

# Evolution of mutation rates in rapidly adapting asexual populations

Benjamin H. Good and Michael M. Desai

*Department of Organismic and Evolutionary Biology, Department of Physics,  
and FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138*

Mutator and antimutator alleles often arise and spread in both natural microbial populations and laboratory evolution experiments. The evolutionary dynamics of these mutation rate modifiers are determined by indirect selection on linked beneficial and deleterious mutations. These indirect selection pressures have been the focus of much earlier theoretical and empirical work, but we still have a limited analytical understanding of how the interplay between hitchhiking and deleterious load influences the fates of modifier alleles. Our understanding is particularly limited when clonal interference is common, which is the regime of primary interest in laboratory microbial evolution experiments. Here, we calculate the fixation probability of a mutator or antimutator allele in a rapidly adapting asexual population, and we show how this quantity depends on the population size, the beneficial and deleterious mutation rates, and the strength of a typical driver mutation. In the absence of deleterious mutations, we find that clonal interference enhances the fixation probability of mutators, even as they provide a diminishing benefit to the overall rate of adaptation. When deleterious mutations are included, natural selection pushes the population towards a stable mutation rate that can be suboptimal for the adaptation of the population as a whole. The approach to this stable mutation rate is not necessarily monotonic, and selection can favor mutator and antimutator alleles that “overshoot” the stable mutation rate by substantial amounts.

## INTRODUCTION

DNA replication occurs with extremely high fidelity, despite taking place in the noisy environment of the cell. For example, laboratory strains of *E. coli* produce a point mutation at a rate of roughly one nucleotide per ten billion copied (Lee et al., 2012; Wielgoss et al., 2011), which implies that hundreds of generations can elapse before a single mutation is introduced into the genome. To achieve such low error rates, bacteria and eukaryotes employ a complex array of cellular machinery, which must be maintained by natural selection.

One explanation for the low observed error rates is that the genome encodes a large number of functions, all of which are essential for survival. Low mutation rates could then emerge from hard selection against these lethal errors. But in practice, observed mutation rates lie far below the levels that would quickly result in extinction. This is evident from the fact that mutator strains, whose mutation rates are 10- to 1000-fold higher than the wild-type, can be propagated for thousands of generations in the laboratory without significant loss of viability (McDonald et al., 2012; Wiser et al., 2013). This suggests that mutation rates are not solely maintained by hard selection, but also by more direct evolutionary competition between strains with different mutation rates. These variants feel the effects of natural selection indirectly, by being linked to other mutations that directly influence fitness.

Strains with higher mutation rates are more likely to be linked to deleterious mutations, and will therefore

experience an effective fitness cost. This cost can be measured in head-to-head competitions between mutator and wildtype strains (Chao and Cox, 1983; Chao et al., 1983; Gentile et al., 2011; Giraud et al., 2001; Thompson et al., 2006; Tröbner and Piechocki, 1981). In the absence of beneficial mutations, natural selection will therefore act to decrease the mutation rate, until it is eventually balanced by genetic drift, mutation, or other physiological costs. A large body of previous theoretical work has explored these dynamics (Dawson, 1998, 1999; Desai and Fisher, 2011; James and Jain, 2015; Johnson, 1999a; Kimura, 1967; Liberman and Feldman, 1986; Lynch, 2008, 2011; Soderberg and Berg, 2011).

When beneficial mutations are available, lineages with higher mutation rates are also more likely to produce and hitchhike with a successful beneficial variant. In the absence of deleterious mutations, natural selection will therefore act to increase the mutation rate, until the supply of beneficial mutations is eventually exhausted. In accordance with these expectations, mutator alleles are often found to spontaneously arise and spread in rapidly adapting microbial populations in the laboratory (Notley-McRobb et al., 2002; Pal et al., 2007; Shaver et al., 2002; Sniegowski et al., 1997; Voordeckers et al., 2015), and are often correlated with pathogenic lifestyles in the wild (Bjorkholm et al., 2001; del Campo et al., 2005; Denamur et al., 2002; Giraud et al., 2002; Labat et al., 2005; LeClerc et al., 1996; Matic et al., 1997; Oliver et al., 2000; Prunier et al., 2003; Richardson et al., 2002; Watson et al., 2004).

In a few cases, laboratory populations that have previ-

ously fixed a mutator allele have been shown to evolve a lower mutation rate later in the experiment, suggesting that mutation rate evolution is an ongoing process (McDonald et al., 2012; Notley-McRobb et al., 2002; Tröbner and Piechocki, 1984; Turrientes et al., 2013; Wielgoss et al., 2013). Because these populations often continue to increase in fitness during this process (Wielgoss et al., 2013), the success of mutator and antimutator alleles must depend on the interplay between beneficial and deleterious mutations, rather than the complete cessation of adaptation. However, our understanding of this interplay remains incomplete. A few simulation studies have shown how these effects can in principle favor either increases or decreases in mutation rates, depending on the specific population parameters (Taddei et al., 1997; Tenaillon et al., 2000, 1999; Travis and Travis, 2002). Some analytical progress has also been made in the case where beneficial mutations are rare, and occur one-by-one in discrete selective sweeps (Andre and Godelle, 2006; Desai and Fisher, 2011; Gillespie, 1981; Ishii et al., 1989; Johnson, 1999b; Leigh, 1970, 1973; Painter, 1975; Tanaka et al., 2003; Wylie et al., 2009) However, our understanding is much more limited in larger populations where beneficial mutations are more common, and multiple adaptive lineages must compete with each other for fixation. This is the regime where the indirect effects of linked selection are likely to be maximized, and it is also the most relevant for understanding the microbial evolution experiments where mutation rate modifiers have been observed to spread. In this clonal interference regime, even the most basic questions about this process remain unanswered: exactly how large can the deleterious load be before it effectively selects against a mutator allele? And how does this depend on the population size, the supply of beneficial mutations, and the magnitude of the mutator phenotype?

Here, we address these questions in the context of a simple model of adaptation, which is motivated by recent microbial evolution experiments. In particular, we focus on a regime in which deleterious mutations have a negligible impact on the rate of adaptation, even though they may have a large effect on the fates of mutation rate modifiers. Within this model, we calculate the fixation probability of an allele that changes the mutation rate by a factor  $r$  (with  $r > 1$  corresponding to mutator alleles and  $r < 1$  corresponding to antimutators). We consider both small changes in mutation rate ( $r \sim 1$ ) as well as substantial changes (where  $|\log r| \gg 1$ ).

We find that clonal interference amplifies the beneficial advantage of mutator alleles, even as it constrains the overall rate of adaptation. For large  $r$  this can be a substantial effect, resulting in greater than 100-fold increases in the probability of fixation. However, clonal interference also amplifies the effects of the deleterious load, re-

sulting in a much narrower window of mutator favorability. We show that this window depends most strongly on the strengths of typical driver mutations, and relatively weakly on the rate at which beneficial mutations arise. With the steady production of such modifier alleles, the mutation rate will evolve towards a stable point at which neither mutator or antimutator alleles are favored. Surprisingly, we find that the approach to this stable point is not necessarily monotonic, even in the absence of epistasis. Instead, alleles that overshoot the stable point can be positively selected, and selection must then act to change the mutation rates in the opposite direction.

## MODEL

Our goal is to understand how beneficial and deleterious mutations influence the fates of mutation rate modifiers in rapidly adapting populations. We focus on the simplest possible model in which we can address this question. Specifically, we consider an asexual haploid population of constant size  $N$  (our analysis also applies to diploids when all mutations have intermediate dominance effects). We assume beneficial and deleterious mutations occur at genome-wide rates  $U_b$  and  $U_d$  respectively. We define  $U = U_b + U_d$  as the total mutation rate and  $\epsilon \equiv \frac{U_b}{U_d}$  as the ratio of beneficial to deleterious mutation rates. We focus on adapting populations, and consider both the successional-mutation regime (where beneficial mutations are rare) and the clonal interference regime.

For most of our analysis we assume that beneficial mutations all confer the same fitness advantage  $s_b$ , though we also comment on how our results can be extended to the case where beneficial mutations have a distribution of fitness effects. We assume throughout that there is no macroscopic epistasis among beneficial mutations, so that  $U_b$  and  $s_b$  remain fixed as the population adapts. We make no specific assumptions about the fitness effects of deleterious mutations, other than that they are *purgeable* — that is, deleterious mutations are sufficiently costly that they are unlikely to fix and do not significantly reduce the overall rate of adaptation of the population. We describe the technical conditions required for this to be true and the qualitative implications of non-purgeable deleterious mutations in more detail in the Discussion.

Our goal is to analyze the fate of a modifier allele that changes the overall mutation rate  $U$  by a factor  $r$  (while leaving  $\epsilon$  and  $s_b$  unchanged). For simplicity, we assume that this allele has no direct influence on fitness, although in the Discussion we show how this assumption can be relaxed. These modifier alleles are produced at rate  $\mu$  from the wildtype background, and we assume that  $\mu$  is sufficiently small that the modifiers can be treated independently. In this case, the substitution rate of the

modifier is  $N\mu p_{\text{fix}}(r)$ , where  $p_{\text{fix}}(r)$  is the fixation probability of a single modifier individual. A natural measure of how selection “favors” or “disfavors” the modifier can be obtained from the scaled fixation probability,  $Np_{\text{fix}}(r)$ . When  $Np_{\text{fix}} > 1$ , the allele is favored by selection, since it substitutes more rapidly than a neutral allele; conversely, the allele is disfavored when  $Np_{\text{fix}} < 1$ . This scaled fixation probability will be our primary focus in the analysis below. In particular, we are interested in determining the parameters where  $Np_{\text{fix}}$  transitions between favored ( $Np_{\text{fix}} > 1$ ) and unfavored ( $Np_{\text{fix}} < 1$ ).

We are primarily interested in situations applicable to microbial populations, and particularly to laboratory evolution experiments. This motivates certain technical assumptions we describe in the Analysis below. It also motivates our focus on asexual populations where recombination can be neglected. This asexual case is particularly relevant because when recombination is absent, modifiers of mutation rate remain perfectly linked to beneficial or deleterious mutations they generate, and hence experience the strongest indirect selection pressures. In principle, we could also apply our analysis to physically linked regions within sexual genomes (“linkage blocks”) that are unlikely to recombine on the appropriate timescales (Good et al., 2014; Neher et al., 2013; Weissman and Hallatschek, 2014). However, the scale of these linkage blocks is a complex problem, and understanding these effects is beyond the scope of this study.

In the remaining sections, we calculate  $Np_{\text{fix}}(r)$  as a function of the underlying model parameters, and then we use these predictions to analyze the dynamics of mutation rate evolution.

## THE SUCCESSIONAL MUTATIONS REGIME

To gain intuition into the forces that determine  $Np_{\text{fix}}(r)$ , it is useful to start by considering the simplest case, where deleterious mutations can be neglected and adaptation proceeds by a sequence of discrete selective sweeps. This “successional mutations” regime applies whenever beneficial mutations are sufficiently rare that they seldom segregate within the population at the same time. This will be true provided that  $NU_b \log(Ns_b) \ll 1$  (Desai and Fisher, 2007). In this regime, beneficial mutations are produced at rate  $NU_b$ , and establish with probability  $p_{\text{est}} = 2s_b$ . Selective sweeps then occur as a Poisson process with rate

$$\frac{1}{\tau_{\text{est}}} = 2NU_b s_b. \quad (1)$$

When a successful beneficial mutation establishes, it starts to grow deterministically and requires roughly  $T_{\text{fix}} = \frac{2}{s_b} \log(Ns_b)$  generations to complete its sweep. By

definition, this time is small compared to the time between sweeps, so the rate of adaptation is just  $v = s_b/\tau_{\text{est}}$ . A modifier allele would increase or decrease this rate by factor of  $r$ , if it was able to fix. We now analyze the dynamics of this process.

When a modifier allele first arises, it starts at an initial frequency  $f_m(0) = 1/N$ , and it will then drift neutrally until the next selective sweep occurs. Most of the time, the lineage will fluctuate to extinction within the first few generations. However, with probability  $\sim 1/t$ , it will survive for  $\sim t$  generations and reach size  $f_m(t) \sim t/N$  (Fisher, 2007). In the absence of selective sweeps, the mutation modifier lineage could only fix by drifting to fixation; this happens with probability  $\sim 1/N$  and occurs over a timescale of order  $\sim N$  generations. Thus, depending on how this drift timescale  $N$  compares with the sweep timescales  $\tau_{\text{est}}$  and  $\tau_{\text{est}}/r$ , two distinct types of fixation dynamics can emerge.

When  $\tau_{\text{est}}$  and  $\tau_{\text{est}}/r$  are both much much larger than  $N$ , then the fate of the modifier lineage will be determined long before the next selective sweep occurs. Drift is the dominant evolutionary force, and  $p_{\text{fix}}(r) = 1/N$ .

In contrast, if either  $\tau_{\text{est}}$  and  $\tau_{\text{est}}/r$  are small compared to  $N$ , the fixation process is instead controlled by genetic draft. In this case, the modifier lineage will only fix if it is lucky enough to produce and hitchhike with the next selective sweep. Since we are primarily interested in understanding these effects, we will focus on this regime for the remainder of our analysis.

The next selective sweep is produced by two competing Poisson processes, corresponding to the beneficial mutants from the wildtype and modifier lineages. These respectively produce sweeps at rates

$$\lambda_0(t) = 2NU_b s_b [1 - f_m(t)], \quad (2a)$$

$$\lambda_m(t) = 2NU_b s_b r f_m(t). \quad (2b)$$

The probability that the next sweep is produced by the modifier lineage is then simply

$$p_{\text{fix}} = \left\langle \int_0^\infty \lambda_m(t) e^{-\int_0^t [\lambda_0(t') + \lambda_m(t')] dt'} dt \right\rangle, \quad (3)$$

where the angled brackets denote an average is over the random lineage trajectory  $f_m(t)$ . In the regime we are considering, the modifier lineage will remain rare until the next selective sweep occurs. The trajectory  $f_m(t)$  is therefore described by the Langevin dynamics,

$$\frac{\partial f_m(t)}{\partial t} = \sqrt{\frac{f_m}{N}} \eta(t), \quad (4)$$

where  $\eta(t)$  is a Brownian noise term (Gardiner, 1985). The solution of Eq. (3) can be complicated, since both the overall rate of sweeps and their probability of arising

in the modifier lineage depend on the random trajectory  $f_m(t)$ . We present an exact solution of these equations in Appendix A, using techniques developed by Weissman et al. (2009).

For the present purposes, however, it will be more useful to focus on an approximation. If the modifier lineage stays sufficiently rare that it does not influence the overall rate of sweeps, then we can approximate  $\lambda_0(t) + \lambda_m(t) \approx 1/\tau_{\text{est}}$ . Equation (3) then reduces to

$$p_{\text{fix}} = \int_0^\infty r \langle f_m(t) \rangle \frac{e^{-t/\tau_{\text{est}}}}{\tau_{\text{est}}} dt. \quad (5)$$

In other words, the fixation probability is given by the average size of the lineage at the time of the next sweep, and this time is unaffected by the presence of the modifier. In this case, the modifier lineage is neutral and  $\langle f_m(t) \rangle = 1/N$ , so that

$$N p_{\text{fix}}(r) = r. \quad (6)$$

Thus, we see that the modifier fixation probability (like the proportional change in the rate of adaptation) is independent of the population size, the mutation rate, and the fitness benefits of the driver mutations.

To derive Eq. (6), we made the approximation that  $\lambda_0(t) + \lambda_m(t) \approx 1/\tau_{\text{est}}$ . Substituting our expressions for  $\lambda_0(t)$  and  $\lambda_m(t)$  in Eq. (2), we see that this is equivalent to the assumption that  $(r-1)f_m(t) \ll 1$ . However, since  $f_m$  and  $t$  are both random variables, we cannot determine the validity of this assumption based on the average values alone. We must instead consider the *typical* dynamics that contribute to the average in Eq. (5): with probability  $\sim 1/\tau_{\text{est}}$ , the modifier lineage survives for  $\tau_{\text{est}}$  generations and reaches size  $f_m(\tau_{\text{est}}) \sim \tau_{\text{est}}/N$ . This ensures that the average size of the lineage is just  $\frac{\tau_{\text{est}}}{N} \cdot \frac{1}{\tau_{\text{est}}} \sim \frac{1}{N}$ , as expected. It also shows that the typical values of  $(r-1)f_m(t)$  will remain small compared to one provided that  $r\tau_{\text{est}} \ll N$ . For simplicity, we will assume that this condition holds for the remainder of our analysis. For moderate  $r$ , this is actually the same assumption we already made in focusing on the genetic-draft regime. For large  $r$  it imposes an upper limit  $r \sim N/\tau_{\text{est}} \gg 1$  on the range of modifier effects that we consider. However, this is not a fundamental problem for the analysis; we consider such large-effect modifiers in Appendix A.

### Incorporating purgeable deleterious mutations

A similar picture applies in the presence of strongly deleterious mutations. In particular, we assume that the typical fitness costs are larger than  $s_b$ , so that a driver mutation can only fix in a background that is free of deleterious mutations. If the costs are also much larger than  $U_d$ ,

then the vast majority of the population will be of this mutation-free type. We refer to these as *purgeable* mutations, and we will assume that all deleterious mutations are purgeable in the analysis that follows.

When a new beneficial mutation occurs, the wildtype will be in mutation-selection balance with respect to its deleterious mutations, so the mean fitness of the population is  $-U_d$ . Even if it arises on a mutation-free background, the new beneficial lineage will continue to produce loaded individuals at rate  $U_d$ . This exactly cancels the growth advantage of the mean fitness of the population, so that the establishment probability is still just  $p_{\text{est}} = 2(s_b + U_d - U_d) = 2s_b$ . Since most genetic backgrounds are in this unloaded state, the overall rate of sweeps is unchanged from before. This means that if a modifier allele were to fix, it would still change the rate of adaptation by a simple factor of  $r$ .

However, while the deleterious mutations do not affect the rate of adaptation, they can still have a large effect on the modifier lineage while it is rare. The modifier produces doomed lineages at rate  $rU_d$ , which is not completely cancelled by the mean fitness of the wildtype. Instead, mutator alleles will feel an effective cost of magnitude  $U_d(r-1)$ , while antimutators will experience an effective fitness advantage. Generalizing Eq. (4), the mutation-free portion of the modifier lineage will then evolve as

$$\frac{\partial f_m(t)}{\partial t} = -U_d(r-1)f_m(t) + \sqrt{\frac{f_m(t)}{N}}\eta(t). \quad (7)$$

Beneficial mutations produced by this lineage will likewise have an effective fitness  $s_b - U_d(r-1)$ , and will establish with probability

$$p_{\text{est}}^m = 2(s_b - U_d(r-1)). \quad (8)$$

In exactly the same manner as before, the next selective sweep is produced by two competing Poisson processes with rates

$$\lambda_0(t) = \frac{1}{\tau_{\text{est}}}[1 - f_m(t)], \quad (9a)$$

$$\lambda_m(t) = \frac{r}{\tau_{\text{est}}} \left[ 1 - \frac{U_d(r-1)}{s_b} \right] f_m(t), \quad (9b)$$

and the fixation probability is given by the average in Eq. (3). If we again make the assumption that the modifier lineage has a negligible influence on the overall sweep rate ( $\lambda_m + \lambda_0 \approx 1/\tau_{\text{est}}$ ), this reduces to

$$p_{\text{fix}} = r \left[ 1 - \frac{U_d(r-1)}{s_b} \right] \int_0^\infty \langle f_m(t) \rangle \cdot \frac{e^{-t/\tau_{\text{est}}}}{\tau_{\text{est}}} dt. \quad (10)$$

In this case, the average lineage size is now  $\langle f_m(t) \rangle =$



$e^{-U_d(r-1)t}/N$ , and we obtain

$$Np_{\text{fix}}(r) = \frac{r \left[ 1 - \frac{U_d(r-1)}{s_b} \right]}{1 + U_d\tau_{\text{est}}(r-1)}. \quad (11)$$

The validity of this expression depends on our assumption that  $\lambda_0(t) + \lambda_m(t) \approx 1/\tau_{\text{est}}$ . In the same manner as above, we can establish the region of validity of this approximation by considering the typical dynamics behind the averages in Eq. (10). These will depend on the sign of  $U_d(r-1)$ .

For a mutator allele ( $r > 1$ ), the modifier lineage will feel an effective fitness cost  $U_d(r-1)$ . If this cost is small compared to  $1/\tau_{\text{est}}$ , then selection will barely have time to influence  $f_m(t)$  before the next sweep occurs, and mutators will continue to fix with probability  $Np_{\text{fix}} \sim r$ . On the other hand, if  $U_d(r-1) \gg \tau_{\text{est}}$ , then this fitness cost exponentially suppresses the probability that the mutator survives until the next sweep. In particular, the mutator will survive for  $t$  generations with probability  $\sim U_d(r-1)e^{-U_d(r-1)t}$ , and provided that it does so, it will no longer grow indefinitely, but will instead reach a maximum size of order  $1/NU_d(r-1)$ . If the time between sweeps was deterministic, this would simply suppress the fixation probability by a factor  $e^{-U_d(r-1)\tau_{\text{est}}}$ . However, this is not actually the case, since the next sweep is generated by the random Poisson process above. Thus, the intervals between sweeps are not only random, but they are drawn from an exponential distribution, which has important consequences for the dynamics. The broad distribution of  $t$  does not lead to a simple exponential suppression of  $Np_{\text{fix}}(r)$ , but rather a power law decay  $\propto 1/U_d\tau_{\text{est}}(r-1)$ , which is dominated by anomalously early sweeps for which  $t \sim 1/U_d(r-1) \ll \tau_{\text{est}}$ . For a strong-effect mutator ( $r \gg 1$ ), this power-law decay exactly balances the lineage's increased beneficial mutation rate, so that the  $r$ -dependence of  $p_{\text{fix}}$  is primarily mediated through the reduced establishment probability,  $p_{\text{est}}^m$ .

In contrast, antimutators will have an effective fitness advantage relative to the wildtype, of magnitude  $U_d(1-r)$ . If  $U_d(1-r) \ll 1/\tau_{\text{est}}$ , this fitness advantage will still have a negligible influence on  $Np_{\text{fix}}(r)$ . However, when  $U_d(1-r)\tau_{\text{est}} \gg 1$ , the antimutator will establish with probability  $U_d(1-r)$  and will start to grow deterministically as  $f_m(t) \sim e^{U_d(r-1)t}/NU_d(r-1)$ . Again, if the time to the next sweep was deterministic, this would result in a simple exponential enhancement of  $Np_{\text{fix}}$ . But because  $t$  is exponentially distributed, it interacts with the exponential growth of the lineage in a strong way. In particular, as  $U_d(1-r)\tau_{\text{est}}$  approaches 1, the average in Eq. (11) will be dominated by anomalously late sweep times for which  $f_m(t)$  is no longer rare, and our approximation that  $\lambda_0 + \lambda_m \approx 1/\tau_{\text{est}}$  breaks down. Thus, Eq. (11) will only be valid provided that

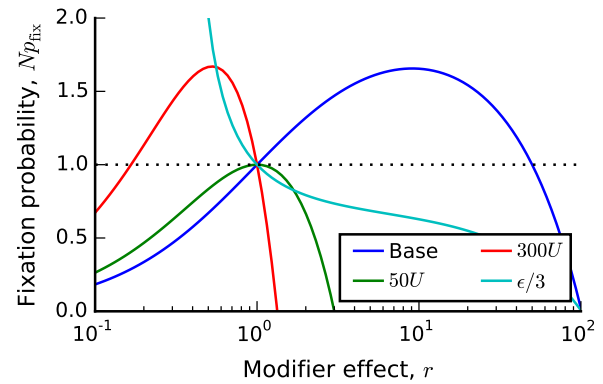


FIG. 1 The fixation probability of a mutation rate modifier in the successive mutations regime. Solid lines depict Eq. (11) for four sets of parameters, which illustrate the four characteristic shapes of  $Np_{\text{fix}}(r)$ . In all four cases, the base parameters are  $N = 10^7$ ,  $s_b = 10^{-2}$ ,  $U = 10^{-4}$  and  $\epsilon = 10^{-5}$ , with modifications listed in the legend.

$1 - U_d(1-r)\tau_{\text{est}} \gg 1/\log(N/\tau_{\text{est}})$ . Outside this regime, the antimutator lineage will appear to sweep to fixation on its own, without the help of an additional driver mutation.

The fixation probability in Eq. (11) is illustrated in Fig. 1 for several choices of parameters. Its overall shape is determined by the key parameter

$$U_d\tau_{\text{est}} = \frac{1}{2\epsilon N s_b}, \quad (12)$$

which depends only on  $N$ ,  $s_b$ , and  $\epsilon \equiv U_b/U_d$ , and is independent of the overall mutation rate  $U$ . Depending on the value of  $U_d\tau_{\text{est}}$ , Eq. (11) takes on one of two characteristic shapes. If  $U_d\tau_{\text{est}} > 1$ , the fixation probability is a monotonically decreasing function of  $r$  (see Fig. 1), and mutators will never be favored to invade. Instead, antimutators will be positively selected. Note that this happens even though the antimutator actually causes a *decrease* in the overall rate of adaptation.

On the other hand, for  $U_d\tau_{\text{est}} < 1$ , the fixation probability takes on a paraboloid shape (Fig. 1). It crosses  $Np_{\text{fix}}(r) = 1$  at exactly two points: once at  $r = 1$  and once at

$$r^* = \frac{s_b}{U_d} (1 - U_d\tau_{\text{est}}). \quad (13)$$

When  $r^* > 1$ , modifiers will be favored in the range  $1 < r < r^*$ , and disfavored elsewhere (i.e., some mutators will be favored). Conversely, when  $r^* < 1$ , modifiers will be favored in the range  $r^* < r < 1$  (i.e., some antimutators will be favored). It is also useful to rewrite  $r^*$  in terms of the mutation rate  $U_m = Ur$  of the modifier lineage:

$$U_m^* \equiv Ur^* = s_b(1 + \epsilon) \left( 1 - \frac{1}{2\epsilon N s_b} \right). \quad (14)$$

Note that this is always positive (since  $1/2Ns_b\epsilon < 1$ ), and it is independent of  $U$ .

### Dynamics of mutation rate evolution

We are now in a position to analyze how mutation rates evolve over time. Imagine that we start with some particular values of  $N$ ,  $U$ ,  $\epsilon$ , and  $s_b$ . This will correspond to a particular value of  $\tau_{\text{est}}$  and  $U_m^*$ . If  $U_d\tau_{\text{est}} > 1$ , mutators are disfavored and antimutator alleles will be positively selected instead. After an antimutator allele fixes, both  $U_b$  and  $U_d$  are reduced, but  $U_d\tau_{\text{est}}$  is unchanged, so natural selection will continue to select for lower mutation rates until this force is balanced by drift, mutational pressure, or other physiological costs.

On the other hand, if  $U_d\tau_{\text{est}} < 1$  and  $U < U_m^*$ , mutators will be favored provided that their resulting mutation rate is less than  $U_m^*$ . If a mutator allele fixes,  $U_b$  and  $U_d$  will be increased, but  $U_d\tau_{\text{est}}$  and  $U_m^*$  will remain constant. Thus, natural selection will continue to favor increased mutation rates until  $U$  reaches a special value,

$$\hat{U} = U_m^* = s_b(1 + \epsilon) \left( 1 - \frac{1}{2\epsilon N s_b} \right), \quad (15)$$

which is stable against further changes in the mutation rate. If instead the initial mutation rate  $U > \hat{U}$ , antimutators will be favored provided that their resulting mutation rate is greater than  $\hat{U}$ , and natural selection will continue to favor increased mutation rates until  $U$  reaches  $\hat{U}$ . We describe these dynamics in more detail in the Discussion.

Of course, this analysis crucially depends on the assumption that any mutation rate increases will still lie within the successive mutations regime. This will be true provided that  $N\hat{U}_b \sim \epsilon N s_b - 1 \ll 1/\log(N s_b)$ . But in order for a nonzero stable mutation rate to exist in the first place, we previously required that  $\epsilon N s_b \sim 1/U_d\tau_{\text{est}} > 1$ . These two conditions can only be satisfied if  $\epsilon = 1/2N s_b + \delta$ , where

$$0 < \delta \ll \frac{1}{\log(N s_b)}. \quad (16)$$

This is an extremely small region of parameter space, and it grows increasingly narrow as  $N s_b$  increases. As a result, the stable mutation rate in Eq. (15) is actually a rapidly varying function of  $N$ ,  $s_b$ , and  $\epsilon$ . For larger values of  $\epsilon$  that violate the stringent constraints in Eq. (16), mutator alleles will generically drive mutation rates into regimes where selective sweeps begin to interfere with each other. We now turn to an analysis of this case.

### THE CLONAL INTERFERENCE REGIME

When multiple beneficial mutations segregate at the same time, many potential drivers are lost due to competition with other, fitter genetic backgrounds. This reduces the rate of *successful* drivers in a way that depends on the relative values of  $N s_b$  and  $N U_b$ . In the regime most relevant for our current study, Desai and Fisher (2007) have shown that  $\tau_{\text{est}}$  is reduced to

$$\frac{1}{\tau_{\text{est}}} \sim \frac{2s_b \log(N s_b)}{\log^2(s_b/U_b)}, \quad (17)$$

which is valid provided that  $1 \ll \log(s_b/U_b) \ll 2\log(N s_b)$  and  $2\log(N s_b) \ll \log^2(s_b/U_b)$ . Similar expressions can be obtained for other parameter regimes, all of which share the weak dependence on  $N U_b$  (Fisher, 2013). Since adaptation is no longer mutation-limited, one might guess that mutators will be less strongly favored in this regime. However, previous simulation studies (Tenaillon et al., 1999) and heuristic reasoning (Desai and Fisher, 2011) suggest that the opposite can actually be true: clonal interference enhances the fixation of mutator alleles, even as they provide a diminishing overall benefit for the rate of adaptation.

In the following subsection, we introduce a traveling-wave formalism for calculating  $N p_{\text{fix}}(r)$  in the presence of clonal interference. Before doing so, however, it will be useful to consider this process from a heuristic perspective. This will allow us to identify the key forces and parameters involved, and will provide intuition for the more rigorous analysis that follows.

#### Heuristic analysis

In the absence of deleterious mutations, clonal interference alters the dynamics of fixation in two main ways. First, successful mutations can only occur in the most highly-fit genetic backgrounds in the population. These individuals lie in the extreme right tail or “nose” of the population fitness distribution, which steadily increases in fitness as the population adapts. In the regime described by Eq. (17), the relative fitness of the nose is given by

$$x_c \sim \frac{1}{\tau_{\text{est}}} \log \left( \frac{s_b}{U_b} \right), \quad (18)$$

which is much larger than the size of a single driver mutation. These individuals have already acquired  $q = x_c/s_b \gg 1$  more adaptive mutations than the average individual, but they are not yet destined to fix. The reason is that there are still enough individuals in the nose that they will collectively produce *multiple* additional driver

mutations in the next  $\tau_{\text{est}}$  generations, and these will occur in a relatively narrow time window  $\Delta t \ll \tau_{\text{est}}$  (Desai and Fisher, 2007). By definition, all but one of these mutations must eventually be outcompeted. But this means that the process of fixation within the nose takes place over *multiple* establishment intervals, each of length  $\tau_{\text{est}}$ .

Since  $x_c \tau_{\text{est}} \gg 1$ , genetic drift is not directly relevant for most of this fixation process. Random fluctuations in the lineage sizes are still important, but these are now driven by *genetic draft*, which arises from slight differences in the relative order of the next round of driver mutations. During most of this process, the lineages founded by different driver mutations make up less than  $\sim 1/q$  of the total size of the nose. However, approximately once every  $\sim q$  establishments, an anomalously early driver mutation will occur and reach an  $\mathcal{O}(1)$  fraction of the nose (Desai et al., 2013). This roughly coincides with a fixation event. In order to fix, a lineage must therefore (i) arise in the nose, (ii) persist for  $\sim q$  additional establishment intervals, and (iii) be lucky enough to hitchhike with the special “jackpot” driver event.

Mutation-rate modifiers can be incorporated into this picture in a straightforward way. The key quantities  $\tau_{\text{est}}$  and  $q$  are only weakly dependent on the mutation rate. Provided that  $|\log(r)| \ll \log(s_b/U_b)$ , the modifier versions of these parameters will be similar to the wildtype, and the competition between these lineages will play out according to the same dynamics as above. The modifier is no more or less likely to arise in the nose compared to a neutral mutation. However, provided that it does occur in one of these special genetic backgrounds, a mutator lineage is  $r$  times more likely than a neutral lineage to generate an additional driver, while an antimutator lineage is  $r$  times less likely to do so. Since the modifier lineage must generate  $\sim q$  additional drivers to fix, its overall fixation probability is given by

$$Np_{\text{fix}}(r) \sim r^q. \quad (19)$$

Thus, we see that clonal interference increases the fixation probability of mutators by  $\sim q$  factors of  $r$ . This increase can be substantial for large  $r$ , even when  $q \approx 2$  (see Fig. 2). From our discussion above, we see that this increase is primarily driven by the fact that *multiple* additional drivers are required for fixation.

In the presence of deleterious mutations, a mutator lineage again feels an effective cost of  $U_d(r-1)$ , so it will tend to decline in frequency relative to the other individuals in the nose. In particular, when the next burst of driver mutations arises, the mutator lineage will have decreased in frequency by a factor of  $e^{-U_d(r-1)\tau_{\text{est}}}$ , which makes it that much less likely to survive to the next round. Similarly, an antimutator lineage will have increased in frequency by an analogous factor of  $e^{U_d(1-r)\tau_{\text{est}}}$ . The overall fixation probability then be-

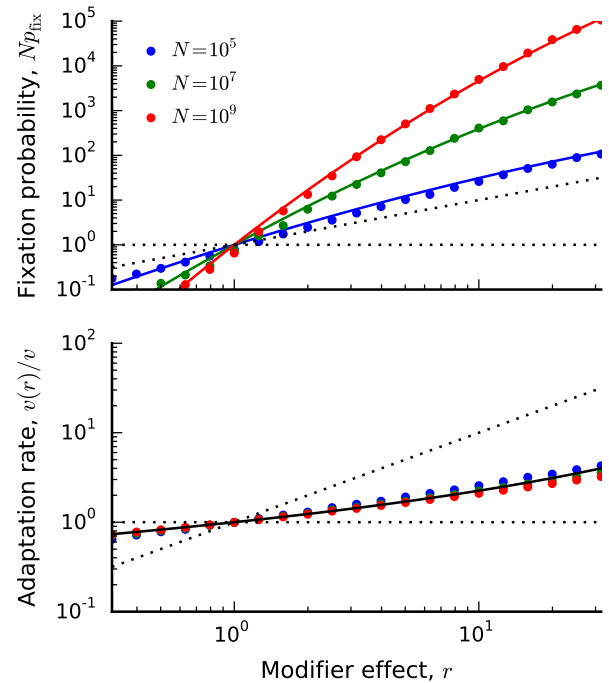


FIG. 2 (Top) The fixation probability of a mutation rate modifier in the clonal interference regime when  $U_d = 0$ . Symbols denote the results of forward-time solutions (described in Appendix C) for  $s_b = 10^{-2}$ ,  $U_b = 10^{-5}$ , and  $N \in \{10^5, 10^7, 10^9\}$ . Solid lines denote the theoretical predictions in Eq. (38). For comparison, the successive mutations prediction ( $Np_{\text{fix}} \approx r$ ) and neutrality ( $Np_{\text{fix}} \approx 1$ ) are shown in the dashed lines. (Bottom) The rate of adaptation of a successful modifier lineage, relative to that of the wildtype, for the same set of populations above. The solid black line denotes the asymptotic prediction,  $v(r)/v \approx \log^2(s_b/U_b)/\log^2(s_b/rU_b)$ , which is independent of  $N$ . For comparison, the successive mutations prediction ( $v(r)/v \approx r$ ) and no change ( $v(r)/v \approx 1$ ) are shown in the dashed lines.

comes

$$Np_{\text{fix}}(r) \sim \left( r e^{-(r-1)U_d\tau_{\text{est}}} \right)^q. \quad (20)$$

We discuss the regimes of validity of this expression in our more rigorous analysis below. For the purposes of our heuristic analysis, it will be more useful to focus on the implications of Eq. (20).

Similar to the selective sweeps case, the direction of selection in Eq. (20) is again determined by the product  $U_d\tau_{\text{est}}$ . There are three characteristic regimes of behavior. When  $U_d\tau_{\text{est}} \ll 1$ , mutators will be favored provided that  $U_m = Ur$  is less than a maximum value,

$$U_m^* = \left( \frac{1+\epsilon}{\tau_{\text{est}}} \right) \log \left( \frac{1}{U_d\tau_{\text{est}}} \right). \quad (21)$$

Conversely, when  $U_d\tau_{\text{est}} \gg 1$ , antimutators will be fa-

vored above a minimum value

$$U_m^* = \left( \frac{1 + \epsilon}{\tau_{\text{est}}} \right) U_d \tau_{\text{est}} e^{-U_d \tau_{\text{est}}}. \quad (22)$$

Similar behavior is obtained for  $U_d \tau_{\text{est}}$  is close to one, except that  $U_m^*$  is now given by

$$U_m^* = U \left( 1 + \frac{1 - U_d \tau_{\text{est}}}{2} \right). \quad (23)$$

When  $U_d \tau_{\text{est}} = 1$ , the range of favorable modifiers vanishes, and mutators and antimutators are both selected against. The evolutionarily stable mutation rate is therefore given by

$$\hat{U} = \frac{1 + \epsilon}{\tau_{\text{est}}}. \quad (24)$$

Note that this is actually an implicit relation for  $\hat{U}$ , since  $\tau_{\text{est}}$  depends on  $\hat{U}_b = \epsilon \hat{U} / (1 + \epsilon)$ . Substituting our expression for  $\tau_{\text{est}}$  in Eq. (17) and solving for  $\hat{U}$ , we find that

$$\hat{U} \sim \frac{2s_b \log(Ns_b)}{\log^2\left(\frac{1}{\epsilon}\right)}, \quad (25)$$

where the region of validity for Eq. (17) [and hence Eq. (25)] now becomes

$$1 \ll \sqrt{2 \log(Ns_b)} \ll \log(1/\epsilon) \ll \log 2(Ns_b). \quad (26)$$

This is still a restrictive parameter range for  $\epsilon$ , although it is much broader than Eq. (16) in the successive mutations regime, and it grows larger with increasing  $Ns_b$ .

Provided that these conditions are met, we see that the stable mutation rate in Eq. (25) is only weakly dependent on  $N$  and  $\epsilon$ , and is much more strongly influenced by  $s_b$ . This contrasts with the behavior we found in the successive mutations regime. In addition, we see that unlike in the successive mutations regime, the stable mutation rate ( $\hat{U}$ ) does not necessarily coincide with the maximally permitted mutator or antimutator allele ( $U_m^*$ ) when  $U \neq \hat{U}$ . This can have important consequences for the dynamics of mutation rate evolution in this regime. If the mutation rate starts far above or below  $\hat{U}$ , natural selection can favor modifier alleles that overshoot the stable value by a substantial amount, which can lead to a nonmonotonic approach to the stable point. We will revisit these dynamics in more detail in the Discussion.

### Formal analysis

We now turn to a more formal derivation of the results described above. To do so, we make use of “traveling-wave” formalism developed in previous work (Fisher,

2013; Good and Desai, 2014; Good et al., 2012; Halatschek, 2011; Neher and Shraiman, 2011; Neher et al., 2010). This formalism focuses on the distribution of relative fitness within the population, which we denote by  $f(x)$ , and the fixation probability of a (wildtype) individual with relative fitness  $x$ , which we denote by  $w(x)$ . In the absence of mutator or antimutator alleles, we and others have previously shown that  $f(x)$  and  $w(x)$  satisfy the partial differential equations,

$$-v \frac{\partial f}{\partial x} = (x - U_d - U_b)f + U_b f(x - s_b), \quad (27a)$$

$$v \frac{\partial w}{\partial x} = (x - U_d - U_b)w + U_b w(x + s_b) - \frac{w^2}{2}, \quad (27b)$$

where  $v \equiv s_b / \tau_{\text{est}}$  is the average rate of adaptation (Good and Desai, 2014). Note that Eq. (27) implicitly assumes that the deleterious mutations are purgeable, i.e., they do not fix and constitute a negligible fraction of the population. The rate of adaptation (or equivalently,  $\tau_{\text{est}} \equiv s_b / v$ ) is set by the self-consistency condition,

$$\int f(x)w(x) dx = \frac{1}{N}, \quad (28)$$

which is just another way of saying that the fixation probability of a neutral mutation must be equal to  $1/N$ . For a detailed derivation of these equations, see Good and Desai (2014). Note that because we have assumed that deleterious mutations are purgeable,  $U_d$  enters into Eqs. (27) and (28) only as an overall shift in the mean fitness of the population. If we measure fitnesses using the shifted variable  $\tilde{x} = x - U_d$  (i.e., fitness relative to the average *unloaded* individual), then the evolution equations revert back to the purely beneficial case. We will adopt this convention from now on.

Even in the simplified model of Eq. (27), there are many possible parameter regimes that one may consider (Fisher, 2013). Here we will focus on a particular approximate solution which is thought to be relevant for many microbial evolution experiments. In this regime, the fitnesses of unloaded individuals are approximately normally distributed,

$$f(\tilde{x}) \approx \frac{1}{\sqrt{2\pi v}} e^{-\frac{\tilde{x}^2}{2v}} \theta(x_c - \tilde{x}), \quad (29)$$

with a sharp cutoff at  $\tilde{x} = x_c$ . This corresponds to the “nose” of the fitness distribution described above. Meanwhile, the survival probability  $w(\tilde{x})$  can be approximated by the piecewise form

$$w(\tilde{x}) \approx \begin{cases} 2\tilde{x} & \text{if } \tilde{x} > x_c, \\ 2x_c e^{\frac{\tilde{x}^2 - x_c^2}{2v}} & \text{if } x_c - s_b \leq \tilde{x} \leq x_c, \\ 0 & \text{if } \tilde{x} \leq x_c - s_b. \end{cases} \quad (30)$$



In this context,  $x_c$  can also be thought of as an *interference threshold*. Lineages that are at relative fitness  $x > x_c$  will fix provided they survive drift and establish; this occurs with probability  $2x$ . Below  $x_c$  the fixation probability drops off rapidly because lineages can establish but still be lost to interference. Once a lineage is more than  $s_b$  below  $x_c$ , it can essentially never catch up with the remaining lineages at the nose, and it will therefore have a negligible fixation probability. The location of  $x_c$  is determined by the auxiliary condition

$$1 \approx \frac{U_b}{s_b} \left[ 1 - \frac{s_b}{x_c} \right]^{-1} e^{\frac{x_c s_b}{v} - \frac{s_b^2}{2v}}. \quad (31)$$

After substituting our expressions for  $f(x)$  and  $w(x)$  into Eq. (28), the self-consistency condition becomes

$$\frac{2x_c s_b e^{-\frac{x_c^2}{2v}}}{\sqrt{2\pi v}} \approx \frac{1}{N}. \quad (32)$$

Together with Eq. (31), this completely determines  $v$  and  $x_c$  as a function of the underlying parameters. Asymptotic formulae for these quantities are given in Eqs. (17) and (18); more accurate estimates can be obtained by solving Eqs. (31) and (32) numerically.

We have previously shown that this solution applies whenever  $x_c - s_b \gg \sqrt{v}$ ,  $s_b \gg \sqrt{v}$ , and  $U_b \ll s_b$  (Good and Desai, 2014). The first of these conditions says that the parents of successful lineages must be highly-fit (i.e., that clonal interference is indeed present). The second condition (which also implies the third) says that the vast majority of individuals in the population have roughly the same number of driver mutations (i.e., mutation pressure is not too strong). In terms of the underlying parameters, these conditions become  $1 \ll \log(s_b/U_b) \ll \log(Ns_b) \ll \log^2(s_b/U_b)$ , as we stated after Eq. (17) above.

Once we have set up this travelling-wave formalism, mutation-rate modifiers can be added in a straightforward way. Since the fate of any allele is determined while it is rare, we can neglect the effects of the modifier on the wildtype population, so that  $f(x)$  and  $v$  are unchanged from above. When a modifier allele occurs, it will arise on a genetic background whose fitness is drawn from  $f(x)$ . We then introduce a new function  $w_m(x)$  describing the fixation probability of the modifier allele as a function of its initial fitness. This function satisfies a similar equation as  $w(x)$ , except with a different mutation rate:

$$v \frac{\partial w_m}{\partial \tilde{x}} = [\tilde{x} - (r-1)U_d - rU_b]w_m + rU_b w_m(\tilde{x} + s_b) - \frac{w_m^2}{2}. \quad (33)$$

Thus, in addition to the increase in  $U_b$ , we see that a mutator lineage experiences an effective fitness cost  $U_d(r-1)$

corresponding to its increased deleterious load. This cost enters as a shift in the overall location of  $w_m(\tilde{x})$ , which we can account for by defining a shifted function,  $w_m(\tilde{x}) = \tilde{w}_m(\tilde{x} - U_d(r-1))$ . After this transformation, Eq. (33) is of the same form as Eq. (27b). Then, provided that the modifier mutation rate  $rU_b$  is still in the same regime as  $U_b$ , the solution for the shifted function  $\tilde{w}_m(\tilde{x})$  will have the same form as  $w(\tilde{x})$ ,

$$\tilde{w}_m(\tilde{x}) = \begin{cases} 2\tilde{x} & \text{if } \tilde{x} > x_{cm}, \\ 2x_{cm} e^{\frac{\tilde{x}^2 - x_{cm}^2}{2v}} & \text{if } x_{cm} - s_b \leq \tilde{x} \leq x_{cm}, \\ 0 & \text{if } \tilde{x} \leq x_{cm} - s_b, \end{cases} \quad (34)$$

except with a different interference threshold  $x_{cm}$ , which satisfies

$$1 = \frac{rU_b}{s_b} \left[ 1 - \frac{s_b}{x_{cm}} \right]^{-1} e^{\frac{x_{cm} s_b}{v} - \frac{s_b^2}{2v}}. \quad (35)$$

Like the wildtype interference threshold  $x_c$ , the location of  $x_{cm}$  is independent of  $U_d$ . Since mutators are favored when  $U_d = 0$ , we expect that  $x_{cm} < x_c$  whenever  $r > 1$ . In other words, we expect that mutators have a lower interference threshold, since these lineages generate beneficial mutations more rapidly and, once in the nose, are therefore less likely to be lost to clonal interference.

In order for this solution to apply, it must satisfy the same conditions as the wildtype population:  $x_{cm} - s_b \gg \sqrt{v}$  and  $rU_b \ll s_b$ . This will be true provided that  $x_{cm}$  is still close to  $x_c$ , or equivalently, that the fractional difference  $\delta_x = x_{cm}/x_c - 1$  is small compared to 1. Given this assumption, we can divide Eq. (35) by Eq. (31) and solve for  $\delta_x$ :

$$\delta_x = -\frac{v}{x_c s_b} \log \left[ \frac{r(1 + \delta_x)}{1 + \frac{x_c \delta_x}{x_c - s_b}} \right] \approx -\frac{v}{x_c s_b} \log(r). \quad (36)$$

Since  $x_c s_b/v \sim \log(s_b/U_b)$  in this regime, we see that this approximation will hold provided that  $|\log(r)| \ll \log(s_b/U_b)$ . This places an upper limit on the range of modifier effects that we can consider. But since  $s_b/U_b \gg 1$ , this includes many realistically-large modifiers of order  $r \sim 100$ .

Given our solution for  $w_m(\tilde{x})$ , we can compute the marginal fixation probability of the modifier by averaging over all the fitness backgrounds that the modifier could have arisen on:

$$p_{\text{fix}}(r) = \int f(\tilde{x}) w_m(\tilde{x}) d\tilde{x}. \quad (37)$$

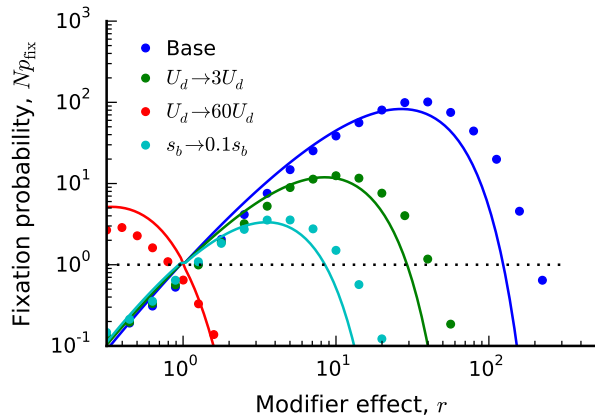


FIG. 3 The fixation probability of a mutation rate modifier in the clonal interference regime when  $U_d > 0$ . Symbols denote the results of forward-time solutions with the base parameters  $N = 10^7$ ,  $s_b = 10^{-2}$ ,  $U_b = 10^{-5}$ ,  $U_d = 10^{-4}$ , and  $s_d = 10^{-1}$ , with modifications listed in the legend. Solid lines denote the theoretical predictions in Eq. (38).

We evaluate this integral in Appendix B. We find that

$$Np_{\text{fix}}(r) \approx e^{q[\log(r) - U_d \tau_{\text{est}}(r-1)]} \times e^{-\frac{\nu}{2}[\log(r) - U_d \tau_{\text{est}}(r-1)]^2} \times \frac{e^{U_d \tau_{\text{est}}(r-1)} - 1}{U_d \tau_{\text{est}}(r-1)}, \quad (38)$$

where we have employed the short-hand notation  $q = x_c/s_b$ ,  $\tau_{\text{est}} = s_b/v$ , and  $\nu = 1/s_b \tau_{\text{est}}$ . In the asymptotic regime we are considering,  $1/q$  and  $\nu$  are both small parameters. To leading order, Eq. (38) converges to the simpler expression from our heuristic analysis, which we now recognize as the technical statement that

$$\lim_{\substack{q \rightarrow \infty \\ \nu \rightarrow 0}} \frac{\log Np_{\text{fix}}(r)}{q} = \log(r) - U_d \tau_{\text{est}}(r-1). \quad (39)$$

However, since the errors are multiplied by a large number  $q$  and exponentiated, the full version in Eq. (38) is often required for quantitative accuracy.

We illustrate the full expression in Eq. (38) and compare it to forward-time simulations in Figs 2 and 3. The accuracy is generally good, although there are some systematic deviations for large  $r$  when the effects of the deleterious load are particularly costly. These are ultimately the result of two factors: (i) inaccuracies in our approximate solution for  $w(x)$  for  $x_c - s_b < x < x_c$  and (ii) the fact that  $\log(r)$  is starting to approach  $\log(s_b/U_b)$ . More accurate approximations for  $w(x)$  derived by Fisher (2013) could be used to improve the quantitative accuracy for these parameters; we leave this for future work.

## DISCUSSION

In rapidly adapting populations, the fates of mutator and antimutator alleles depend on a careful balance between the benefits of hitchhiking and the cost of a deleterious load. Our analysis shows how the fixation probabilities of these mutation-rate modifiers depend on the population size, the beneficial and deleterious mutation rates, and the strength of selection. By searching for parameters where mutators and antimutators are both disfavored by selection, we can identify particular mutation rates  $\hat{U}$  which are stable under future evolution. At these mutation rates, the benefits of hitchhiking are exactly balanced by the costs of the deleterious load, so that neither mutators nor antimutators are favored. We now consider how a population will approach the stable mutation rates, and the implications of this mutation rate evolution for the rate of adaptation within the population. Finally, we describe the approximations we have made throughout our analysis and the resulting limitations of our approach.

### Approach to the stable mutation rate

Consider a situation in which the population starts with a mutation rate  $U_0 \neq \hat{U}$ , and fixes a sequence of modifier alleles with mutation rates  $U_1, U_2, \dots, U_n$ . What are the typical values of this mutation rate trajectory, under the hypothesis that each step was favored by selection?

In both the successional mutations and clonal interference regimes, we found that  $\hat{U}$  depends only on  $N$ ,  $s_b$ , and  $\epsilon$ . This implies that the mutation rate will always evolve towards this unique stable point:  $U_n \rightarrow \hat{U}$ . However, the manner in which the population approaches  $\hat{U}$  can be strongly influenced by clonal interference.

In the successional mutations regime, selection favors a monotonic approach to  $\hat{U}$ . Consider, e.g., a mutation rate  $U_0 < \hat{U}$ . In this case, we have shown that mutators will be able to fix provided that the new mutation rate  $U_m = rU$  lies in the range  $U_0 < U_m < \hat{U}$ . When one of these alleles fixes, the new mutation rate will be given by  $U_1 = U_m$ , and the process will repeat itself. Mutators will still be favored to fix provided that  $U_1 < U_m < \hat{U}$ , which will lead to a  $U_2 > U_1$ , and so on. Thus, evolution will favor a monotonic sequence of mutator alleles until  $U_n = \hat{U}$ , after which the mutation rate will be stable to further changes. An exactly analogous conclusion holds if  $U_0$  starts above  $\hat{U}$ : here selection will tend to fix antimutator alleles that lie in the range  $\hat{U} < U_m < U_i$  until  $U_n = \hat{U}$ . In both cases, mutator or antimutator alleles that “overshoot” the stable mutation rate are always disfavored. This is true even if such alleles would move the mutation rate closer to the stable rate; e.g., an allele

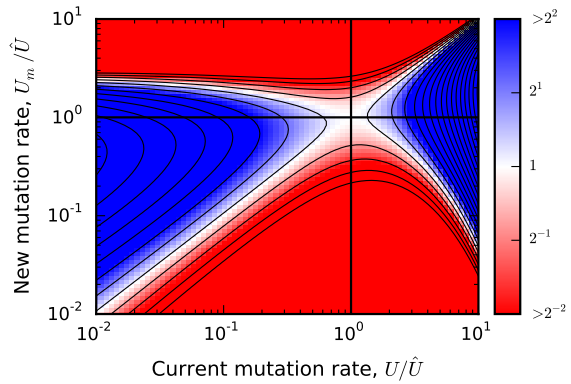


FIG. 4 The predicted fixation probability of a modifier allele when  $Ns_b = 10^6$  and  $\epsilon = 10^{-6}$ . Grid points are colored according to the value of  $Np_{\text{fix}}$  from Eq. (38), and are capped at a maximum value of  $|\log_2 Np_{\text{fix}}| = 2$  to maintain contrast. For comparison, the solid lines denote twenty log-spaced contours that range from  $10^{-1}$  to  $10^4$ .

that takes a population from  $0.1\hat{U} \rightarrow 1.1\hat{U}$  would not be positively selected.

When clonal interference is present, the approach to the stable mutation rate is more complex. This is easiest to see if we rewrite Eq. (20) in terms of  $U_m$ ,  $U$ , and  $\hat{U}$ . Provided that  $|\log(U_i/\hat{U})| \ll \log(1/\epsilon)$ , the quantities  $\tau_{\text{est}}(U)$  and  $q(U)$  will stay roughly constant over the relevant range of mutation rates, and we can approximate  $\tau_{\text{est}}(U) \approx \tau_{\text{est}}(\hat{U}) \equiv 1/\hat{U}$ . This yields a simple heuristic formula,

$$Np_{\text{fix}}(U_m) \approx \left( \frac{U_m e^{-U_m/\hat{U}}}{U e^{-U/\hat{U}}} \right)^q, \quad (40)$$

which depends only on the scaled mutation rates  $U/\hat{U}$ ,  $U_m/\hat{U}$ , and  $q$ . A more accurate (but more cumbersome) version can be obtained by substituting the full expressions for  $\tau_{\text{est}}(U)$  and  $q(U)$  into Eq. (38). We illustrate this full expression in Fig. 4 for a particular combination of  $Ns_b$  and  $\epsilon$ , although the important qualitative features are already contained in Eq. (40).

If the population starts at a mutation rate  $U_0 \sim \hat{U}$ , then only modest changes in the mutation rate will be favorable. However, if the population starts at a mutation rate  $U_0 \ll \hat{U}$ , Eq. (40) shows that mutators will be positively selected provided that  $U_0 < U_m < \hat{U} \log(\hat{U}/U_0)$ . While the most strongly beneficial modifier has  $U_m = \hat{U}$ , the range of favored mutators also includes values of  $U_m$  that are *larger* than  $\hat{U}$  by a factor  $\log(\hat{U}/U_0) \gg 1$ . This means that selection can favor mutator alleles that overshoot the stable mutation rate by a substantial amount. If a mutator allele of this maximal strength fixes, the new mutation rate will be  $U_1 = \hat{U} \log(\hat{U}/U_0) > \hat{U}$ , and

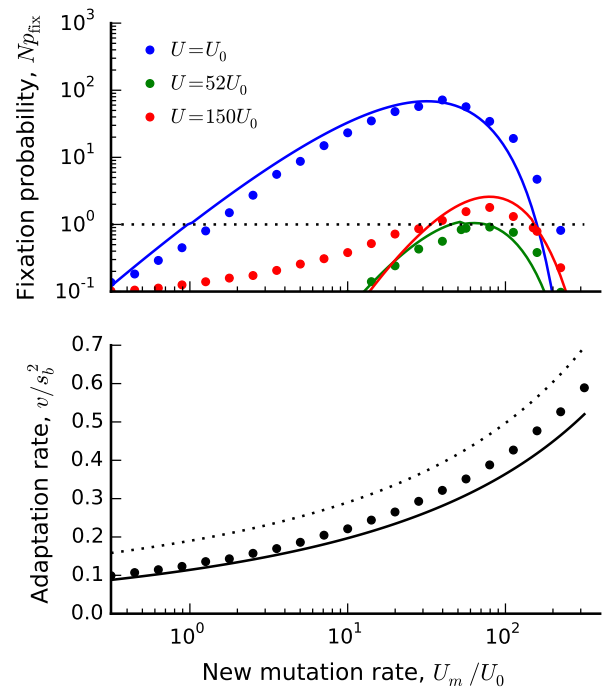


FIG. 5 (Top) A vertical “slice” of Fig. 4 for three different values of  $U$ . Symbols denote the results of forward time simulations with  $NU_0 = 5780$ , and  $s_d = 5s_b$ ; other parameters are the same as Fig. 4. Solid lines denote theoretical predictions from Eq. (38). (Bottom) The scaled rate of adaptation as a function of the mutation rate for the same set of parameters. Symbols denote the results of forward-time simulations, the solid lines show the theoretical predictions obtained by solving Eqs. (31) and (32) numerically, and the dashed line shows the asymptotic formula,  $v/s_b^2 \sim 2 \log(Ns_b)/\log(s_b/\epsilon U_m)^2$ .

selection will immediately favor the fixation of antimutator alleles, despite the fact that the location of  $\hat{U}$  has not changed (see Fig. 5). If  $U \gg \hat{U}$ , these antimutator alleles can overshoot the stable mutation rate in the other direction, by a factor  $U/\hat{U} e^{-U/\hat{U}} \ll 1$ . In the example above, this could then lead to  $U_2 = U_0 \log(\hat{U}/U_0)$ , which is larger than the initial mutation rate  $U_0$  but still much smaller than  $\hat{U}$ . Thus while the population will still approach the stable mutation rate over time, this approach does not have to be monotonic, even in the absence of epistasis or environmental variation. Moreover, the functional form of Eq. (40) implies that this behavior is highly asymmetric: an antimutator allele can overshoot  $\hat{U}$  by a larger amount (and with a larger  $Np_{\text{fix}}$ ) than a mutator allele with the same value of  $|\log(U/\hat{U})|$ . This can also be seen in Figs 4 and 5.

## Application to a long-term evolution experiment in *E. coli*

Several studies have observed the spread of mutator alleles in microbial populations in the laboratory. One of the best-studied examples is Lenski's long-term evolution experiment (the LTEE), where 12 replicate populations of *E. coli* have been propagated constant conditions for more than 60,000 generations (Lenski et al., 1991, 2015). Mutator phenotypes have fixed in 6 of the 12 populations over the course of this experiment, and have typically increased the mutation rate by  $\sim 100$ -fold. The rate of adaptation in the mutator populations increased less than 2-fold, suggesting that clonal interference plays an important role (Wiser et al., 2013). In one of the 12 replicates, the dynamics of mutation rate evolution have been studied in more detail. Wielgoss et al. (2013) have shown that, soon after the fixation of the mutator allele, an antimutator phenotype fixed, which lowered the mutation rate by a factor of two. The antimutator phenotype had two independent origins, suggesting that both it and the original mutator allele were favored by selection. Given these findings, it is interesting to compare various explanations for this behavior in the context of our theoretical analysis above.

One potential explanation for these findings is suggested by the "overshooting" behavior illustrated in Figs 4 and 5. Although the precise evolutionary parameters of these populations are not known, Fig. 5 shows that it is at least feasible to overshoot the optimum by a factor of 2 when the population starts at a mutation rate of  $U_0 \sim 0.01\hat{U}$ . Note, however, that  $Np_{\text{fix}}$  is not necessarily large in the reverse direction, so it remains to be seen whether this effect could efficiently select for lower mutation rates on the rapid timescales observed.

Another potential explanation for the mutation-rate reversal, initially suggested by Wielgoss et al. (2013), is that long-term epistasis reduced the effective advantage of hitchhiking later in the experiment, thereby shifting the location of  $\hat{U}$  below its initial value. In the context of our model, this epistasis can manifest itself in one of two ways. First, it could reduce  $\epsilon$ , reflecting a diminished supply of beneficial mutations (or an increasing supply of deleterious mutations) as the population adapts. Our formula in Eq. (25) suggests that rather drastic reductions in  $\epsilon$  are required to cause a 2-fold reduction in  $\hat{U}$ . Alternatively, diminishing-returns epistasis could reduce the magnitude of  $s_b$ , while leaving the beneficial mutation rate unchanged. Equation (25) suggests that this has much stronger effect on  $\hat{U}$ . Wiser et al. (2013) have recently proposed and fit a concrete model for how  $s_b$  declines with fitness in the LTEE. These estimates suggest that  $s_b$  declines by  $\lesssim 20\%$  in the 4,000 generations that separate the mutator and antimutator alleles. This would seem to be too small to account for a  $\gtrsim 50\%$  reduction

in  $\hat{U}$ , although the magnitudes are sufficiently close that a careful comparison with simulations is required. This would be an interesting avenue for future work.

A third potential explanation is a direct fitness benefit to either the mutator or antimutator allele. While such benefits are difficult to measure directly (due to the confounding effects of the deleterious load), they can be incorporated into our model in a straightforward way. We can simply replace  $U_d(r-1) \rightarrow U_d(r-1) - s_m$  in our formulae for  $p_{\text{fix}}(r)$ , where  $s_m$  is the direct benefit of the modifier allele. Since the effect of the deleterious load is typically of order  $\hat{U} < s_b$ , even a small direct benefit ( $s_m < s_b$ ) can shift the balance between hitchhiking and load in important ways.

## The evolution of mutation rates and the average rate of adaptation

Because we have assumed that deleterious mutations are purgeable, increasing  $U$  always increases the rate of adaptation, even though these increases may be small in the presence of clonal interference. Our results therefore suggest that mutation rates will evolve toward a stable mutation rate that is less than what would be optimal for the population (which, in this case, is technically  $U_{\text{opt}} = \infty$ ). Of course, as we increase mutation rates, the assumption that deleterious mutations are purgeable will eventually fail. Somewhere above this point, there will be an optimal mutation rate  $U_{\text{opt}} < \infty$  that maximizes the rate of adaptation (see, e.g. Orr (2000)). To confirm that the stable mutation rate is indeed below  $U_{\text{opt}}$ , we must verify that  $\hat{U}$  is still in the purgeable deleterious mutations regime. For the example in Fig. 5, we can verify this directly, since the rate of adaptation continues to increase as the mutation rate passes through  $\hat{U}$ . This behavior will hold more generally provided that the typical cost of a deleterious mutation,  $s_d$ , is much larger than  $\hat{U}$ . Since  $\hat{U}$  is typically smaller than  $s_b$  in the regimes that we consider, this will indeed be the case whenever  $s_d \gtrsim s_b$ .

In these cases, the stable mutation rate is below the optimal mutation rate, which implies that the dynamics of the evolutionary process can sometimes favor changes in mutation rate that slow the adaptation of the population. Antimutator alleles can be favored even when their fixation will ultimately reduce the overall rate of adaptation. Conversely, mutator alleles can be disfavored even when their fixation would have increased the rate of adaptation.

Although our results are limited to the purgeable regime, a similar distinction between  $\hat{U}$  and  $U_{\text{opt}}$  may also apply even when deleterious mutations are no longer purgeable. We cannot prove this conjecture in our present framework, but it is an interesting hypothesis



for future work.

### Limitations to our analysis

We have made a number of key assumptions throughout our analysis. Most crucially, we have assumed that deleterious mutations are purgeable. For this to be true, two conditions must be met: the deleterious mutations cannot fix, nor can they affect the overall rate of adaptation by reducing the fixation probabilities of beneficial mutations. This is a slightly stronger condition than the “ruby in the rough” approximation used by earlier authors (Charlesworth, 1994; Peck, 1994). Our previous work on the effects of deleterious mutations in adapting populations provides a detailed analysis of when these conditions will hold, and of the effects of deleterious mutations on adaptation when the purgeable assumption fails (Good and Desai, 2014). This earlier work suggests that deleterious mutations are likely to be purgeable in most microbial evolution experiments, though recent experimental work hints that this may not always be the case (McDonald et al., 2016). Here we first summarize the conditions under which deleterious mutations are purgeable, and then describe the potential effects of deleterious mutations when this condition is violated.

In our earlier work, we showed that deleterious mutations with fitness cost  $s_d$  will typically not fix provided that  $s_d \gg 1/T_c$ , where  $T_c \sim 1/q\tau_{\text{est}}$  is the coalescence timescale (Good and Desai, 2014). Since  $s_b \gg 1/T_c$  in all the regimes that we consider, this will be true provided that  $s_d$  is not much smaller than  $s_b$ . Deleterious mutations can also hinder the fixation of beneficial mutations that arise in less-fit backgrounds. However, provided that  $s_d \gg U_d$ , deleterious variants will be rapidly eliminated from the population, and most genetic backgrounds will be free of deleterious mutations. Assuming typical values of  $U_d \sim 10^{-4}$  and  $s_d \sim 10^{-2} - 10^{-1}$  for microbial evolution experiments (Wloch et al., 2001), this condition will often be met.

When deleterious mutations have small enough fitness costs that they are not purgeable, our quantitative results all break down. However, sufficiently weakly deleterious mutations cannot affect mutator or antimutator dynamics on the relevant timescales, so they are effectively neutral from the point of view of mutation rate evolution. On the other hand, sufficiently strongly deleterious mutations are purgeable. Thus it is only some range of deleterious mutations between these limits whose effects are both important to mutation rate evolution and not described by our analysis. Since these mutations are less deleterious than purgeable mutations, they must by definition be less unfavorable to mutator alleles (and less favorable to antimutators). Thus they do not affect

the fate of a modifier of mutation rates as strongly as a corresponding purgeable mutation would. This suggests that we can qualitatively understand the effects of non-purgeable deleterious mutations in adapting populations by weighting them less heavily than purgeable ones, so that the total effective deleterious mutation rate is actually somewhat less than  $U_d$ . Of course, this is only a rough intuition. To analyze the effects of non-purgeable deleterious mutations more fully, we need to include them in our traveling wave framework. Our earlier work explains how these mutations affect  $f(x)$ ,  $w(x)$ ,  $x_c$ , and  $v$ , which provides a potential starting point for these calculations (Good and Desai, 2014). More generally, we may wish to consider the evolution of mutation rates in the limit where non-purgeable mutations become so important that the population no longer adapts, and instead Muller’s ratchet plays a crucial role. These are interesting but complex directions for future work.

In addition to these key limitations of our analysis, we have also made a number of more technical assumptions. For example, we have assumed that beneficial mutations all provide the same fitness benefit  $s_b$ . In reality, beneficial mutations will have a range of different fitness effects, drawn from some distribution of fitness effects. However, earlier work has shown that in this case the evolutionary dynamics can be summarized using a single effective beneficial fitness effect and corresponding effective beneficial mutation rate (Desai and Fisher, 2007; Fisher, 2013; Good et al., 2012). Thus our analysis can be applied to this situation using the appropriate effective  $s_b$  and  $U_b$ .

In our analysis of clonal interference, we focused on a regime in which  $1 \ll \log(s_b/U_b) \ll 2 \log(Ns_b) \ll \log^2(s_b/U_b)$ . The latter condition implies that clonal interference is not exceptionally strong, and is often a good approximation for microbial evolution experiments at wildtype mutation rates. For some large effect mutators, we start to approach the boundary of this regime, and our quantitative expressions become less accurate (see Fig. 5). In principle, we could extend our analysis to the “high-speed” [ $2 \log(Ns_b) \gg \log^2(s_b/U_b)$ ] or “mutational-diffusion” regimes [ $U_b \gg s_b$ ] by using analogous solutions for  $f(x)$  and  $w(x)$  derived by Fisher (2013) or Hallatschek (2011). We leave this for future work.

Throughout our analysis, we have assumed that modifier effect sizes can be large but not exceptionally so [e.g., in the clonal interference regime,  $|\log(r)| \ll \log(s_b/U_b)$ ]. Since  $U_b \ll s_b$ , this includes many realistically large mutators of order  $r \sim 100$ , and in practice, our expressions appear to work reasonably well even when  $\log(r) \sim \log(s_b/U_b)$  (see Fig. 3). For sufficiently large  $r$ , we may also encounter situations where modifiers switch from one regime to another (e.g., from the clonal interference to successive mutations regime, or from  $U_b \ll s_b$  to  $U_b > s_b$ ). Our quantitative predictions for  $Np_{\text{fix}}$  break

down in all of these cases. However, the evolutionarily stable mutation rates are defined by the behavior of  $Np_{\text{fix}}$  in the local neighborhood of  $r \approx 1$ , so the location of  $\hat{U}$  is not influenced by this problem. If the population starts sufficiently far from  $\hat{U}$ , then the fixation probabilities of the first few  $U_i$  are not well-described by our analysis, but we know that the mutation rate must still approach  $\hat{U}$  (in a possibly non-monotonic manner). Eventually, the mutation rate will become close enough to  $\hat{U}$  that our expressions start to apply, and the remaining steps of the mutation trajectory can be predicted.

Finally, we have assumed throughout that modifier mutations are sufficiently rare that their fates are determined independently. In other words, we have neglected clonal interference *between* different modifier mutations. For deleterious or weakly beneficial modifiers, this will often be a good approximation. However, for strongly beneficial modifiers, this requires that the establishment time  $1/N\mu p_{\text{fix}}(r)$  between successive mutators is large compared to the fixation time. In the clonal interference regime,  $T_{\text{fix}} \sim \log(s_b/U_b)/s_b$ , and this condition becomes  $\mu \ll s_b/\log(s_b/U_b)Np_{\text{fix}}(r)$ . This can sometimes be violated for strongly beneficial mutator alleles with a large target size (e.g., loss-of-function mutations in multiple genes). In these cases, our quantitative predictions become inaccurate, although the overall direction of selection [i.e., whether  $Np_{\text{fix}} > 1$  or  $Np_{\text{fix}} < 1$ ] will remain unchanged.

## Conclusions

Our analysis has explained how the interplay between beneficial and deleterious mutations in adapting populations creates indirect selection pressures on modifiers of mutation rates, and we have shown how these indirect selection pressures affect the fates of mutator and antimutator alleles. Our analysis of the successional mutations regime follows the logic of earlier work (Andre and Godelle, 2006; Desai and Fisher, 2011), balancing the probability a new beneficial mutation arises in a mutator or antimutator background with the effects of the modifier allele on the accumulation of deleterious load. We have also studied rapidly adapting populations where clonal interference is widespread. Our approach to this question builds on the traveling wave framework we and others have recently introduced (Fisher, 2013; Good and Desai, 2014; Good et al., 2012; Hallatschek, 2011; Neher and Shraiman, 2011; Neher et al., 2010). In this framework, analyzing the fate of an allele modifying mutation rate is similar in spirit to calculating the fixation probability of any lineage: we must solve the same equation for  $w(x)$ , except instead of a mutation changing the fitness  $x$ , it changes the mutation rate  $U$ . This general

framework can also be applied to analyze indirect selection pressures that act on other modifiers of the evolutionary process. For example, we could analyze the fate of a mutation that changes the distribution of fitness effects of new mutations by solving for  $w(x)$  for an allele that modifies  $\rho(s)$ . Our analysis does not need to be limited to adapting populations, since the traveling wave framework applies whenever interference selection is widespread, even if the population is not adapting on average. Of course, in practice, some modifiers and parameter regimes will lead to equations for  $w(x)$  that are analytically tractable, while others will not. Thus further work is needed to more fully understand both the limits and promise of this approach.

## ACKNOWLEDGMENTS

We thank Ivana Cvijović for useful discussions and helpful comments on the manuscript. Simulations in this article were run on the Odyssey cluster supported by the FAS Division of Science Research Computing Group at Harvard University. This work was supported in part by the Simons Foundation (grant 376196), grant PHY 1313638 from the NSF, and grant GM104239 from the NIH.

## REFERENCES

- ANDRE, J.-B. AND GODELLE, B. 2006. The evolution of mutation rate in finite asexual populations. *Genetics* 172:611–626.
- BJORKHOLM, B., SJOLUND, M., FALK, P. G., BERG, O. G., ENGSTRAND, L., AND *et al* 2001. Mutation frequency and biological cost of antibiotic resistance in *helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* 98:14607–14612.
- CHAO, L. AND COX, E. C. 1983. Competition between high and low mutating strains of *Escherichia coli*. *Evolution* 37:125–134.
- CHAO, L., VARGAS, C., SPEAR, B. B., AND COX, E. C. 1983. Transposable elements as mutator genes in evolution. *Nature* 303:633–635.
- CHARLESWORTH, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research* 63:213–227.
- DAWSON, K. J. 1998. Evolutionarily stable mutation rates. *J Theor Biol* 194:147–157.
- DAWSON, K. J. 1999. The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theor Popul Biol* 55:1–22.
- DEL CAMPO, R., MOROSINI, M. I., DE LA PEDROSA, E. G. G., FENOLL, A., MUNOZ-ALMAGRO, C., AND *et al* 2005. Population structure, antimicrobial resistance, and mutation frequencies of *Streptococcus pneumoniae* isolates from cystic fibrosis patients. *J. Clin. Microbiol* 43:2207–2214.

- DENAMUR, E., BONACORSI, S., GIRAUD, A., DURIEZ, P., HILALI, F., AND *et al* 2002. High frequency of mutator strains among human uropathogenic *Escherichia coli* isolates. *J. Bacteriol.* 184:605–609.
- DESAI, M. M. AND FISHER, D. S. 2007. Beneficial mutation selection balance and the effect of genetic linkage on positive selection. *Genetics* 176:1759–1798.
- DESAI, M. M. AND FISHER, D. S. 2011. The balance between mutators and nonmutators in asexual populations. *Genetics* 188:997–1014.
- DESAI, M. M., WALCZAK, A. M., AND FISHER, D. S. 2013. Genetic diversity and the structure of genealogies in rapidly adapting populations. *Genetics* 193:565–585.
- EWENS, W. J. 2004. *Mathematical Population Genetics*. Springer-Verlag, New York, second edition.
- FISHER, D. S. 2007. Evolutionary dynamics, pp. 395–446. In M. M. Jean-Philippe Bouchaud and J. Dalibard (eds.), *Complex Systems*, volume 85 of *Les Houches*. Elsevier.
- FISHER, D. S. 2013. Asexual evolution waves: fluctuations and universality. *J Stat Mech* 2013:P01011.
- GARDINER, C. 1985. *Handbook of Stochastic Methods*. Springer, New York.
- GENTILE, C. F., YU, S.-C., SERRANO, S. A., GERRISH, P. J., AND SNIEGOWSKI, P. D. 2011. Competition between high- and higher-mutating strains of *Escherichia coli*. *Biol Lett* 7:422–424.
- GILLESPIE, J. H. 1981. Mutation modification in a random environment. *Evolution* 35:468–476.
- GIRAUD, A., MATIC, I., RADMAN, M., FONS, M., AND TADDEI, F. 2002. Mutator bacteria as a risk factor in treatment of infectious diseases. *Antimicrob. Agents Chemother.* 46:863–865.
- GIRAUD, A., MATIC, I., TENAILLON, O., CLARA, A., RADMAN, M., AND *et al* 2001. Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* 291:2606–2608.
- GOOD, B. H. AND DESAI, M. M. 2014. Deleterious passengers in adapting populations. *Genetics* 198:1183–1208.
- GOOD, B. H., ROUZINE, I. M., BALICK, D. J., HALLATSCHKEK, O., AND DESAI, M. M. 2012. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl. Acad. Sci.* 109:4950–4955.
- GOOD, B. H., WALCZAK, A. M., NEHER, R. A., AND DESAI, M. M. 2014. Genetic diversity in the interference selection limit. *PLoS Genetics* 10:e1004222.
- HALLATSCHKEK, O. 2011. The noisy edge of traveling waves. *Proc. Natl. Acad. Sci. USA* 108:1783–1787.
- ISHII, K., MATSUDA, H., IAWASA, Y., AND SASAKI, A. 1989. Evolutionarily stable mutation rate in a periodically changing environment. *Genetics* 121:163–174.
- JAMES, A. AND JAIN, K. 2015. Fixation probability of rare nonmutator and evolution of mutation rates. *Ecology and Evolution* 6:755–764.
- JOHNSON, T. 1999a. The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *Proc. R. Soc. Lond. Ser. B* 266:2389–2397.
- JOHNSON, T. 1999b. Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. *Genetics* 151:1621–1631.
- KIMURA, M. 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genet Res* 9:23–34.
- LABAT, F., PRADILLON, O., GARRY, L., PEUCHMAUR, M., FANTIN, B., AND *et al* 2005. Mutator phenotype confers advantage in *Escherichia coli* chronic urinary tract infection pathogenesis. *FEMS Immunol. Med. Microbiol.* 44:317–321.
- LECLERC, J. E., LI, B. G., PAYNE, W. L., AND CEBULA, T. A. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274:1208–1211.
- LEE, H., POPODI, E., TANGA, H., AND FOSTER, P. L. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc Natl Acad Sci* 109:E2774–E2783.
- LEIGH, E. G. 1970. Natural selection and mutability. *Am Nat* 104:301.
- LEIGH, E. G. 1973. The evolution of mutation rates. *Genetics* 73:1–18.
- LENSKI, R. E., ROSE, M. R., SIMPSON, S. C., AND TADLER, S. C. 1991. Long-term experimental evolution in *Escherichia coli*. i. adaptation and divergence during 2,000 generations. *The American Naturalist* 138:1315–1341.
- LENSKI, R. E., WISER, M. J., RIBECK, N., BLOUNT, Z. D., NAHUM, J. R., AND *et al* 2015. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with *Escherichia coli*. *Proceedings of the Royal Society, London B* 282:20152292.
- LIBERMAN, U. AND FELDMAN, M. W. 1986. Modifiers of mutation rate: a general reduction principle. *Theor Popul Biol* 30:125–142.
- LYNCH, M. 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180:933–943.
- LYNCH, M. 2011. The lower bound to the evolution of mutation rates. *Genome Biol. Evol.* 3:1107–1118.
- MATIC, I., RADMAN, M., TADDEI, F., PICARD, B., DOIT, C., BINGEN, E., DENAMUR, E., AND ELION, J. 1997. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* 277:1833–1834.
- MCDONALD, M. J., HSIEH, Y. Y., YU, Y. H., CHANG, S. L., AND LEU, J. Y. 2012. The evolution of low mutation rates in experimental mutator populations of *Saccharomyces cerevisiae*. *Current Biology* 22:1235–1240.
- MCDONALD, M. J., RICE, D. P., AND DESAI, M. M. 2016. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* 531:233–236.
- NEHER, R. AND SHRAIMAN, B. I. 2011. Genetic draft and quasi-neutrality in large facultatively sexual populations. *Genetics* 188:975–976.
- NEHER, R. AND SHRAIMAN, B. I. 2012. Fluctuations of fitness distributions and the rate of muller’s ratchet. *Genetics* 191:1283–1293.
- NEHER, R., SHRAIMAN, B. I., AND FISHER, D. S. 2010. Rate of adaptation in large sexual populations. *Genetics* 184:467–481.
- NEHER, R. A., KESSINGER, T. A., AND SHRAIMAN, B. I. 2013. Coalescence and genetic diversity in sexual populations under selection. *Proc Nat Acad Sci* 110:15836–15841.
- NOTLEY-MCROBB, L., SEETO, S., AND FERENCI, T. 2002. Enrichment and elimination of mutator mutants in *Escherichia coli* populations. *Genetics* 162:1055–1062.
- OLIVER, A., CANTON, R., CAMPO, P., BAQUERO, F., AND BLAZQUEZ, J. 2000. High frequency of hypermutable *Pseu-*

- domonas aeruginosa* in cystic fibrosis lung infection. *Science* 288:1251–1253.
- ORR, H. A. 2000. The rate of adaptation in asexuals. *Genetics* 155:961–968.
- PAINTER, P. R. 1975. Mutator genes and selection for the mutation rate in bacteria. *Genetics* 79:649–660.
- PAL, C., MACIA, M. D., OLIVER, A., SCHACHAR, I., AND BUCKLING, A. 2007. Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* 450:1079–1081.
- PECK, J. R. 1994. A ruby in the rubbish. *Genetics* 137:597–606.
- PRUNIER, A. L., MALBRUNY, B., LAURANS, M., BROUARD, J., DUHAMEL, J. F., AND *et al* 2003. High rate of macrolide resistance in *Staphylococcus aureus* strains from patients with cystic fibrosis reveals high proportions of hypermutable strains. *J. Infect. Dis.* 187:1709–1716.
- RICHARDSON, A. R., YU, Z., POPOVIC, T., AND STOJILJKOVIC, I. 2002. Mutator clones of neisseria meningitidis in epidemic serogroup a disease. *Proc. Natl. Acad. Sci. USA* 99:6103–6107.
- SHAVER, A. C., DOMBROWSKI, P. G., SWEENEY, J. Y., TREIS, T., ZAPPALA, R. M., AND *et al* 2002. Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* 162:557–566.
- SNIEGOWSKI, P. D., GERRISH, P. J., AND LENSKI, R. E. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387:703–705.
- SODERBERG, R. J. AND BERG, O. G. 2011. Kick-starting the ratchet: the fate of mutators in asexual populations. *Genetics* 197:1129–1137.
- TADDEI, F., RADMAN, M., MAYNARD-SMITH, J., TOUPANCE, B., GOUYON, P. H., AND *et al* 1997. Role of mutator alleles in adaptive evolution. *Nature* 387:700–702.
- TANAKA, M. M., BERGSTROM, C. T., AND LEVIN, B. R. 2003. The evolution of mutator genes in bacterial populations: the roles of environmental change and timing. *Genetics* 164:843–854.
- TENAILLON, O., NAGARD, H. L., GODELLE, B., AND TADDEI, F. 2000. Mutators and sex in bacteria: conflict between adaptive strategies. *Proc Natl Acad Sci USA* 97:10465–10470.
- TENAILLON, O., TOUPANCE, B., NAGARD, H. L., TADDEI, F., AND GODELLE, B. 1999. Mutators, population size, adaptive landscape, and the adaptation of asexual populations of bacteria. *Genetics* 152:485–493.
- THOMPSON, D. A., DESAI, M. M., AND MURRAY, A. W. 2006. Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. *Curr Biol* 16:1581–1590.
- TRAVIS, J. AND TRAVIS, E. 2002. Mutator dynamics in fluctuating environments. *Proc Biol Sci* 269:591–597.
- TRÖBNER, W. AND PIECHOCKI, R. 1981. Competition growth between *Escherichia coli* mut<sup>l</sup> and mut<sup>+</sup> in continuously growing cultures. *Z Allg Mikrobiol* 21:347–349.
- TRÖBNER, W. AND PIECHOCKI, R. 1984. Selection against hypermutability in *Escherichia coli* during long term evolution. *Mol Gen Genet* 198:177–178.
- TURRIENTES, M.-C., BAQUERO, F., LEVIN, B. R., MARTINEZ, J.-L., RIPOLL, A., AND *et al* 2013. Normal mutation rate variants arise in a mutator (mut s) *Escherichia coli* population. *PLoS ONE* 8:e72963.
- VOORDECKERS, K., KOMINEK, J., DAS, A., ESPINOSA-CANTÚ, A., MAEYER, D. D., ARSLAN, A., AND *et al* 2015. Adaptation to high ethanol reveals complex evolutionary pathways. *PLoS Genet* 11:e1005635.
- WATSON, M. E., BURNS, J. L., AND SMITH, A. L. 2004. Hypermutable haemophilus influenzae with mutations in muts are found in cystic fibrosis sputum. *Microbiol. Sgm.* 150:2947–2958.
- WEISSMAN, D. AND HALLATSCHKE, O. 2014. The rate of adaptation in large sexual populations with linear chromosomes. *Genetics* 196:1167–1183.
- WEISSMAN, D. W., DESAI, M. M., FISHER, D. S., AND FELDMAN, M. W. 2009. The rate at which asexual populations cross fitness valleys. *Theoretical Population Biology* 75:286–300.
- WIELGOSS, S., BARRICK, J. E., TENAILLON, O., CRUVEILLER, S., CHANE-WOON-MING, B., MÉDIGUE, C., LENSKI, R. E., SCHNEIDER, D., AND ANDREWS, B. J. 2011. Mutation rate inferred from synonymous substitutions in a long-term evolution experiment with *Escherichia coli*. *G3 (Bethesda)* 1:183–186.
- WIELGOSS, S., BARRICK, J. E., TENAILLON, O., WISER, M. J., DITTMAR, J., CRUVEILLER, S., CHANE-WOON-MING, B., MÉDIGUE, C., LENSKI, R. E., AND SCHNEIDER, D. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci* 110:222227.
- WISER, M. J., RIBECK, N., AND LENSKI, R. E. 2013. Long-term dynamics of adaptation in asexual populations. *Science* 342:1364–1367.
- WLOCH, D. M., SZAFRANIEC, K., BORTS, R. H., AND KORONA, R. 2001. Direct estimate of the mutation rate and the distribution of fitness effects in the yeast *Saccharomyces cerevisiae*. *Genetics* 159:441–452.
- WYLIE, C. S., GHIM, C.-M., KESSLER, D., AND LEVINE, H. 2009. The fixation probability of rare mutators in finite asexual populations. *Genetics* 181:1595–1612.



### Appendix A: Exact solution of Eq. (3)

In this section, we evaluate the integral

$$Np_{\text{fix}} = \left\langle \int_0^\infty \lambda_m(t) e^{-\int_0^t [\lambda_0(t') + \lambda_m(t')] dt'} dt \right\rangle, \quad (\text{A1})$$

where  $\lambda_0 = 2NU_b s_b [1 - f_m(t)]$  and  $\lambda_m(t) = 2NU_b s_b r f_m(t)$ , and  $f_m(t)$  satisfies the Langevin dynamics

$$\frac{\partial f_m}{\partial t} = \sqrt{\frac{f_m}{N}} \eta(t), \quad (\text{A2})$$

with  $f_m(0) = 1/N$ . Using integration by parts, we can rewrite this integral as

$$Np_{\text{fix}} = \frac{r}{r-1} \left[ 1 - \int_0^\infty \frac{dt}{\tau_{\text{est}}} e^{-\frac{t}{\tau_{\text{est}}}} \left\langle e^{-\frac{r-1}{\tau_{\text{est}}} \int_0^t f_m(t') dt'} \right\rangle \right], \quad (\text{A3})$$

where we have used the short-hand  $\tau_{\text{est}} = 1/2NU_b s_b$ . To evaluate this integral, we must first solve for the generating function  $H(y, t)$  for the *weight*  $\int_0^t f_m(t') dt'$ , which is defined by

$$H(y, t) = \left\langle e^{-y \int_0^t f_m(t') dt'} \right\rangle. \quad (\text{A4})$$

This can be done using techniques described in Weissman et al. (2009). Briefly, we first introduce the joint generating function,

$$H(z, y, t) = \left\langle e^{-z \int_0^t f_m(t') dt' - y \int_0^t f_m(t') dt'} \right\rangle. \quad (\text{A5})$$

By Taylor expanding  $H(z, y, t + dt)$  and using the Langevin dynamics in Eq. (A2), we can show that  $H(z, y, t)$  obeys the partial differential equation

$$\frac{\partial H}{\partial t} = \left[ y - \frac{z^2}{2N} \right] \frac{\partial H}{\partial z}, \quad (\text{A6})$$

which can be solved using the method of characteristics. The characteristic curves satisfy

$$\frac{\partial y}{\partial \sigma} = 0, \quad \frac{\partial z}{\partial \sigma} = \frac{z^2}{2N} - y, \quad \frac{\partial H}{\partial \sigma} = 0, \quad (\text{A7a})$$

and the solution is given by

$$H(z, y, t) = 1 - \frac{\sqrt{2Ny}}{N} \tanh \left[ \frac{t}{2N} \sqrt{2Ny} + \tanh^{-1} \left( \frac{z}{\sqrt{2Ny}} \right) \right]. \quad (\text{A8})$$

The marginal generating function for  $\int_0^t f_m(t') dt'$  can then be obtained by setting  $z = 0$ :

$$H(y, t) = 1 - \frac{\sqrt{2Ny}}{N} \tanh \left( \frac{t}{2N} \sqrt{2Ny} \right). \quad (\text{A9})$$

Substituting this expression into the integral in Eq. (A3), we find that

$$Np_{\text{fix}} = r \int_0^\infty e^{-z} \sqrt{\frac{2N}{\tau_{\text{est}}(r-1)}} \tanh \left( \sqrt{\frac{\tau_{\text{est}}(r-1)}{2N}} z \right) dz, \quad (\text{A10})$$

$$= r \left[ \frac{\Psi \left( \frac{1}{2} + \sqrt{\frac{N}{8\tau_{\text{est}}(r-1)}} \right) - \Psi \left( \sqrt{\frac{N}{8\tau_{\text{est}}(r-1)}} \right) - \sqrt{\frac{2\tau_{\text{est}}(r-1)}{N}}}{\frac{\tau_{\text{est}}(r-1)}{N}} \right], \quad (\text{A11})$$

where  $\Psi(z)$  is the digamma function. Asymptotic expansions for small and large arguments yield

$$Np_{\text{fix}} \sim \begin{cases} r \left[ 1 - \mathcal{O} \left( \frac{\tau_{\text{est}}(r-1)}{N} \right) \right] & \text{if } \tau_{\text{est}} r \ll N, \\ \sqrt{Nr/\tau_{\text{est}}} \left[ 1 - \mathcal{O} \left( \sqrt{\frac{N}{\tau_{\text{est}} r}} \right) \right] & \text{if } \tau_{\text{est}} r \gg N. \end{cases} \quad (\text{A12})$$

## Appendix B: Derivation of Eq. (38)

To obtain an expression for  $Np_{\text{fix}}(r)$ , we must evaluate the integral

$$p_{\text{fix}}(r) = \int f(\tilde{x})w_m(\tilde{x})d\tilde{x} = \int f(\tilde{x})\tilde{w}_m(\tilde{x} - U_d(r-1))d\tilde{x} = \int f(\tilde{x} + U_d(r-1))\tilde{w}_m(\tilde{x})d\tilde{x}, \quad (\text{B1})$$

using the expressions for  $f(\tilde{x})$  and  $\tilde{w}_m(\tilde{x})$  in Eqs. (29) and (34) in the main text. This is straightforward, although the algebra is somewhat tedious. Since  $f(\tilde{x})$  and  $\tilde{w}_m(\tilde{x})$  are piecewise functions, the form of the integrand will depend on whether  $x_c - U_d(r-1)$  is greater or less than  $x_{cm}$  and  $x_{cm} - s_b$ .

We first consider the case where  $x_c - U_d(r-1) > x_{cm}$ . In this case, the mutator interference threshold is reached before the nose, so that

$$p_{\text{fix}}(r) \approx \int_{x_{cm}-s_b}^{x_{cm}} \frac{e^{-\frac{(\tilde{x}+U_d(r-1))^2}{2v}}}{\sqrt{2\pi v}} \cdot 2x_{cm}e^{\frac{\tilde{x}^2-x_{cm}^2}{2v}} d\tilde{x} + \int_{x_{cm}+U_d(r-1)}^{x_c} \frac{e^{-\frac{\tilde{x}^2}{2v}}}{\sqrt{2\pi v}} \cdot 2[\tilde{x} - U_d(r-1)] d\tilde{x}. \quad (\text{B2})$$

After multiplying both sides by  $N = \frac{\sqrt{2\pi v}}{2x_c s_b} e^{\frac{x_c^2}{2v}}$ , we obtain

$$\begin{aligned} Np_{\text{fix}}(r) &\approx e^{\frac{x_c^2-x_{cm}^2}{2v} - \frac{(U_d(r-1))^2}{2v} - \frac{x_{cm}U_d(r-1)}{v}} \frac{x_{cm}}{x_c s_b} \int_0^{s_b} e^{\frac{U_d(r-1)z}{v}} dz + \frac{e^{\frac{x_c^2}{2v}}}{x_c s_b} \int_{x_{cm}+U_d(r-1)}^{x_c} ze^{-\frac{z^2}{2v}} dz \\ &\quad - 2NU_d(r-1) \int_{x_{cm}+U_d(r-1)}^{x_c} \frac{e^{-\frac{z^2}{2v}}}{\sqrt{2\pi v}} dz. \end{aligned} \quad (\text{B3})$$

Evaluating the remaining integrals and substituting  $x_{cm} = x_c - \frac{v}{s_b} \log(r)$ , we find that

$$\begin{aligned} Np_{\text{fix}}(r) &\approx e^{\frac{x_c}{s_b} \left[ \log(r) - \frac{U_d s_b (r-1)}{v} \right] - \frac{v}{2s_b^2} \left[ \log(r) - \frac{U_d s_b (r-1)}{v} \right]^2} \left[ \frac{e^{\frac{U_d s_b (r-1)}{v}} - 1}{\frac{U_d s_b (r-1)}{v}} \right] \\ &\quad + \frac{v}{x_c s_b} \left[ e^{\frac{x_c}{s_b} \left[ \log(r) - \frac{U_d s_b (r-1)}{v} \right] - \frac{v}{2s_b^2} \left[ \log(r) - \frac{U_d s_b (r-1)}{v} \right]^2} - 1 \right] \\ &\quad - 2NU_d(r-1) \cdot \frac{\text{Erfc} \left( x_c - \frac{v}{s_b} \log(r) + U_d(r-1)/\sqrt{2v} \right) - \text{Erfc} \left( x_c/\sqrt{2v} \right)}{2}, \end{aligned} \quad (\text{B4})$$

where  $\text{Erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-t^2} dt$ . The last two terms are only relevant for extremely high deleterious mutation rates where  $U_d(1-r) \sim x_c$ . After neglecting these terms, we obtain Eq. (11) in the main text.

Now we consider the case where  $x_c - U_d(r-1) < x_{cm}$ . In this case, the edge of the fitness distribution is reached before the mutator interference threshold. If  $x_c - U_d(r-1) < x_{cm} - s_b$ , then the edge of the fitness distribution never even makes it to the region where  $\tilde{w}_m(\tilde{x})$  is positive, and  $p_{\text{fix}}(r) \approx 0$ . On the other hand, if  $x_c - U_d(r-1)$  lies between  $x_{cm}$  and  $x_{cm} - s_b$ , then

$$p_{\text{fix}}(r) = \int_{x_{cm}-s_b}^{x_c-U_d(r-1)} \frac{e^{-\frac{(\tilde{x}+U_d(r-1))^2}{2v}}}{\sqrt{2\pi v}} \cdot 2x_{cm}e^{\frac{\tilde{x}^2-x_{cm}^2}{2v}} d\tilde{x}. \quad (\text{B5})$$

Multiplying by  $N = \frac{\sqrt{2\pi v}}{2x_c s_b} e^{\frac{x_c^2}{2v}}$  and evaluating the integrals, we obtain

$$Np_{\text{fix}}(r) = e^{\frac{x_c}{s_b} \left[ \log(r) - \frac{U_d s_b (r-1)}{v} \right] + \frac{U_d s_b (r-1)}{v} - \frac{v}{2s_b^2} \left[ \log(r) - \frac{U_d s_b (r-1)}{v} \right]^2} \left[ \frac{1 - e^{-\frac{U_d(r-1)}{v} (x_c - U_d(r-1) - x_{cm} + s_b)}}{\frac{U_d s_b (r-1)}{v}} \right]. \quad (\text{B6})$$

In practice, the differences between Eqs. (B4) and (B6) do not become very pronounced until  $Np_{\text{fix}}(r)$  is already quite small. For simplicity, we will therefore use Eq. (B4) for the full range of  $r$ .

### Appendix C: Forward-time simulations

We validate several of our key approximations by comparing our predictions with the results of forward-time simulations similar to the standard Wright-Fisher model. In these simulations, the population is divided into fitness classes depending on the number of beneficial mutations ( $n_b$ ) and the number of deleterious mutations ( $n_d$ ) that each individual possesses. The simulation starts with a homogeneous population of  $N$  individuals with  $n_b = 0$  and  $n_d = 0$ . To form the next generation, each individual produces a Poisson number of identical offspring with mean  $C_t(1 + s_b n_b - s_d n_d)(1 - U_d - U_b)$ , a Poisson number of beneficial offspring ( $n_b \rightarrow n_b + 1$ ) with mean  $C_t(1 + s_b n_b - s_d n_d)U_b$ , and a Poisson number of deleterious offspring ( $n_d \rightarrow n_d + 1$ ) with mean  $C_t(1 + s_b n_b - s_d n_d)U_d$ , where  $C_t$  is a constant recalculated at each generation to ensure that the expected number of offspring for the entire population is  $N$ . Starting from the initial population, all simulations are allowed to “burn-in” for  $\Delta t = 5 \times 10^4$  generations before any measurements are made.

To measure the fixation probability of a modifier allele, we modify this basic algorithm so that new modifier offspring are produced from wildtype individuals at rate  $\mu$ . These modifier lineages reproduce the same way as wildtype individuals except with  $U_b \rightarrow rU_b$  and  $U_d \rightarrow rU_d$ . Further modifications to the mutation rate or reversion to the wildtype are not allowed. To minimize the effects of initial conditions,  $\mu$  is artificially fixed at zero until the burn-in period has elapsed. We then record the number of generations  $T$  between the end of the burn-in period and the fixation of the modifier phenotype. In the limit that  $\mu \rightarrow 0$ , this is related to the fixation probability through the relation

$$Np_{\text{fix}}(r) = \frac{1}{\mu T}. \quad (\text{C1})$$

To ensure that  $\mu$  is chosen to be small enough for Eq. (C1) to apply, we repeat this process  $M = 60$  times with a sequence of modifier mutation rates  $\mu_1, \mu_2, \dots, \mu_M$ , and a sequence of fixation times  $T_1, T_2, \dots, T_M$ . The mutation rate at step  $i$  is chosen based on the previous  $T_{i-1}$ , so that the predicted value of  $T_i$  is  $10^4$  generations:

$$\mu_i = \min \left\{ \frac{\mu_{i-1} T_{i-1}}{10^4}, 10^{-2} \right\}. \quad (\text{C2})$$

The value  $10^4$  is chosen because, for the parameters considered here,  $10^4 \gg \frac{1}{s_b} \log(s_b/U_b)$ , and the mutation rates are capped at  $10^{-2}$  to minimize the correlated effects of Muller’s ratchet (Neher and Shraiman, 2012). The sequence is started with  $\mu_1 = 10^{-4}$ , and is allowed to “burn-in” for 10 iterations before  $T_i$ ’s are recorded. The fixation probability is calculated from the maximum likelihood estimator,

$$\widehat{Np_{\text{fix}}}(r) = \frac{1}{\sum_{i=10}^{60} \mu_i T_i}. \quad (\text{C3})$$

A copy of our implementation in Python is available at [https://github.com/benjaminhgood/mutator\\_simulations](https://github.com/benjaminhgood/mutator_simulations).