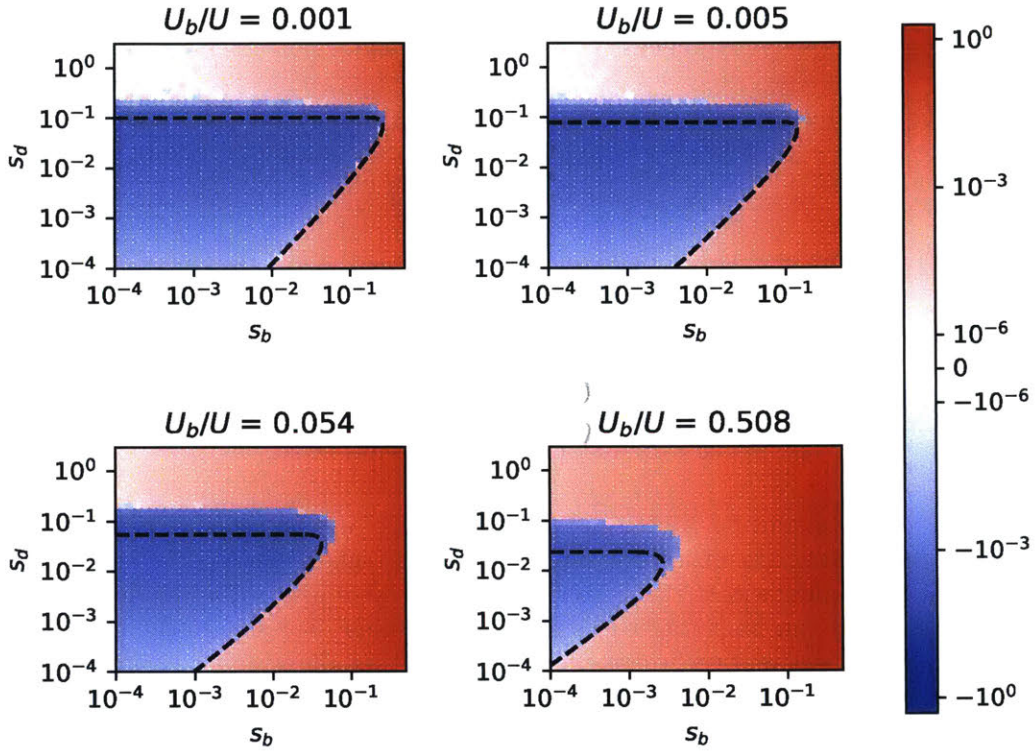

Figure 3.2: Cross sections of an "evolvability phase diagram" in the space of unscaled beneficial and unscaled deleterious fitness effects, with the color of each point denoting the average rate of adaptation as measured in a Wright-Fisher simulation, with $NU = 10^4$. The dashed line is the theoretical prediction for the fixed-point curve given by the numerical solution outlined in Section 3.3.

An asymptotic expression for s_d as a function of s_b on the fixed-point curve can be obtained in the limiting case that $\delta \gg 1$. To leading order, the fixed-point curve is given by

$$s_d = \frac{U_b s_b}{U_d} \exp\left[-\frac{3}{2} W(z_g)\right] \tag{3.72}$$

with

$$z_g \equiv -\frac{2}{3}\left(\frac{6 U_b s_b \log N U_d}{U_d^2}\right)^{1/3} \tag{3.73}$$

The Lambert W-function $W(z)$ attains two real values for $-1/e < z < 0$; in our case, the W_0 branch corresponds to the portion of the fixed-point curve with weaker deleterious

effects and the W_{-1} branch corresponds to the portion of the fixed-point curve with stronger deleterious effects. The inflection point (s_b^*, s_d^*) is given by the intersection of the two branches, where $z_g = -1/e$; this point is located at

$$\frac{s_b^*}{U_b} = \frac{9}{16e^3 \log NU_d} \left(\frac{U_d}{U_b}\right)^2 \tag{3.74}$$

$$\frac{s_d^*}{U_d} = \frac{9}{16e^{3/2} \log NU_d} \tag{3.75}$$

Note that $s_d^*/s_b^* = \frac{U_b}{U_d} e^{3/2}$, so this result for the location of the inflection point is consistent with the approximation $\delta \gg 1$ when $U_b > \frac{U_d}{e^{3/2}}$.

An alternative way to visualize the boundary surface is to consider fixed fitness effects s_b and $-s_d$, which need not be the same, as well as a fixed total mutation rate U, and then identify the ratio $\epsilon \equiv U_b/U$ of beneficial mutations to total mutations at which $v = 0$. There are three relevant cases, which we will consider separately: $U_b \ll U_d$ (mostly deleterious mutations), $U_b \approx U_d$, and $U_b \gg U_d$ (mostly beneficial mutations). In the case that $U_b \ll U_d$,

$$\epsilon \approx \frac{s_d}{s_b} \exp\left[\frac{3(s_b + s_d)}{s_d} W(z)\right] \tag{3.76}$$

with

$$z \equiv -\frac{1}{3}\left(\frac{6s_d^2 \log Nc}{U(s_b + s_d)}\right)^{1/3} \tag{3.77}$$

Alternatively, if beneficial and deleterious mutations occur at nearly the same rate ($U_b \approx U_d$), then

$$\epsilon \approx \frac{1}{4}\left[3 - \frac{s_b}{s_d} \exp\left[\frac{6(s_b + s_d)}{s_b - s_d} W(z)\right]\right] \tag{3.78}$$

with

$$z = \frac{s_b - s_d}{3s_b}\left(\frac{9s_b^5 \log^2(Nc)}{4s_d(s_b + s_d)^2 U^2}\right)^{1/6} \tag{3.79}$$

In the special case that $s_b = s_d \equiv s$ and $Ns \gg U/s$, this can be shown to reproduce the result

$$\epsilon \approx \frac{1}{2} - \left(\frac{3}{4\lambda} \log Ns\right)^{1/3} \tag{3.80}$$

given in [20] for the fast-ratchet regime ($Nse^{-U/s} \ll 1$). Finally, if most of the mutations are beneficial, then

$$\epsilon \approx 1 - \frac{s_b}{s_d} \exp\left[\frac{3(s_b + s_d)}{s_b} W(z)\right] \tag{3.81}$$

with

$$z = \frac{1}{3}\left(\frac{6s_b^2 \log Nc}{U(s_b + s_d)}\right)^{1/3} \tag{3.82}$$

In summary, we find that for any beneficial effect s_b and deleterious effect s_d, there is a fixed-point ϵ at which the rate of fitness increase from beneficial mutations exactly cancels the rate of fitness decline due to Muller's ratchet. For fixed NU and s_d, this ratio ϵ is a decreasing function of s_b—if the strength of beneficial mutations is increased, a smaller ratio of beneficial to total mutations is necessary to ensure positive adaptation. However, we find that, with s_b held fixed, ϵ is *not*, in general, a monotonically increasing (or decreasing) function of s_d. For a particular s_b, across a broad range of parameters, an intermediate value of s_d maximizes ϵ—we denote this maximum ϵ by $\epsilon_c(s_b)$. $\epsilon_c(s_b)$ can be interpreted as the ratio of beneficial mutations to total mutations that is necessary to guarantee positive adaptation across all possible deleterious effect sizes (provided that beneficial mutations confer a fitness effect s_b, and mutations occur at a population-wide rate NU). In Figure 3.3 we compare our theoretical predictions for ϵ as well as ϵ_c to results obtained from simulations.

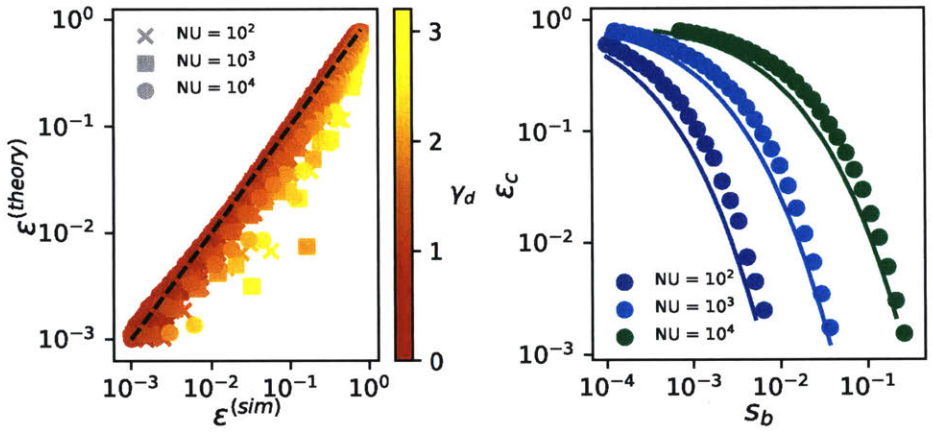

Figure 3.3: (Left) Predicted fixed-point ϵ vs. simulated fixed-point ϵ, over a grid of s_b and s_d values at $NU = 10^2$, $NU = 10^3$, and $NU = 10^4$. For each combination of s_b and s_d, the simulated ϵ is obtained by interpolating between the largest simulated ϵ with a negative rate and the lowest simulated ϵ with a positive rate. The corresponding prediction for the fixed-point ϵ is obtained using the numerical solution outlined in the previous section. For some combinations of s_b and s_d, simulations over the entire specified range of ϵ values all resulted in $v > 0$ or all resulted in $v < 0$; these choices of s_b and s_d are omitted from the figure. (Right) ϵ_c as a function of s_b for a range of values of NU. For a particular s_b, ϵ_c is defined as the minimum ϵ above which $v > 0$ regardless of s_d. Curves denote predicted values (using numerical solution) and points denote simulation results.

Chapter 4

Distributions of fitness effects

THESE predictions can be extended to account for distributions of fitness effects in a straightforward way. We will first discuss the evolvability phase diagrams with fitness effects scaled by coalescence rates, for which results are much simpler analytically. We will then proceed to make use of the travelling-wave formalism to compute pairwise coalescence times and obtain phase diagrams in s_d vs. s_b space.

■ 4.1 Evolvability phase diagrams in the space of scaled fitness effects

We'll first assume that effect sizes of beneficial and deleterious mutations are both drawn from (possibly distinct) gamma distributions. Let $\alpha_b > 0$ and $\beta_b > 0$ be the shape and rate parameters, respectively, of the distribution of beneficial fitness effects. Similarly define parameters α_d and β_d as shape and rate parameters of the distribution of deleterious fitness effects. We then have

$$U\rho(s) = \begin{cases} U_d \frac{\beta_d^{\alpha_d}}{\Gamma(\alpha_d)} |s|^{\alpha_d - 1} e^{-\beta_d |s|} & s < 0 \\ U_b \frac{\beta_b^{\alpha_b}}{\Gamma(\alpha_b)} |s|^{\alpha_b - 1} e^{-\beta_b |s|} & s > 0 \end{cases} \tag{4.1}$$

We'll gain the most intuition about distributions by eliminating rate parameters β_b and β_d, treating shape parameters and mean effect sizes as the independent parameters. (At a fixed shape parameter α, changing the mean of a gamma distribution is just a scale transformation. This is reflected by the fact that the inverse of the "rate" parameter β is also referred to as the "scale" parameter. At fixed mean effect size, decreasing the shape parameter amounts to broadening the DFE.) Additionally, throughout the rest of this section we'll denote $\frac{1}{2} T_2 \langle s_b \rangle$ by γ_b and likewise for γ_d.

The expected fitness increase per coalescence timescale is proportional to the scaled rate of adaptation $\nu \equiv \frac{1}{2} T_2 v$. Assuming that fixation probabilities are related to the coalescence timescale as in (2.5),

$$\nu = \frac{U_b \gamma_b}{\left(1 - \frac{\gamma_b}{\alpha_b}\right)^{\alpha_b + 1}} - \frac{U_d \gamma_d}{\left(1 + \frac{\gamma_d}{\alpha_d}\right)^{\alpha_d + 1}} \tag{4.2}$$

provided that $\gamma_b < \alpha_b$. If $\gamma_b > \alpha_b$, then (2.5) must be violated at least in some portion of the support of $\rho(s)$—otherwise ν would diverge. Equating the right-hand

side of Equation 4.2 to zero predicts the phase boundary surface, for the case in which beneficial mutations and deleterious mutations confer gamma-distributed effect sizes. These predictions, along with phase-diagram cross sections obtained from simulations, are displayed in Figure 4.1.

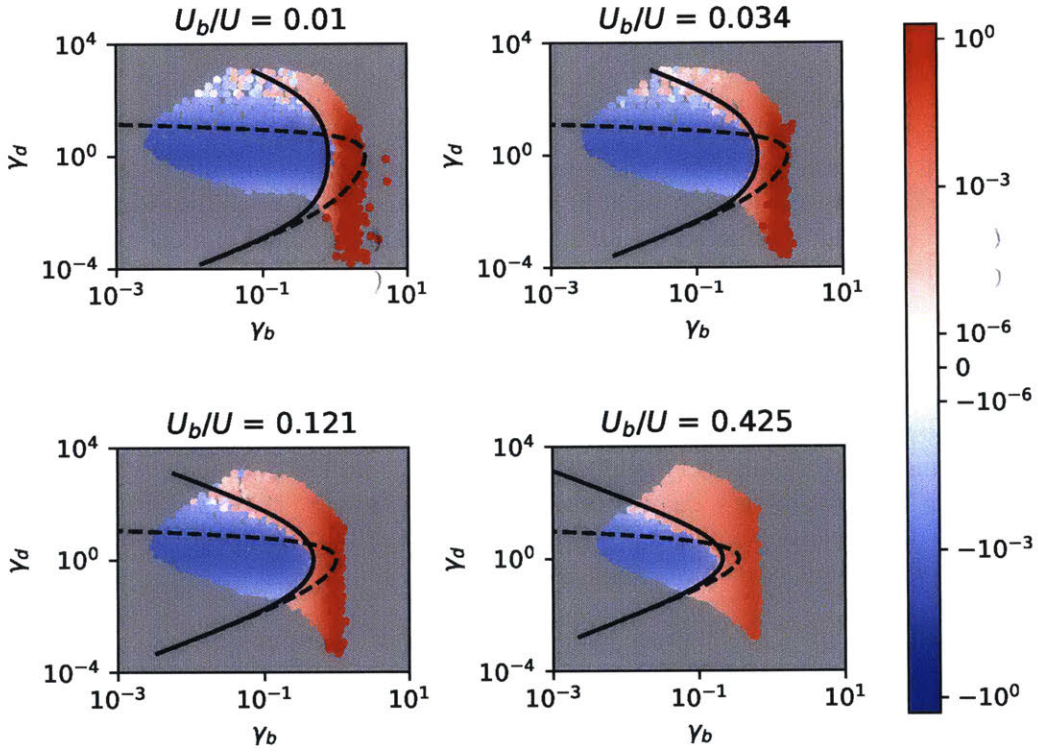

Figure 4.1: Cross sections of an evolvability phase diagram, in the space of scaled fitness effects, for gamma-distributed beneficial and deleterious mutations, with $\alpha_b = \alpha_d = 1$ (exponential distributions). The solid lines are the fixed-point curve predictions that account for the finite shapes of the distributions of fitness effects, given by equating the right-hand side Equation 4.2 to zero; the dashed lines are the theoretical predictions given by Equation 2.8 for a single scaled beneficial effect size and a single scaled deleterious effect size. The population-wide mutation rate was specified as $NU = 10^4$.

It will be useful to define ν_b as the scaled rate of adaptation due to beneficial mutations, and ν_d as the scaled rate of (negative) adaptation due to deleterious mutations, so that $\nu \equiv \nu_b - \nu_d$. The quantity ν_d is maximized with respect to γ_d at $\gamma_d = 1$, and phase boundary curves in (γ_b, γ_d) space (with fixed mutation rates and distributional

shapes) possess a single inflection point located at (γ_b^*, γ_d^*), which is determined by

$$\frac{\gamma_b^*}{\left(1 - \frac{\gamma_b^*}{\alpha_b}\right)^{\alpha_b+1}} = \frac{U_d}{U_b}\left(\frac{\alpha_d}{\alpha_d + 1}\right)^{\alpha_d+1} \tag{4.3}$$

$$\gamma_d^* = 1 \tag{4.4}$$

We can see that the result $\gamma_d^* = 1$ for single-effect distributions is insensitive with respect to broadening either the beneficial DFE or deleterious DFE (so long as the resulting DFEs are within the family of gamma distributions), while the corresponding result for γ_b^* is not. Note that the predicted value of γ_b^*, and hence all values of γ_b along the fixed-point curve, satisfy $\gamma_b < \alpha_b$, a necessary condition for the validity of Equation 4.2. In Figure 4.2, the location of the inflection point, as obtained from simulations, is compared to the predictions in Equation 4.3 and Equation 4.4, over a range of values of U_b/U and population-wide mutation rates NU.

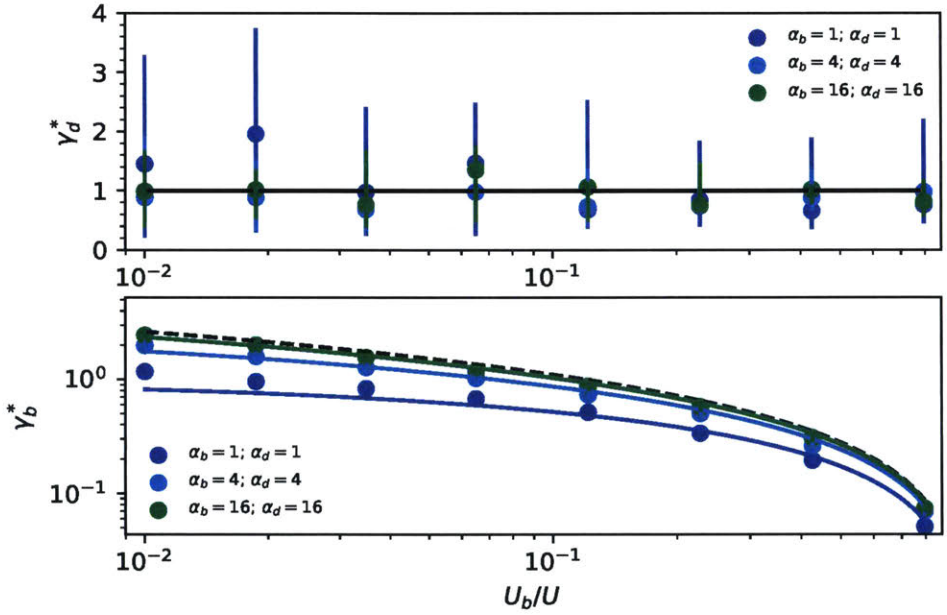

Figure 4.2: Scaling of γ_b^* and γ_d^* as a function of U_b/U. Both beneficial and deleterious effect sizes were drawn from gamma distributions, with various choices of the shape parameters α_b and α_d. The population-wide mutation rate was specified as $NU = 10^4$. Points are obtained from simulations; solid curves denote theoretical predictions given by Equations 4.3 and 4.4.

Simplifications to the equations defining fixed-point curves and γ_b^* can be made in a few special cases. If beneficial effect sizes are exponentially distributed ($\alpha_b = 1$), then

the fixed-point curve is given by

$$\gamma_b(\gamma_d) = 1 - \frac{1}{2\xi(\gamma_d)}\left(\sqrt{1+4\xi(\gamma_d)} - 1\right) \tag{4.5}$$

and γ_b^* is given by $\gamma_b(\gamma_d)$ evaluated at $\gamma_d = 1$, where

$$\xi(\gamma_d) \equiv \frac{U_d}{U_b}\frac{\gamma_d}{\left(1+\frac{\gamma_d}{\alpha_d}\right)^{\alpha_d+1}} \tag{4.6}$$

In particular, if both beneficial and deleterious effect sizes are exponentially distributed, then

$$\gamma_b^* = \frac{1-\sqrt{\frac{U_b}{U}}}{1+\sqrt{\frac{U_b}{U}}} \tag{4.7}$$

(A larger relative frequency of beneficial mutations to deleterious mutations entails that a weaker scaled beneficial effect is necessary to ensure positive adaptation across all scaled deleterious effects—this intuitive fact is reflected in the fact that γ_b^* is a monotonically decreasing function of U_b/U.)

Finally, if beneficial mutations only confer a single effect ($\alpha_b \to \infty$), but deleterious mutations are gamma-distributed with an arbitrary shape parameter α_d, then the fixed-point curve is given by

$$\gamma_b(\gamma_d) = W\left[\frac{U_d}{U_b}\frac{\gamma_d}{\left(1+\frac{\gamma_d}{\alpha_d}\right)^{\alpha_d+1}}\right] \tag{4.8}$$

and the inflection point γ_b^* by

$$\gamma_b^* = W\left[\frac{U_d}{U_b}\left(\frac{\alpha_d}{\alpha_d+1}\right)^{\alpha_d+1}\right] \tag{4.9}$$

Now suppose one of the parameters is perturbed away from the $v = 0$ surface: does that region correspond to adaptation, or to fitness decline? Examining (4.2) it is clear that perturbing U_b/U_d above the $v = 0$ surface corresponds to positive adaptation; negative perturbations to U_b/U_d result in fitness decline. Similarly, ν_b is an increasing function of γ_b, provided that $\gamma_b < \alpha_b$, so the region in which γ_b is larger than the $v = 0$ surface corresponds to positive adaptation and the region in which γ_b is less than the $v = 0$ surface corresponds to negative adaptation. For each $\gamma_b < \gamma_b^*$, an intermediate range of γ_d values will correspond to negative adaptation; if $\gamma_b > \gamma_b^*$ then $v > 0$ regardless of the value of γ_d. All of this matches the qualitative results obtained for single-effect distributions.

We can also analyze whether perturbations of α_b and α_d away from the boundary surface result in positive or negative adaptation. We can see that ν_b is a decreasing

function of α_b: for a given γ_b, broadening the distribution of beneficial effect sizes always increases the (scaled) rate of adaptation. Fixation probabilities of beneficial mutations in the large-effect tail are sufficiently amplified to compensate for the larger prevalence of weak-effect mutations.

The behavior of ν_d under changes in α_d is qualitatively different than the behavior of ν_b under changes in α_b. For $\gamma_d < 2$, ν_d is an increasing function of α_d—the delta-distribution is the gamma distribution that maximizes fitness decline. For $\gamma_d > 2$, a gamma distribution of finite shape α_d^* maximizes fitness decline, with α_d^* determined by

$$\gamma_d = \alpha_d^* \left(\exp \left[W_0(-\frac{\alpha_d^*+1}{\alpha_d^*} e^{-\frac{\alpha_d^*+1}{\alpha_d^*}}) + \frac{\alpha_d^*+1}{\alpha_d^*} \right] - 1 \right) \tag{4.10}$$

We can interpret this in the following way: when the mean deleterious effect is small enough to fix relatively frequently (with a fixation probability larger than $\frac{1}{e^2 N}$) the maximum contribution to ν_d is achieved when all deleterious mutations confer that same effect. As the mean deleterious effect becomes large enough so that it would fix extremely rarely (compared to a neutral mutation) the maximum contribution to ν_d is achieved for a broader distribution, so that a larger fraction of deleterious mutations have a reasonable chance of fixing. From equation (4.10) it is also possible to see that α_d^* is a decreasing function of γ_d (and in fact rapidly decays to zero as $\gamma_d \to \infty$) so that for substantially large values of γ_d, broadening the distribution of deleterious effects always decreases the overall rate of adaptation. A fixed-point curve can be drawn through (α_b, α_d) space with γ_b, γ_d and other parameters held fixed; this curve would have a single inflection point if $\gamma_d > 2$ and no inflection point if $\gamma_d < 2$.

Above we considered fixed-point curves and inflection points in spaces spanned by scale transformations—(γ_b, γ_d) space—and shape transformations—(α_b, α_d) space—of the beneficial and deleterious DFEs, under the assumption that both were gamma distributions. We will now generalize some of our results regarding scale transformations to arbitrary DFEs. Again defining $\nu_b - \nu_d \equiv \nu \equiv \frac{1}{2} T_2 v$ but here also defining the variable $\gamma = \frac{1}{2} T_2 s$ and distributions of *scaled* fitness effects $\tilde{\rho}(\gamma) d\gamma = \rho[s(\gamma)] ds$ we have

$$\nu = U_b \int_0^\infty \tilde{\rho}_b(\gamma) \gamma e^{\gamma} d\gamma - U_d \int_0^\infty \tilde{\rho}_d(-\gamma) \gamma e^{-\gamma} d\gamma \tag{4.11}$$

Suppose that (normalization-preserving) scale transformations are performed on $\tilde{\rho}_b(\gamma)$ and $\tilde{\rho}_d(\gamma)$ so that $\tilde{\rho}_b(\gamma) \to C_b^{-1} \tilde{\rho}_b(C_b^{-1}\gamma)$ and $\tilde{\rho}_d(\gamma) \to C_d^{-1} \tilde{\rho}_d(C_d^{-1}\gamma)$—note that $C_b > 1$ or $C_d > 1$ denotes a transformation that *increases* the scale of beneficial effects or deleterious effects, respectively. Unsurprisingly, ν_b increases with the scale of the beneficial DFE C_b. The deleterious contribution to the rate of adaptation ν_d depends non-monotonically on the scale of the deleterious DFE, and is extremized by the C_d such that

$$\int_0^\infty C_d^{-1} \tilde{\rho}_d(-C_d^{-1}\gamma) e^{-\gamma} \gamma(\gamma - 1) d\gamma = 0 \tag{4.12}$$

Distributions of deleterious fitness effects that extremize ν with respect to scale transformations satisfy the following property: the exponentially-damped mean-squared effect size equals the exponentially-damped mean effect size (with the damping coefficient proportional to the pairwise coalescence timescale T_2). It's possible to check that this result is consistent with our result in Equation 2.9 for single-effect distributions as well as our result in Equation 4.4 for gamma distributions.

■ 4.2 Evolvability phase diagrams in the space of unscaled fitness effects

Here we will provide simplified expressions for the moments $M_p \equiv \frac{U}{p!}\int \rho(s)s^p e^{as}ds$ along the fixed-point surface, in the case that both beneficial and deleterious effect sizes are drawn from gamma distributions. We will use these moments to obtain the fixed-point surfaces, focusing in particular on the case in which beneficial and deleterious DFEs have the same shape parameter (but not necessarily the same scale parameter). Provided that $v = 0$,))

$$M_p = \frac{U_b}{p!}\beta_b^{\alpha_b}\left[\frac{\Gamma(\alpha_b+p)}{\Gamma(\alpha_b)}\frac{1}{(\beta_b-a)^{\alpha_b+p}} + (-1)^p\frac{\alpha_b}{\alpha_d}\frac{\Gamma(\alpha_d+p)}{\Gamma(\alpha_d)}\frac{(\beta_d+a)^{1-p}}{(\beta_b-a)^{\alpha_b+1}}\right] \tag{4.13}$$

With the additional assumption that $\alpha_b = \alpha_d \equiv \alpha$, we are able to eliminate the parameter a:

$$as_b = \alpha\left(\frac{y-1}{y+\delta}\right) \tag{4.14}$$

$$M_p = \frac{U\epsilon}{p!}\frac{\Gamma(\alpha+p)}{\Gamma(\alpha)}\beta_b^{-p}\left(\frac{1+\delta}{\delta+y}\right)^{-\alpha-p}\left[1+(-1)^p\left(\frac{y}{\delta}\right)^{1-p}\right] \tag{4.15}$$

(We have defined $y \equiv \left(\frac{\delta(1-\epsilon)}{\epsilon}\right)^{1/(\alpha+1)}$ for convenience.) It follows that

$$(ab)^3 = \frac{U\epsilon}{2}\alpha(\alpha+1)\beta_b\frac{(y-1)^3}{y}\frac{(\delta+y)^\alpha}{(\delta+1)^{\alpha+2}} \tag{4.16}$$

and

$$c = U\left[1-\frac{\epsilon}{\delta}\frac{(\delta+y)^{\alpha+1}}{(1+\delta)^\alpha}\right] \tag{4.17}$$

Fixed-point curves in the space of unscaled mean fitness effects can be obtained as follows:

1. For a given N, U, ϵ, δ, and α, c can be computed using (4.17)
2. Using the asymptotic formula $(ab)^3 = 3\log Nc$ or the numerical solution to the generalized infinitesimal limit, ab can be obtained from c. Note that the relationship between ab and c is unchanged by the presence of a distribution of fitness effects, so the same numerical solver can be used as in the case of single effects.

3. From ab, s_b and therefore s_d can be extracted. By varying δ, a fixed-point curve can be drawn.

Asymptotically, the fixed-point curve is given by

$$s_b = \frac{U\epsilon\alpha^2(\alpha+1)}{6y(1+\delta)^2 \log Nc}\left(\frac{\delta+y}{1+\delta}\right)^{\alpha}(y-1)^3 \tag{4.18}$$

$$s_d = \frac{U\epsilon\delta\alpha^2(\alpha+1)}{6y(1+\delta)^2 \log Nc}\left(\frac{\delta+y}{1+\delta}\right)^{\alpha}(y-1)^3 \tag{4.19}$$

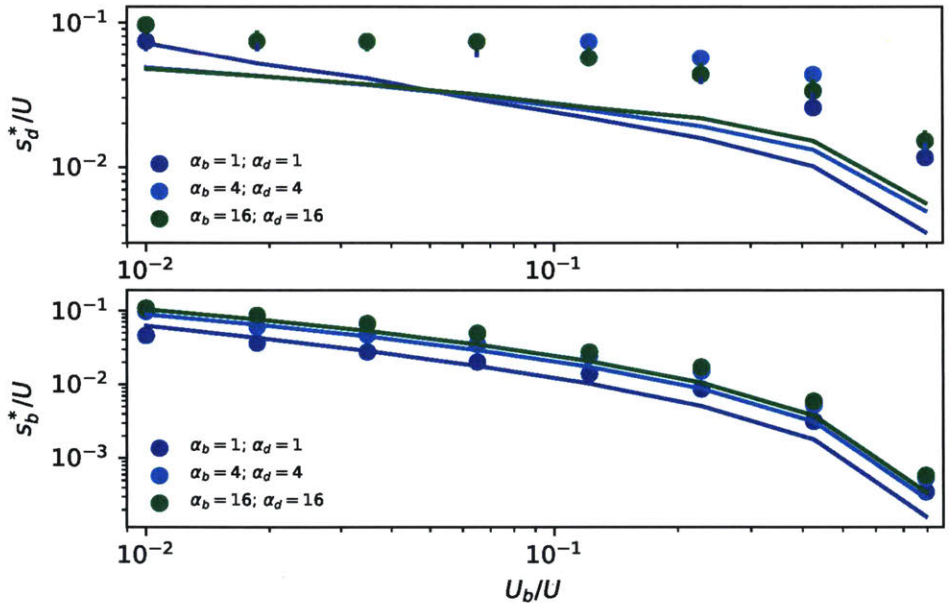

Figure 4.3: Scaling of s_b^* and s_d^* as a function of U_b/U. Both beneficial and deleterious effect sizes were drawn from gamma distributions, with various choices of the shape parameters α_b and α_d. The population-wide mutation rate was specified as $NU = 10^4$. Points are obtained from simulations; solid curves denote theoretical predictions obtained using the numerical solution outlined above.

Chapter 5

Fitness-mediated epistasis

IMPLICATIONS of a given model of fitness-mediated epistasis follow immediately from our knowledge of evolvability phase diagrams. We will first consider models of fitness-mediated epistasis in which γ_b and/or γ_d vary monotonically with fitness. We will also consider fitness-mediated epistatic interactions in which γ_b and γ_d remain fixed, but U_b/U_d varies with fitness, as well as interactions involving changes to the *shape* of the distribution of fitness effects. Of course, since T_2 depends on U_b/U_d, s_b and s_d in general, the assumption that γ_b and/or γ_d remain fixed involves an implicit assumption that s_b and/or s_d vary with fitness. This additional assumption can be removed by considering fitness-mediated epistatic interactions in which *unscaled* fitness effects are held constant or vary monotonically with fitness, instead of scaled fitness effects. While the quantitative results are considerably more complicated, much of the same qualitative behavior is predicted in these analogous scenarios.

As a first example, suppose that scaled fitness effects γ_b and γ_d remain constant, while U_b/U_d decreases with increasing fitness. As previously discussed, a single fixed-point exists, at $U_b/U_d = \frac{\gamma_d e^{-\gamma_d}}{\gamma_b e^{\gamma_b}}$. A larger relative frequency of beneficial to deleterious mutations would correspond to a positive rate of adaptation, while a smaller relative frequency of beneficial to deleterious mutations would correspond to a negative rate of adaptation—therefore, this fixed point is stable. This result is reminiscent of the result in [20]; exactly one fixed point exists, and it is always a stable attractor.

Alternatively, consider diminishing-returns epistasis for beneficial mutations, with fixed U_b/U_d and γ_d, but a variable γ_b that decreases monotonically with increasing fitness. Examining (2.8) we see that for the given parameters, exactly one value of $\gamma_b = W\left(\frac{U_d}{U_b}\gamma_d e^{-\gamma_d}\right)$ lies on the boundary surface. If the population starts out with $\gamma_b > W\left(\frac{U_d}{U_b}\gamma_d e^{-\gamma_d}\right)$, the rate of adaptation will be positive, so that γ_b will decrease over time until $v = 0$ and $\gamma_b = W\left(\frac{U_d}{U_b}\gamma_d e^{-\gamma_d}\right)$. The same argument can be given if the population initially starts out with $\gamma_b < W\left(\frac{U_d}{U_b}\gamma_d e^{-\gamma_d}\right)$ to show that the population is attracted to a *stable* evolutionary fixed point over time. This result is also reminiscent of the result in [20]; at least with respect to the existence and stability of evolutionary fixed-points, interactions involving changes to the beneficial effect size are qualitatively similar to interactions involving changes to the beneficial mutation rate.

An orthogonal model of fitness-mediated epistasis is plausible: suppose that U_b/U_d and γ_b remain constant, but γ_d increases monotonically with increasing fitness. Roughly speaking, as individuals become more fit, a single deleterious mutation might be able to cause more damage. Earlier we saw that if $\gamma_b > W(\frac{U_d}{eU_b})$, then $v > 0$ regardless of the value of γ_d–therefore, if this condition holds, no fixed points exist, and γ_d will increase indefinitely. Alternatively, if $\gamma_b < W(\frac{U_d}{eU_b})$, then *two* fixed points exist, corresponding to the 0 and -1 branches of $-W\left[-\frac{U_b}{U_d}\gamma_b e^{\gamma_b}\right]$, which we'll denote by $\gamma_{d,0}$ and $\gamma_{d,-1}$. It's easy to see that $v < 0$ for $\gamma_{d,0} < \gamma_d < \gamma_{d,-1}$ and $v > 0$ otherwise; thus, if $\gamma_d < \gamma_{d,-1}$ initially, then the population will reach the $\gamma_d = \gamma_{d,0}$ fixed point. If $\gamma_d > \gamma_{d,-1}$ initially, then the population will adapt in fitness, so that γ_d will increase, further pushing the population *away* from the $\gamma_{d,-1}$ fixed point. In conclusion, this model predicts substantially different qualitative behavior, including a bifurcation of the $v = 0$ fixed point into a (lower γ_d) stable fixed point and a (larger γ_d) unstable fixed point.

It remains unclear, however, whether γ_d generally increases monotonically with fitness. One might imagine that γ_d *decreases* with increasing fitness, while U_b/U_d and γ_d remain fixed. Under this assumption, there would still be two fixed points if $\gamma_b < W(\frac{U_d}{eU_b})$ and no fixed points if $\gamma_b > W(\frac{U_d}{eU_b})$. However, the stability of the two fixed points would be interchanged: the $\gamma_{d,0}$ fixed point would be unstable and the $\gamma_{d,-1}$ fixed point would be stable.

We can also posit models of fitness-mediated epistasis in which the *shape* of the DFE varies monotonically with fitness. As in the previous section, we'll limit our analysis to DFEs that consists of a gamma distribution of beneficial effects and another gamma distribution of deleterious effects. For instance, suppose that as fitness increases, the distribution of *beneficial* fitness effects narrows, while its mean remains fixed. Also assume that the distribution of deleterious fitness effects remains unchanged. Under this assumption, a $v = 0$ fixed point exists and is *stable*, since narrowing the distribution of beneficial effects decreases the rate of adaptation. Assuming the opposite relationship between the shape of the beneficial DFE and fitness, the fixed point would be unstable.

Fitness-mediated epistatic interactions involving the shape of the deleterious DFE give rise to some more complicated behavior. If $\gamma_d < 2$, the behavior is relatively simple: broadening the DFE increases the rate of adaptation (a single-effect DFE is optimally deleterious to the rate of adaptation). A stable fixed point would exist if the deleterious DFE narrows with increasing fitness; an unstable fixed point would exist if the deleterious DFE broadens with increasing fitness. However, if $\gamma_d > 2$, then two fixed points would exist: one stable and one unstable. If the deleterious DFE narrows with increasing fitness, then the fixed point corresponding to a broader deleterious DFE will be stable; the other fixed point will be unstable. Additionally, this model gives rise to a bifurcation of the two fixed points, with no fixed point existing if $\alpha_b > \alpha_b^*$, where

$$\left(1 - \frac{\gamma_b}{\alpha_b^*}\right)^{\alpha_b^*+1} = \frac{U_b\gamma_b}{U_d\gamma_d}\left(1 + \frac{\gamma_d}{\alpha_d^*}\right)^{\alpha_d^*+1} \tag{5.1}$$

with α_d^* given by (4.10).

All of the above examples involved epistatic models in which one parameter varies monotonically with fitness and all others remain constant over long evolutionary timescales. This analysis can be straightforwardly generalized to cases in which multiple parameters vary with fitness, if we suppose that the population is constrained to lie on a particular curve in population-genetic parameter space, with its direction of motion at each point of the curve dictated by the corresponding sign of the rate of adaptation at that point. Intersections of the curve with the $v = 0$ surface are fixed-points, whose stability can easily be determined.

To illustrate the approach that can be taken, we will introduce a toy example. First, suppose that U_b/U_d is independent of fitness, and that

$$\gamma_b = cf(X) \tag{5.2}$$

$$\gamma_d = f(X) \tag{5.3}$$

with $f(X)$ a monotonically decreasing function of the absolute fitness X. For concreteness, assume that there exists some absolute fitness X_0 above which both $\gamma_b = 0$ and $\gamma_d = 0$. Existence of a fixed-point depends on the relative magnitude cU_b to U_d: if U_d exceeds cU_b, then a stable fixed-point exists at $\gamma_d = \frac{1}{1+c} \log\left(\frac{U_d}{cU_b}\right)$; otherwise, no fixed-point exists.

))

))

Chapter 6

Discussion

THE evolutionary process does not always result in increasing fitness; in this work we have mapped out an "evolvability phase diagram" that describes which population-genetic parameters give rise to positive adaptation, and which parameters give rise to fitness decline. Our analysis is motivated by the fact that the inverse coalescence timescale sets the fitness effect threshold above which mutational fates are significantly impacted by selection. In particular, we provide evidence from simulations that the simple relation (2.5) predicts the $v = 0$ boundary surface over a broad range of population-genetic parameters, provided that $NU \gg 1$. We then use (2.5) to make an ansatz to the evolutionary equations in order to obtain the scaling of the $v = 0$ boundary surface in the space of unscaled fitness effects.

Especially near the inflection point, prediction of the $v = 0$ boundary surface in the space of scaled fitness effects is found to be more quantitatively accurate than in the space of unscaled fitness effects, though our analysis correctly predicts many qualitative features of the phase diagrams in both the space of scaled fitness effects and the space of unscaled fitness effects. In particular, the following features are observed: for sufficiently weak fitness effects (much smaller than T_2^{-1}), mutations accumulate neutrally, and the boundary surface is approximately planar, determined by equality of $U_b s_b$ and $U_d s_d$. For larger deleterious fitness effects, the boundary surface deviates from this plane toward weaker beneficial effects, and for any fixed N, U_b and U_d, attains a single maximum scaled beneficial effect. This maximum beneficial effect corresponds to a deleterious effect size; for larger deleterious effect sizes, the boundary surface rapidly decays to very weak beneficial effects as Muller's ratchet becomes suppressed. Increasing the relative frequency of beneficial mutations to deleterious mutations largely amounts to a shift in fixed-point curves towards weaker beneficial effects.

Arguably, the distribution of scaled fitness effects is more readily observable than the distribution of unscaled fitness effects. By sequencing samples of two diverged species, and comparing the levels of polymorphism and divergence among synonymous and nonsynonymous mutations, the distribution of scaled selection coefficients can be estimated [24, 42]. In contrast, several efforts have been made to measure distribution of *unscaled* fitness effects using laboratory evolution [12]. The relative frequency of beneficial to deleterious mutations can also be obtained from population-genetic data or experimental evolution; given this information, our results can be used to identify

whether a population is adapting or declining in fitness over time. In particular, our results can shed light on whether a population has approached a $v = 0$ steady state. Convergence to this $v = 0$ steady state would be consistent with modes of fitness-mediated epistasis that give always rise to a stable fixed point, such as running out of mutations and diminishing-returns for beneficial mutations.

The tug-of-war between beneficial and deleterious mutations has been shown to play an important role in cancer progression [32]. In [33], the authors assume a population with a fitness-dependent size is subject to both beneficial (driver) and deleterious mutations occurring at fixed rates and with fixed selection coefficients, and identify a critical population size above which a population will adapt (and below which it will decline in fitness). In a similar vein, a critical mutation rate is found above which population-mean fitness will decline. In this model, positive adaptation of a precancerous lesion entails an increase in size and progression to cancer; remission would presumably occur if a population is below the critical population size or above the critical mutation rate and declines in fitness. In this sense, a population precisely at the critical population size and corresponding critical mutation rate lies at a fixed-point, which is stable by the same arguments given in [20]. Central to this analysis is the assumption that different sites fix independently of one another, with fixation probabilities given by the Kimura formula, with interference among beneficial mutations neglected. Our analysis can be applied to predict the onset of cancer progression in the case that clonal interference and hitchhiking play a more significant role in the dynamics, as has been suggested by a recent empirical study [2] and a theoretical study [31].

In [40], the authors consider the equilibrium distribution of fitness effects that would result purely from the consideration that, after a mutation occurs, a back mutation with an effect size of the same magnitude but opposite sign becomes available. Once the DFE approaches this equilibrium DFE, the population will be in a state of detailed balance, such that, for each effect magnitude, the substitution rate of beneficial mutations exactly matches that of deleterious mutations (thus, the population approaches the $v = 0$ state). Here we focus on an alternative case in which a population does not begin to exhaust its supply of beneficial mutations, so that back mutations are not the primary drivers of changes to the DFE. Instead, changes to the DFE primarily consist of epistatic interactions among different mutations, with these changes bearing a general correlation with fitness. This will perhaps be the dominant form of epistasis at intermediate timescales, with the considerations of [40] becoming increasingly relevant at particularly long evolutionary timescales, when the supply of beneficial mutations begins to be exhausted.

A limitation to our analysis is that we assume that changes to the DFE occur over long evolutionary timescales, so that at any point in time, all individuals in a population share the same DFE, and the population DFE varies negligibly over the course of fixation of a single mutation. This *quasistatic* approximation, also employed in [20], is a large simplification, because it enables the mapping of the problem to one without any epistasis. More realistically, a lineage may be subject to fitness-mediated epistasis

immediately after acquiring a mutation. The immediate action of epistasis may alter the distribution of fixed fitness effects, relative to the distribution of fixed fitness effects that follows from the model we consider. For example, if a lineage acquires a deleterious mutation, it may immediately be subject to a larger beneficial mutation rate and a stronger typical beneficial selective effect; these effects may reduce its extinction probability, relative to the corresponding extinction probability predicted by our model. In general, if the DFE varies sufficiently rapidly with fitness, these effects could become rather pronounced and may alter the $v = 0$ boundary surface. A quantitative understanding of rapid epistasis, in which immediate changes to the DFE might affect fixation probabilities, remains an important topic for future study.

While our analysis accounts for the role of genetic drift in determining whether a mutation will fix or go extinct, we neglect fluctuations in the bulk of the fitness distribution. Analysis of these fluctuations has demonstrated that fluctuations in the number of particularly fit individuals can propagate to fluctuations in the bulk of the fitness distribution, after a delay [37]. The delay between the correlated fluctuations speeds the ratchet; this is one potential explanation for the discrepancy between predicted and observed values of s_d^*, especially if this effect is most pronounced for larger deleterious effects.

We have confined our attention to asexual populations. In recombining populations, the phase diagram should be qualitatively similar, although Muller's ratchet would click much less frequently, so the fixed-point may be well outside the range of realistic population-genetic parameters. Our analysis could be extended to consider recombination using the approach of [17, 38], identifying the self-consistent length of a chromosomal "linkage block" on which the dynamics are essentially asexual.

In conclusion, here we present a computational framework for estimating the coalescence timescale and rate of adaptation over a broad range of population-genetic parameters. We provide arguments that the approximations used are particularly effective in capturing the qualitative features of "evolvability phase diagrams" and generating the implications of any particular model of fitness-mediated epistasis. Simple models of fitness-mediated epistasis, such as diminishing-returns for beneficial mutations and running out of beneficial mutations, give rise to a long-term evolutionary attractor. Other simple interactions, such as diminishing-returns for deleterious mutations, and changes to the shape of the distribution of fitness effects, give rise to more complicated behavior—fixed points may be stable or unstable, and may not exist in certain parameter regimes. Given observations of fitness trajectories or DNA sequence data, these predictions could be used to place constraints on the underlying patterns of fitness-mediated epistasis that are active within a population.

Appendix A

Simulation Methods

We implemented individual-based, discrete-time Wright-Fisher simulations of a population of constant size. In all cases, we considered a population with $N = 10^4$ individuals, and varied the mutation rates U_b and U_d as well as the distribution of fitness effects $\rho(s)$. Each generation, the total number of non-neutral mutations conferred on the population is drawn from a Poisson distribution with mean $N(U_b + U_d)$—these mutations are then randomly assigned to the individuals in the population with fitness effects independently drawn from $\rho(s)$. We do not track the identities of non-neutral mutations possessed by individuals in the population; we only track the absolute fitness X of each individual. The absolute fitness X is defined as the sum of fitness effects of mutations acquired by an individual and its ancestors. Each generation, after mutations are distributed among individuals in a population, we implement a reproductive step. Individuals are sampled with replacement, with probabilities proportional to $e^{X-\bar{X}}$, where $\bar{X}$ is the average absolute fitness of the population.

Populations were also subject to neutral mutations at a rate $NU_n = 500$. These neutral mutations were tracked to obtain the pairwise heterozygosity π of the population, as well as to establish an appropriate sampling interval for the population. More specifically, the population is initialized to be clonal. We measured the number of generations T_{sweep} after initialization until a single neutral mutation was present in all individuals in a population. We then sampled the average absolute fitness $\bar{X}$ of the population as well as the pairwise heterozygosity π every T_{sweep} generations, for a specified number of sampling intervals. For delta-distributed fitness effects, we collected 192 samples for each set of parameters. For gamma-distributed fitness effects with $\alpha = 1$ or with $\alpha = 2$ we collected 48 samples; for gamma-distributed fitness effects with $\alpha = 4$ or with $\alpha = 16$ we collected 192 samples.

Because the population was initialized clonally, transient dynamics at the beginning of our simulation runs may differ from the long-term equilibrium behavior of the mean fitness that we wish to predict. To focus on this equilibrium behavior, in measuring the rate of adaptation, we considered only the change in absolute fitness throughout the second half of collected samples. In measuring the pairwise coalescence timescale T_2 via the pairwise heterozygosity π, we averaged over all collected samples.

For simulations involving mutations of a single beneficial effect size s_b and a single deleterious effect size s_d, measurement of the pairwise coalescence timescale T_2 deter-

mines the corresponding scaled fitness effect sizes. We determined fixed-point curves (Figure 2.2) and the locations of inflection points (Figure 2.3) in the space of scaled fitness effects using the following method. Simulations were performed at $NU = 10^2$, $NU = 10^3$ and $NU = 10^4$, with the values of U_b/U as denoted in Figure 2.3. At each combination of NU and U_b/U, simulations were conducted along the entire grid of unscaled fitness effects displayed in Figure 3.2. At each unscaled deleterious effect (and for each value of U_b/U and NU), linear interpolation between the largest beneficial effect with a negative rate and the smallest beneficial effect with a positive rate determines the corresponding fixed-point unscaled beneficial effect. Linear interpolation of the measured T_2 values at these two points also determines the corresponding fixed point in the space of scaled fitness effects. For each simulated combination of values of U_b/U and NU, the value of γ_d^* is obtained by identifying the largest measured scaled deleterious effect corresponding to a negative measured rate of adaptation, out of all points in the above-mentioned grid of specified unscaled fitness effects. (The corresponding γ_b^* is just the scaled beneficial effect measured at that same point in the grid of unscaled fitness effects.) For simulations involving gamma distributions of fitness effects, a similar approach was taken, but a grid of *mean* fitness effect sizes was specified.

Code is available upon request.

Bibliography

[1] J. Arjan G., M. de Visser, Clifford W. Zeyl, Philip J. Gerrish, Jeffrey L. Blanchard, and Richard E. Lenski. Diminishing returns from mutation supply rate in asexual populations. *Science*, 283(5400):404–406, 1999. ISSN 0036-8075. doi:10.1126/science.283.5400.404. URL http://science.sciencemag.org/content/283/5400/404.

[2] Ann-Marie Baker, Trevor A Graham, and Nicholas A Wright. Pre-tumour clones, periodic selection and clonal interference in the origin and progression of gastrointestinal cancer: potential for biomarker development. *The Journal of Pathology*, 229(4):502–514, 2013. ISSN 1096-9896. doi:10.1002/path.4157. URL http://dx.doi.org/10.1002/path.4157.

[3] Jeffrey E. Barrick, Mark R. Kauth, Christopher C. Strelioff, and Richard E. Lenski. Escherichia coli rpob mutants have increased evolvability in proportion to their fitness defects. *Molecular Biology and Evolution*, 27(6):1338, 2010. doi:10.1093/molbev/msq024. URL +http://dx.doi.org/10.1093/molbev/msq024.

[4] Nick H Barton, Alison M Etheridge, and Amandine Véber. The infinitesimal model. *bioRxiv*, 2016. doi:10.1101/039768. URL http://biorxiv.org/content/early/2016/02/15/039768.

[5] Hsin-Hung Chou, Hsuan-Chao Chiu, Nigel F. Delaney, Daniel Segrè, and Christopher J. Marx. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science*, 332(6034):1190–1192, 2011. ISSN 0036-8075. doi:10.1126/science.1203799. URL http://science.sciencemag.org/content/332/6034/1190.

[6] Elisheva Cohen, David A. Kessler, and Herbert Levine. Front propagation up a reaction rate gradient. *Phys. Rev. E*, 72:066126, Dec 2005. doi:10.1103/PhysRevE.72.066126. URL https://link.aps.org/doi/10.1103/PhysRevE.72.066126.

[7] Alejandro Couce and Olivier A. Tenaillon. The rule of declining adaptability in microbial evolution experiments. *Frontiers in Genetics*, 6:99, 2015. ISSN 1664-8021. doi:10.3389/fgene.2015.00099. URL http://journal.frontiersin.org/article/10.3389/fgene.2015.00099.

[8] Michael M. Desai and Daniel S. Fisher. Beneficial mutation–selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798, 2007. ISSN 0016-6731. doi:10.1534/genetics.106.067678. URL http://www.genetics.org/content/176/3/1759.

[9] A. Etheridge, P. Pfaffelhuber, and A. Wakolbinger. How often does the ratchet click? Facts, heuristics, asymptotics. *ArXiv e-prints*, September 2007.

[10] Daniel S Fisher. Asexual evolution waves: fluctuations and universality. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01011, 2013. URL http://stacks.iop.org/1742-5468/2013/i=01/a=P01011.

[11] Ronald Aylmer Fisher et al. The distribution of gene ratios for rare mutations. In *Proc. R. Soc. Edinb*, volume 50, pages 205–220, 1930.

[12] Evgeni M. Frenkel, Benjamin H. Good, and Michael M. Desai. The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. *Genetics*, 196(4):1217–1226, 2014. ISSN 0016-6731. doi:10.1534/genetics.113.160069. URL http://www.genetics.org/content/196/4/1217.

[13] Benjamin H. Good and Michael M. Desai. Deleterious passengers in adapting populations. *Genetics*, 198(3):1183–1208, 2014. ISSN 0016-6731. doi:10.1534/genetics.114.170233. URL http://www.genetics.org/content/198/3/1183.

[14] Benjamin H. Good and Michael M. Desai. The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics*, 199(1):177–190, 2015. ISSN 0016-6731. doi:10.1534/genetics.114.172460. URL http://www.genetics.org/content/199/1/177.

[15] Benjamin H. Good and Michael M. Desai. Evolution of mutation rates in rapidly adapting asexual populations. *Genetics*, 2016. ISSN 0016-6731. doi:10.1534/genetics.116.193565. URL http://www.genetics.org/content/early/2016/09/19/genetics.116.193565.

[16] Benjamin H. Good, Igor M. Rouzine, Daniel J. Balick, Oskar Hallatschek, and Michael M. Desai. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proceedings of the National Academy of Sciences*, 109(13):4950–4955, 2012. doi:10.1073/pnas.1119910109. URL http://www.pnas.org/content/109/13/4950.abstract.

[17] Benjamin H. Good, Aleksandra M. Walczak, Richard A. Neher, and Michael M. Desai. Genetic diversity in the interference selection limit. *PLOS Genetics*, 10 (3):1–1, 03 2014. doi:10.1371/journal.pgen.1004222. URL https://doi.org/10.1371/journal.pgen.1004222.

[18] Isabel Gordo and Brian Charlesworth. The degeneration of asexual haploid populations and the speed of muller's ratchet. *Genetics*, 154(3):1379–1387, 2000. ISSN 0016-6731. URL http://www.genetics.org/content/154/3/1379.

[19] Roy G Gordon. New method for constructing wavefunctions for bound states and scattering. *The Journal of Chemical Physics*, 51(1):14–25, 1969.

[20] Sidhartha Goyal, Daniel J. Balick, Elizabeth R. Jerison, Richard A. Neher, Boris I. Shraiman, and Michael M. Desai. Dynamic mutation–selection balance as an evolutionary attractor. *Genetics*, 191(4):1309–1319, 2012. ISSN 0016-6731. doi:10.1534/genetics.112.141291. URL http://www.genetics.org/content/191/4/1309.

[21] John Haigh. The accumulation of deleterious genes in a population—muller's ratchet. *Theoretical Population Biology*, 14(2):251 – 267, 1978. ISSN 0040-5809. doi:http://dx.doi.org/10.1016/0040-5809(78)90027-8. URL http://www.sciencedirect.com/science/article/pii/0040580978900278.

[22] Oskar Hallatschek. The noisy edge of traveling waves. *Proceedings of the National Academy of Sciences*, 108(5):1783–1787, 2011. doi:10.1073/pnas.1013529108. URL http://www.pnas.org/content/108/5/1783.abstract.

[23] Katy C Kao and Gavin Sherlock. Molecular characterization of clonal interference during adaptive evolution in asexual populations of saccharomyces cerevisiae. *Nat Genet*, 40(12):1499–1504, 12 2008. URL http://dx.doi.org/10.1038/ng.280.

[24] Peter D. Keightley and Adam Eyre-Walker. What can we learn about the distribution of fitness effects of new mutations from dna sequence data? *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 365 (1544):1187–1193, 2010. ISSN 0962-8436. doi:10.1098/rstb.2009.0266. URL http://rstb.royalsocietypublishing.org/content/365/1544/1187.

[25] MOTOO KIMURA. Evolutionary rate at the molecular level. *Nature*, 217(5129): 624–626, 02 1968. URL http://dx.doi.org/10.1038/217624a0.

[26] J.F.C. Kingman. The coalescent. *Stochastic Processes and their Applications*, 13(3):235 – 248, 1982. ISSN 0304-4149. doi:http://dx.doi.org/10.1016/0304-4149(82)90011-4. URL http://www.sciencedirect.com/science/article/pii/0304414982900114.

[27] Sergey Kryazhimskiy, Daniel P. Rice, Elizabeth R. Jerison, and Michael M. Desai. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344(6191):1519–1522, 2014. ISSN 0036-8075. doi:10.1126/science.1250939. URL http://science.sciencemag.org/content/344/6191/1519.

[28] Gregory I. Lang, Daniel P. Rice, Mark J. Hickman, Erica Sodergren, George M. Weinstock, David Botstein, and Michael M. Desai. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500(7464): 571–574, 08 2013. URL http://dx.doi.org/10.1038/nature12344.

[29] Sasha F. Levy, Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519(7542):181–186, 03 2015. URL http://dx.doi.org/10.1038/nature14279.

[30] R. C. MacLean, G. G. Perron, and A. Gardner. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in pseudomonas aeruginosa. *Genetics*, 186(4):1345–1354, 2010. ISSN 0016-6731. doi:10.1534/genetics.110.123083. URL http://www.genetics.org/content/186/4/1345.

[31] Erik A Martens, Rumen Kostadinov, Carlo C Maley, and Oskar Hallatschek. Spatial structure increases the waiting time for cancer. *New Journal of Physics*, 13(11): 115014, 2011. URL http://stacks.iop.org/1367-2630/13/i=11/a=115014.

[32] Christopher D. McFarland, Kirill S. Korolev, Gregory V. Kryukov, Shamil R. Sunyaev, and Leonid A. Mirny. Impact of deleterious passenger mutations on cancer progression. *Proceedings of the National Academy of Sciences*, 110(8):2910–2915, 2013. doi:10.1073/pnas.1213968110. URL http://www.pnas.org/content/110/8/2910.abstract.

[33] Christopher D. McFarland, Leonid A. Mirny, and Kirill S. Korolev. Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *Proceedings of the National Academy of Sciences*, 111(42):15138–15143, 2014. doi:10.1073/pnas.1404341111. URL http://www.pnas.org/content/111/42/15138.abstract.

[34] Rosario Miralles, Philip J. Gerrish, Andrés Moya, and Santiago F. Elena. Clonal interference and the evolution of rna viruses. *Science*, 285(5434):1745–1747, 1999. ISSN 0036-8075. doi:10.1126/science.285.5434.1745. URL http://science.sciencemag.org/content/285/5434/1745.

[35] R. A. Neher, B. I. Shraiman, and D. S. Fisher. Rate of adaptation in large sexual populations. *Genetics*, 184(2):467–481, 2010. ISSN 0016-6731. doi:10.1534/genetics.109.109009. URL http://www.genetics.org/content/184/2/467.

[36] Richard A. Neher and Oskar Hallatschek. Genealogies of rapidly adapting populations. *Proceedings of the National Academy of Sciences*, 110(2):437–442, 2013. doi:10.1073/pnas.1213113110. URL http://www.pnas.org/content/110/2/437.abstract.

[37] Richard A. Neher and Boris I. Shraiman. Fluctuations of fitness distributions and the rate of muller's ratchet. *Genetics*, 191(4):1283–1293, 2012. ISSN 0016-6731. doi:10.1534/genetics.112.141325. URL http://www.genetics.org/content/191/4/1283.

[38] Richard A. Neher, Taylor A. Kessinger, and Boris I. Shraiman. Coalescence and genetic diversity in sexual populations under selection. *Proceedings of the National Academy of Sciences*, 110(39):15836–15841, 2013. doi:10.1073/pnas.1309697110. URL http://www.pnas.org/content/110/39/15836.abstract.

[39] TOMOKO OHTA. Slightly deleterious mutant substitutions in evolution. *Nature*, 246(5428):96–98, 11 1973. URL http://dx.doi.org/10.1038/246096a0.

[40] Daniel P. Rice, Benjamin H. Good, and Michael M. Desai. The evolutionarily stable distribution of fitness effects. *Genetics*, 200(1):321–329, 2015. ISSN 0016-6731. doi:10.1534/genetics.114.173815. URL http://www.genetics.org/content/200/1/321.

[41] Igor M. Rouzine, John Wakeley, and John M. Coffin. The solitary wave of asexual evolution. *Proceedings of the National Academy of Sciences*, 100(2):587–592, 2003. doi:10.1073/pnas.242719299. URL http://www.pnas.org/content/100/2/587.abstract.

[42] Stanley A. Sawyer, Rob J. Kulathinal, Carlos D. Bustamante, and Daniel L. Hartl. Bayesian analysis suggests that most amino acid replacements in drosophila are driven by positive selection. *Journal of Molecular Evolution*, 57(1):S154–S164, Aug 2003. ISSN 1432-1432. doi:10.1007/s00239-003-0022-3. URL http://dx.doi.org/10.1007/s00239-003-0022-3.

[43] Olin K Silander, Olivier Tenaillon, and Lin Chao. Understanding the evolutionary fate of finite populations: The dynamics of mutational effects. *PLOS Biology*, 5 (4):1–10, 04 2007. doi:10.1371/journal.pbio.0050094. URL https://doi.org/10.1371/journal.pbio.0050094.

[44] John Maynard Smith and John Haigh. The hitch-hiking effect of a favourable gene. *Genetical Research*, 23(1):23?35, 1974. doi:10.1017/S0016672300014634.

[45] Olivier Tenaillon, Alejandra Rodríguez-Verdugo, Rebecca L. Gaut, Pamela McDonald, Albert F. Bennett, Anthony D. Long, and Brandon S. Gaut. The molecular diversity of adaptive convergence. *Science*, 335(6067):457–461, 2012. ISSN

0036-8075. doi:10.1126/science.1212986. URL http://science.sciencemag.org/content/335/6067/457.

[46] Nobuhiko Tokuriki, Colin J. Jackson, Livnat Afriat-Jurnou, Kirsten T. Wyganowski, Renmei Tang, and Dan S. Tawfik. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nature Communications*, 3: 1257 EP –, 12 2012. URL http://dx.doi.org/10.1038/ncomms2246.

[47] Lev S. Tsimring, Herbert Levine, and David A. Kessler. Rna virus evolution via a fitness-space model. *Phys. Rev. Lett.*, 76:4440–4443, Jun 1996. doi:10.1103/PhysRevLett.76.4440. URL https://link.aps.org/doi/10.1103/PhysRevLett.76.4440.

[48] John Wakeley. *Coalescent Theory, an Introduction*. Roberts and Company, Greenwood Village, CO, 2005. URL http://www.coalescentheory.com/.

[49] Michael J. Wiser, Noah Ribeck, and Richard E. Lenski. Long-term dynamics of adaptation in asexual populations. *Science*, 342(6164):1364–1367, 2013. ISSN 0036-8075. doi:10.1126/science.1243357. URL http://science.sciencemag.org/content/342/6164/1364.

[50] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931. ISSN 0016-6731. URL http://www.genetics.org/content/16/2/97.

[51] Andrea Wünsche, Duy M. Dinh, Rebecca S. Satterwhite, Carolina Diaz Arenas, Daniel M. Stoebel, and Tim F. Cooper. Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. *Nature Ecology & Evolution*, 1:0061 EP –, 03 2017. URL http://dx.doi.org/10.1038/s41559-016-0061.