

# The Evolution of Mutator Genes in Bacterial Populations: The Roles of Environmental Change and Timing

Mark M. Tanaka,<sup>\*,1</sup> Carl T. Bergstrom<sup>†</sup> and Bruce R. Levin<sup>\*</sup>

<sup>\*</sup>Department of Biology, Emory University, Atlanta, Georgia 30322 and <sup>†</sup>Department of Zoology, University of Washington, Seattle, Washington 98195

Manuscript received December 9, 2002

Accepted for publication March 7, 2003

## ABSTRACT

Recent studies have found high frequencies of bacteria with increased genomic rates of mutation in both clinical and laboratory populations. These observations may seem surprising in light of earlier experimental and theoretical studies. Mutator genes (genes that elevate the genomic mutation rate) are likely to induce deleterious mutations and thus suffer an indirect selective disadvantage; at the same time, bacteria carrying them can increase in frequency only by generating beneficial mutations at other loci. When clones carrying mutator genes are rare, however, these beneficial mutations are far more likely to arise in members of the much larger nonmutator population. How then can mutators become prevalent? To address this question, we develop a model of the population dynamics of bacteria confronted with ever-changing environments. Using analytical and simulation procedures, we explore the process by which initially rare mutator alleles can rise in frequency. We demonstrate that subsequent to a shift in environmental conditions, there will be relatively long periods of time during which the mutator subpopulation can produce a beneficial mutation before the ancestral subpopulations are eliminated. If the beneficial mutation arises early enough, the overall frequency of mutators will climb to a point higher than when the process began. The probability of producing a subsequent beneficial mutation will then also increase. In this manner, mutators can increase in frequency over successive selective sweeps. We discuss the implications and predictions of these theoretical results in relation to antibiotic resistance and the evolution of mutation rates.

**I**N the face of fluctuating environments, organisms that reproduce asexually have a reduced capacity to generate variation compared to sexual organisms. Bacteria can occasionally undergo genetic exchange due to the wide variety of accessory genetic elements available to them: integrons and gene cassettes, transposons, plasmids, and phages. These elements allow bacteria to exchange genes not only within the narrow bounds of a given species, but also across far wider phylogenetic divides than those typically bridged by sexual eukaryotes (LEVIN and BERGSTROM 2000; OCHMAN *et al.* 2000). As the rapid spread of antibiotic resistance by horizontal transfer so strikingly and frighteningly illustrates, bacteria can and often will acquire beneficial genes from external sources, so long as these genes exist *somewhere* in the prokaryotic world (LEVIN and BERGSTROM 2000). We have suggested elsewhere that spatially and temporally varying environments may be the driving forces behind the evolution and maintenance of many mobile genetic elements (BERGSTROM *et al.* 2000).

What happens when bacteria are faced with novel or fluctuating environmental conditions and no external

sources of beneficial genes are available? In such circumstances, evolution must resort to a second source of variation: *de novo* generation by mutation. But elevated mutation rates come at a cost: they create deleterious changes in the genome. In fact, the classical theory of mutation rate evolution concluded that in sufficiently stable environments, mutation rates evolve to be quite low (KIMURA 1960, 1967; LEIGH 1970, 1973; ESHEL 1973a,b; PAINTER 1975; GILLESPIE 1981; LIBERMAN and FELDMAN 1986; ALTENBERG and FELDMAN 1987). In general, we do indeed observe low mutation rates and elaborate mechanisms of DNA repair, which presumably evolved to keep mutation rates low (DRAKE *et al.* 1998; MILLER 1998; SNIEGOWSKI *et al.* 2000). And, until recently, all was in place: intuition, formal theory, and experimental observations all supported the proposition that although higher rates of mutation can appear occasionally, mutation rates in bacterial populations would remain low and mutator alleles (alleles elevating the genomic mutation rate) would not be selected.

In recent years, however, researchers have begun to uncover high frequencies of mutator genes in natural or clinical (LECLERC *et al.* 1996; MATIC *et al.* 1997; BUCCI *et al.* 1999; OLIVER *et al.* 2000; BJÖRKHOLM *et al.* 2001) and experimental (MAO *et al.* 1997; SNIEGOWSKI *et al.* 1997; FUNCHAIN *et al.* 2000; GIRAUD *et al.* 2001) bacterial populations. What is happening in these populations? A broad but not mechanistic answer from the classical

**We dedicate this article to the memory of Johanna Björkman, an extraordinary scientist and human being.**

<sup>1</sup>Corresponding author: School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney NSW 2052, Australia. E-mail: m.tanaka@unsw.edu.au

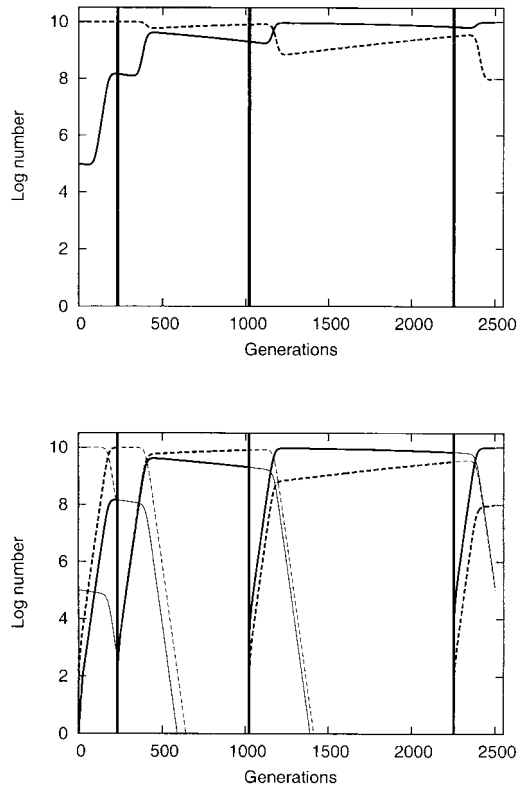


FIGURE 1.—Timing of events during multiple sweeps: simulation with parameters from Table 1. Top, frequencies of mutators (solid line) and nonmutators (dashed line) over time. Thick vertical lines indicate environmental shifts. Bottom, frequencies for the same run of the parent subpopulations from which mutations arise. Thin lines, “unevolved” lineages; thick lines, evolved lineages. Again, solid lines are for mutators and dashed lines for nonmutators. The dynamics are described in *Details of the simulation* in the APPENDIX.

theory is simply that those populations probably faced environments that fluctuated rapidly enough to favor mutator genes and that such fluctuations may be very common for bacteria. These observations prompted theorists to revisit the subject of bacterial mutation rates to understand it in closer detail (TADDEI *et al.* 1997; GERRISH 1998; GERRISH and LENSKI 1998; DE VISSER *et al.* 1999; FIELD *et al.* 1999; JOHNSON 1999a,b; TENAILLON *et al.* 1999, 2000; MASEL and BERGMAN 2003). We propose an alternative to the models developed in this recent renaissance of mutator gene evolution theory and offer a resolution to the following problem.

Both experiments and theory suggest that the fate of mutator alleles will depend critically on their initial frequencies. Whatever their long-term advantages may be, mutators will be at a disadvantage when initially rare. COX and GIBSON (1974), CHAO *et al.* (1983), and CHAO and COX (1983) offer a simple and elegant explanation. For example, consider a rare mutator allele in a large bacterial population. Suppose the mutator gene increases the mutation rate substantially (say, 100-fold) but is initially present at a very low frequency (say,  $1 \times 10^{-5}$ ). In such

a population, beneficial mutations will first appear in a nonmutator individual 99.9% of the time simply because the latter are present in vastly greater numbers. Furthermore, the mutator subpopulation will take, on average, 1000 times as long as the larger nonmutator subpopulation to acquire the same beneficial mutation. Experimental studies confirm this basic intuition: in *Escherichia coli* populations exposed to novel environmental conditions, mutators are unable to invade when rare (COX and GIBSON 1974; CHAO and COX 1983; CHAO *et al.* 1983; GIRAUD *et al.* 2001).

Given this daunting accord between theory and observation, how do mutators ever reach high enough frequencies to be detected, let alone to have an appreciable effect on the process of bacterial evolution in response to environmental variation? In this study we show that one answer lies in the relative *timing* of the generation of beneficial mutations on mutator and nonmutator backgrounds. While the small mutator subpopulation is unlikely to be the first to generate a beneficial mutation, it is likely to generate this beneficial mutation before most similarly sized subsets of the nonmutator subpopulation. This *relative timing* advantage turns out to be sufficient to allow mutators to ascend in many cases. We begin with a simulation model in which the environment changes stochastically at a low rate (SIMULATION MODEL). In ANALYTICAL MODEL we present an analytical approximation to this first model, which examines the details of a single selective sweep. We then discuss the implications of this theoretical study for the problems of acquired antimicrobial resistance and how this theory can be tested experimentally.

#### SIMULATION MODEL

We begin with a simulation model of the population dynamics of mutator genes. Consider a large population of bacteria, composed of both nonmutator and mutator cells. The mutator type is rare; it starts at a frequency of  $10^{-5}$ . Although this may seem like a high frequency of mutators (amounting to  $10^5$  mutator cells in a population of  $10^{10}$ ), it is a reasonable number at mutation-selection balance (JOHNSON 2000).

Both the mutator and nonmutator strains can adapt or evolve to the current environmental conditions by generating a beneficial mutation. Prior to a shift in environmental conditions, we assume that there is a single advantage-conferring phenotype that the nonmutator and mutator subpopulations have the opportunity to acquire through mutation. Mutator cells generate the mutation at a higher rate ( $\mu_m$ ) than nonmutator cells ( $\mu_n$ ). This gives us four possible genotypes: “unevolved nonmutators,” “unevolved mutators,” “evolved nonmutators,” and “evolved mutators.” We track the number of cells of each genotype over time  $t$ , labeling these  $n(t)$ ,  $m(t)$ ,  $N(t)$ , and  $M(t)$ , respectively.

The advantage conferred by the beneficial mutation

TABLE 1

Parameter values used in calculations

$n$	$10^8$ or $10^{10}$ cells
$m_0$	$10^5$ cells
$\mu_n$	$5 \times 10^{-9}$ /cell/generation
$\mu_m$	$4 \times 10^{-7}$ /cell/generation
$b$	0.1/cell/generation
$s$	0.002/cell/generation

is expressed as a growth advantage  $b$  relative to cells without that mutation. We assume bacteria with mutator genes have an intrinsic disadvantage ( $s$ ; because of the higher rate of generating deleterious mutations) both in the ancestral state and when carrying the beneficial mutation. Thus evolved mutators have the growth advantage  $b - s$  over unevaluated nonmutators. With probability 0.002 per generation, the environment changes; each time this occurs, the evolved lineages lose their fitness advantage (and mutators retain their fitness costs). We provide the details of the dynamics in the APPENDIX, in *Details of the simulation*.

In Figure 1, we follow the changes in the nonmutator and mutator populations during the course of a simulation in which the mutators eventually dominated the population. The parameter values used for this simulation are presented in Table 1 and are based on the estimates provided in CHAO and COX (1983). In this run, the first beneficial mutation occurred in the nonmutator subpopulation and, while that mutant began to rise in frequency, the mutator subpopulation declined slowly due to its intrinsic disadvantage. Before the first environmental change, a beneficial mutant appeared in the mutator population, so that by the time of the environment shift, the frequency of the mutator was boosted higher than its initial frequency. Consequently, the probability of a beneficial mutation occurring in the mutator subpopulation increased dramatically. Environmental changes continued to occur, and each time the mutators acquired a beneficial mutation, their absolute numbers as well as their frequency increased; with each increase in size there was a corresponding increase in the probability of acquiring another beneficial mutation. Eventually, the size of the mutator subpopulation increased to the point at which, because of the higher mutation rate, this subpopulation became more likely than the nonmutator subpopulation to acquire a beneficial mutation following an environmental change.

While the mutator population eventually dominated in the simulation depicted in Figure 1, success of the mutator gene occurred in only  $\sim 20\%$  of the 1000 simulation runs performed with this parameter set and initial conditions. In 69% of the runs the subpopulation bearing the mutator gene declined to a frequency of  $10^{-10}$  before the end of the simulation at generation 2500. The mutator persisted without dominating the popula-

TABLE 2

Essential notation used in ANALYTICAL MODEL

$n$	No. of nonmutator cells
$m(t)$	No. of mutator cells at time $t$
$N(t)$	No. of nonmutators with beneficial mutation at time $t$
$M(t)$	No. of mutators with beneficial mutation at time $t$
$m_0$	No. of mutator cells at the beginning of a time cycle
$\mu_n$	Mutation rate of nonmutators per cell division
$\mu_m$	Mutation rate of mutators per cell division
$g$	Strength of mutator ( $\mu_m/\mu_n$ )
$W$	Time until appearance of beneficial mutant from nonmutator subpopulation
$X$	Time until appearance of beneficial mutant from mutator subpopulation
$b$	Growth rate of cells with beneficial mutation
$s$	Cost of being a mutator
$t^*$	Time at the end of a cycle (at completion of selective sweep)
$\bar{t}$	Average length of a cycle
$\lambda_m$	Rate of appearance of beneficial mutation in mutator
$\lambda_n$	Rate of appearance of beneficial mutation in nonmutator
$\bar{v}$	Time from appearance of an evolved mutator that breaks even until the end of cycle
$R$	Log gain made by mutator at the end of cycle

tion in the remaining 11% of runs. Of course, if the mutator-bearing subpopulation were initially larger, the frequency of successful outcomes would have been greater. Conversely, if the mutator subpopulation had been smaller, then even fewer runs would have ended with mutators present at high frequency.

Several features of the trajectory shown in Figure 1 are common to virtually all trajectories in which mutators reached high frequencies. First, the mutator need not be the first to generate the beneficial mutation. It just has to acquire that mutation before its frequency declines to the point where it is too rare to get the beneficial mutation at all. Second, the mutator subpopulation reaches high frequency not in a single selective sweep, but rather in a series of smaller steps, during the course of a number of changes in environmental states. Third, although the mutator can climb in frequency in steps, we note that it occasionally declines in frequency.

#### ANALYTICAL MODEL

To gain a more detailed understanding of the dynamics of this process, we construct a simple analytical model that focuses on the dynamics observed during a *single* environmental cycle as simulated in the previous section. While this approach will not capture everything about the full process as it extends over a series of environmental changes, it will be helpful in characterizing the probability that a mutator phenotype ascends. Our notation is summarized in Table 2 and the process is diagrammed in Figure 2. As in the simulation, we assume the population is composed of nonmutator and mutator cells. The

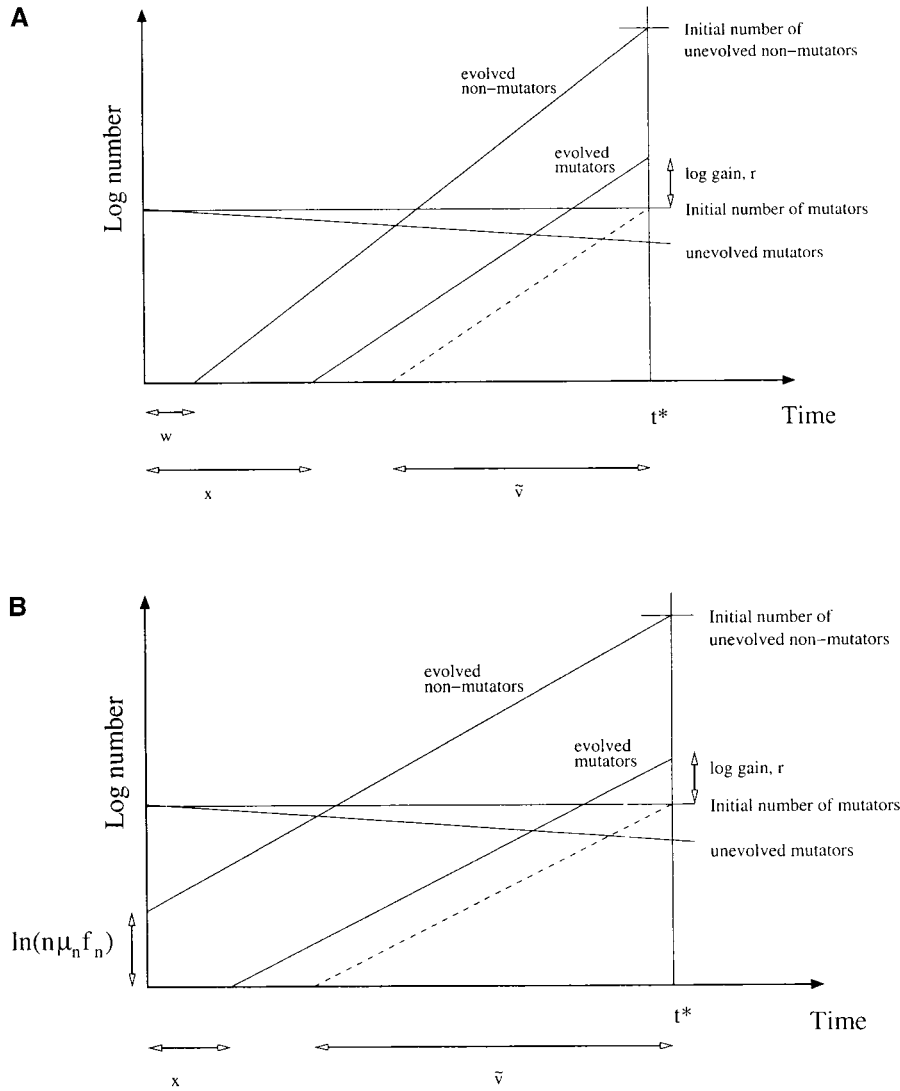


FIGURE 2.—Timing of events during a single selective sweep. A typical time cycle is shown, in which the evolved non-mutator has arisen before the evolved mutator. (A) The case where  $n\mu_n f_n < 1$ . (B) The case where  $n\mu_n f_n > 1$ . The dashed lines indicate the trajectories of mutators that would exactly recover their original density by the end of the time cycle. That is, these are mutator lineages that “break even.” See text and Table 2 for descriptions of notation and time intervals.

major differences between the simulation and this analytical model are that in the latter (i) time is continuous rather than discrete, (ii) the number of unevolved non-mutator cells  $n$  is approximated as constant, and (iii) the numbers of evolved mutators and nonmutators ( $M$  and  $N$ ) increase exponentially.

We limit our analysis to the period of time beginning with the occurrence of an environmental shift and ending once a beneficial mutant from the nonmutator pool has exponentially increased in number (at rate  $b$ ) until it reaches size  $n$ , at time  $t^*$ . We call this period a *time cycle*. The end of the time cycle here corresponds in the simulation model to the time at which the evolved nonmutator reaches an appreciable fraction (e.g., one-half) of the total population. Concomitantly, the unevolved non-mutators slowly decline in the simulation; therefore, as an approximation in this section, we let  $n$  remain constant during the time cycle. In *Definition of time cycle* (APPENDIX) we justify this assumption by showing that the number of unevolved nonmutators stays large ( $>n/2$ ) throughout the time cycle. Hence, there is no need to normalize

the dynamics to keep the population size constant, and the selective advantages enjoyed by the evolved types are the fitnesses relative to the unevolved nonmutator.

**Rate of appearance of successful mutants:** Let  $\lambda_n$  be the rate at which beneficial mutations destined to survive drift appear in the subpopulation of unevolved non-mutators. Similarly,  $\lambda_m$  is the rate at which beneficial mutations surviving drift appear among unevolved mutators. On average,

$$\lambda_n = n\mu_n f_n \tag{1}$$

and

$$\lambda_m = m_e \mu_m f_m, \tag{2}$$

where  $f_n$  is the probability of fixation of each new evolved nonmutator at the time of its appearance,  $f_m$  is the probability of fixation of each new evolved mutator, and  $m_e$  is the effective population size of mutators during a time cycle of average length. We postpone discussing our choice of  $m_e$  until after describing the dynamics of the various cell types.

We use the probability of fixation of a mutant as approximated by diffusion theory (KIMURA 1983),

$$\frac{1 - e^{-2\alpha}}{1 - e^{-2N_{ev}\alpha}}, \quad (3)$$

where  $\alpha$  is the fitness advantage of the mutant over the dominant strain in the population and  $N_{ev}$  is the variance effective population size. Because the number of unevolved nonmutators ( $n$ ) is set to be large throughout the time cycle,  $\alpha = b$  in the case of evolved nonmutators, and  $\alpha = b - s$  for evolved mutators. For bacterial cells, which replicate by binary fission, the distribution of the number of offspring is 0 or 2 with equal probability, the variance in offspring number is 1 (JOHNSON and GERRISH 2002), and  $N_{ev}$  equals the very large total number of cells. Hence, as reasonable approximations we can write

$$\begin{aligned} f_n &= 1 - e^{-2b} \\ f_m &= 1 - e^{-2(b-s)}. \end{aligned} \quad (4)$$

**Dynamics:** Let  $W$  be the time at which the first evolved nonmutator that avoids immediate loss while rare appears in the population, and let  $X$  be the time at which the first evolved mutator avoiding loss appears. We model the appearance of the beneficial mutation in the nonmutator subpopulation in two different ways depending on the magnitude of the supply of successful beneficial mutations per generation. If  $\lambda_n = n\mu_n f_n < 1$  (case 1), the waiting time for a mutation that survives drift is an exponentially distributed random variable. Alternatively, if  $\lambda_n = n\mu_n f_n > 1$  (case 2), the nonmutator subpopulation immediately generates multiple ( $n\mu_n f_n$ ) beneficial mutations that avoid loss by drift. In this case,  $W = 0$ . Evolved mutators that survive drift are assumed to proliferate at rate  $b - s$  until the time cycle ends.

*Case 1:*  $n\mu_n f_n < 1$ . We assume the waiting times until successful mutants arise are distributed as follows:  $W \sim \text{Exp}(\lambda_n)$  and  $X \sim \text{Exp}(\lambda_m)$ . The dynamics are described by

$$m(t) = m_0 e^{-st} \quad (5)$$

$$N(t) = \begin{cases} e^{b(t-W)} & \text{if } W \leq t \\ 0 & \text{if } 0 \leq t < W \end{cases} \quad (6)$$

$$M(t) = \begin{cases} e^{(b-s)(t-X)} & \text{if } X \leq t \\ 0 & \text{if } 0 \leq t < X. \end{cases} \quad (7)$$

We note that this system slightly underestimates the rate of increase of successful mutants because in the cases where the beneficial mutation rises faster than expected by chance due to drift, the allele is more likely to persist (OTTO and BARTON 1997).

*Case 2:*  $n\mu_n f_n > 1$ . Here, multiple beneficial mutations arise in the first generation and survive drift. In fact, there is an ongoing supply of successful beneficial mu-

tants after the first generation. We can approximate the dynamics of  $N$  with the equation

$$\frac{dN}{dt} = bN + n\mu_n f_n. \quad (8)$$

Using the initial condition  $N(0) = n\mu_n f_n$ , the solution to this differential equation is

$$N(t) = n\mu_n f_n (1 + 1/b)e^{bt} - n\mu_n f_n / b, \quad (9)$$

which we use to replace Equation 6 in this case. The appearance and increase of evolved mutators is modeled in the same way as in case 1.

We stress that the frequencies of the evolved cells will not always change in exponential fashion. However, if the evolved mutator arises after this exponential period, this will inevitably lead to the dramatic loss of mutator numbers. In this case we can thus assume that  $M$  is zero at the end of the cycle. Alternatively, if the evolved mutator arises *before* the evolved nonmutator, the approximations may not be an adequate description of the subsequent dynamics, but this is not a concern because at that point, we already know what we wanted to find out—that the mutator has ascended.

We can now return to the effective population size of unevolved mutators  $m_e$ , which we model by using the harmonic mean of  $m(t)$  [with  $m(t)$  as given by Equation 5] over the time cycle of average length ( $\bar{t}$ ):

$$m_e = \frac{m_0 s \bar{t}}{e^{s\bar{t}} - 1}. \quad (10)$$

We use  $m_e$  instead of the initial number of mutators  $m_0$  because  $m(t)$  declines over time, along with the supply rate of beneficial mutations among mutator cells. It is a conservative assumption in that it does not favor mutators.

The average length of a time cycle is the average time it takes for the successful beneficial mutation to arise [ $E(W) = 1/\lambda_n$ ] plus the time it takes for that mutation to reach size  $n$ . That is,

$$\bar{t} = \begin{cases} \frac{1}{n\mu_n f_n} + \frac{\ln(n)}{b} & \text{if } n\mu_n f_n < 1 \\ \frac{1}{b} [\ln(1 + \frac{b}{\mu_n f_n}) - \ln(1 + b)] & \text{if } n\mu_n f_n > 1. \end{cases} \quad (11)$$

**Log change in number of mutators:** We now turn to the statistical properties of the gain made by the mutator subpopulation during the course of a given selective sweep. Define  $R = \ln(m(t^*)) - \ln(m_0)$ ; this is the change in frequency of the mutator over the course of a single time cycle, as measured on a logarithmic scale. If for a given cycle  $R > 0$ , the mutator has increased in frequency. Using the waiting time distributions of  $W$  and  $X$ , we can derive the probability density function of  $R$  (see the APPENDIX: *The distribution of R*).

The expected log change in mutator number is

$$E(R) = (b - s)[\bar{t} - (1/\lambda_m + \bar{v})], \quad (12)$$

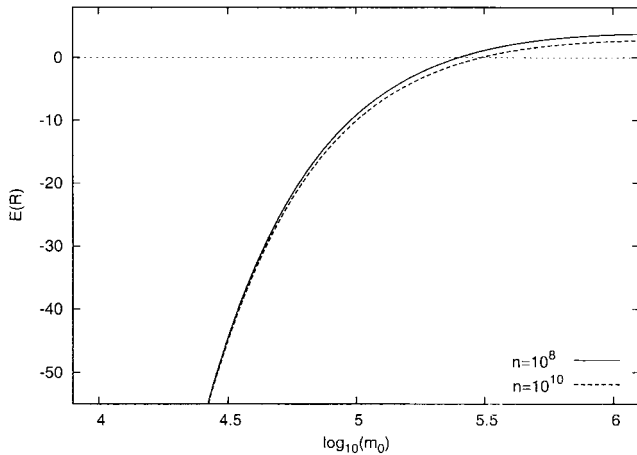


FIGURE 3.— The expected log increase  $E(R)$  of the mutator over the course of a selective sweep, as a function of the initial size of the mutator subpopulation  $m_0$ . The function is given by Equation 12. The parameter values used for this function are given in Table 1.

where  $\bar{t}$  is the average length of the time cycle, as given above, and where  $\bar{v}$  is the time from the appearance of the evolved mutator until the end of a cycle such that the mutator strain would exactly recover its original population size. From (7) and Figure 2 we can write  $\bar{v} = \ln(m_0)/(b - s)$ . The quantity  $(1/\lambda_m + \bar{v})$  is the average time for an evolved mutator to appear plus the amount of time necessary for the mutator to break even. For the parameter values considered (Table 1), this quantity is  $\gg \bar{t}$  because on average the mutator generates the beneficial mutation much later than would be necessary for the mutator to break even. Hence, with  $b - s > 0$  by assumption,  $E(R)$  will almost always be negative. Should the average time for the evolved mutator to appear exactly equal the time required for an evolved mutator to break even, then we would have  $E(R) = 0$ . When mutators are rare, if we consider only the mean change in mutator population size, mutators face a dire fate. Figure 3 shows  $E(R)$  for the parameter values in Table 1 as a function of the initial mutator subpopulation size  $m_0$ .

**Probabilities of increase and jackpot:** While the average change in frequency,  $E(R)$ , suggests a general failure of the mutator to establish itself, if we examine the *probability* of success, a different story emerges. To illustrate this, we consider two probabilities. Let  $P_i = \Pr(R > 0)$  be the probability that the mutator has increased in number after the selective sweep, and let  $P_j = \Pr(R > r_j)$  be the probability that the mutator has increased to at least the “jackpot size” by the end of the selective sweep. We define the jackpot size,  $J$ , as the population size of mutators such that each of the two subpopulations, mutator and nonmutator, has an equal chance of producing the next beneficial mutation. In other words, the jackpot size is  $m$  such that  $\lambda_m = \lambda_n$ ; so  $J = (n\mu_n f_n (e^{\bar{v}} - 1)) / (\bar{v} \mu_m f_m)$  and  $r_j = \ln(\lambda_n/\lambda_m)$ .

These two probabilities are easily computed from the cumulative density function (CDF) of  $R$ , given in *The distribution of  $R$*  in the APPENDIX. Generally,

$$\Pr(R > r) = 1 - \Pr(X > W) e^{-\lambda_n(\bar{v} - \bar{v} - r/(b-s))}, \quad (13)$$

where in case 1 ( $n\mu_n f_n < 1$ )  $\Pr(X > W) = \lambda_n/(\lambda_m + \lambda_n)$  and  $\bar{v} = \ln(n)/b$ , and in case 2 ( $n\mu_n f_n > 1$ )  $\Pr(X > W) = 1$  and  $\bar{v}$  is as given in the second part of Equation 11. Note that  $\bar{v}$  is the duration of time from the appearance of the evolved nonmutator until the end of a time cycle.

We next explore how  $P_i$  and  $P_j$  are affected by key parameters. Figure 4 shows these two quantities as functions of the initial number of mutators  $m_0$ , the fitness advantage of the beneficial mutation  $b$ , and the mutator strength  $g = \mu_m/\mu_n$ . We used two different values for  $n$ , requiring the two alternative cases: when  $n = 10^8$ ,  $n\mu_n f_n < 1$  (Figure 4, left column), and when  $n = 10^{10}$ ,  $n\mu_n f_n > 1$  (Figure 4, right column). To check the analytical model against the simulation model, we ran the simulation multiple times over a single time cycle and kept track of the fraction of times the mutator succeeded. The simulation was run for many different parameter values as shown in the figure.

In natural populations, the frequency of mutators can be very low if selective sweeps in nonmutators periodically eliminate mutators. On the other hand, if mutation rates toward mutators are high, the mutator frequency can be quite high: as high as 2% at “mutation-indirect selection balance” (JOHNSON 2000). These considerations, along with uncertainties inherent in measuring these mutation rates, suggest that a broad range of mutator frequencies should be considered: we use frequencies of  $\sim 10^{-6}$  to  $\sim 10^{-2}$ .

A comparison between Figure 3 and the top two graphs of Figure 4 shows that while the *expected* log gain of the mutator strain can be extremely low when its initial numbers are low, the *probability* of making a gain is non-negligible for exactly the same parameter values. Indeed, when  $n = 10^8$ , for a large proportion of the time the mutator increases, it also hits the jackpot. Although each particular mutator cell is overwhelmingly likely to be lost, the probability that at least one mutator cell in that subpopulation will succeed is substantial. We note the distinction between rarity in numbers and rarity in frequency. In our model, mutators are more likely to produce beneficial mutations when they are in large numbers, even when their frequency is low in a large total population. However, mutators also benefit from being relatively common because nonmutators are then less able to drive them out through selective sweeps.

Curiously, the success of mutators is not greatly affected by the fitness advantage  $b$  conferred by the mutations that they facilitate (see middle two graphs of Figure 4). This relative insensitivity results from the balance of two countervailing factors. On the one hand a rare

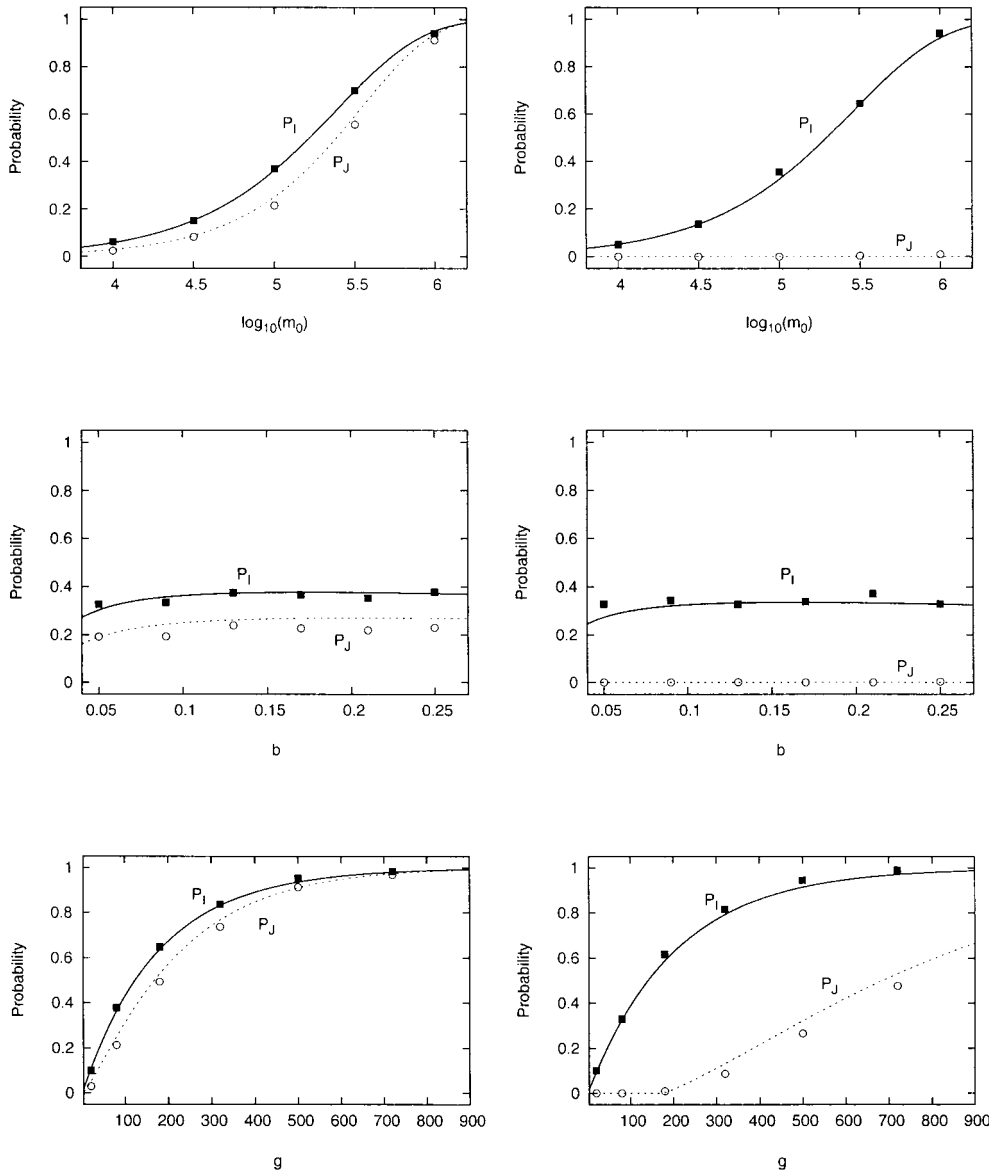


FIGURE 4.—Comparison between simulation and analytical approximation. Left column,  $n = 10^8$  ( $n\mu_n f_n < 1$ ); right column,  $n = 10^{10}$  ( $n\mu_n f_n > 1$ ). The vertical axis represents the probability that by the end of a time cycle the mutator strain has increased in frequency ( $P_I$ ) or hit the jackpot or better ( $P_J$ ). Each point represents 1000 simulations. The variable  $g$  is the mutator strength, *i.e.*,  $\mu_m/\mu_n$ , which is varied while  $\mu_n$  is kept constant. Apart from the varied parameters, the parameter values are given in Table 1.

mutator is better able to survive genetic drift when  $b$  is high, *i.e.*, when beneficial mutations have a large effect. On the other hand, the larger the effect of these beneficial mutations, the faster the evolved nonmutator is able to sweep through the population—and thus the shorter the time window during which the mutator has the opportunity to generate its own beneficial mutation.

In our model, the strength of the mutator  $g = \mu_m/\mu_n$  has the effect of monotonically increasing mutator success. We note, however, that this effect is an artifact of the way our model is set up. In reality, increasing  $g$  would also increase the deleterious effects of the mutator (*i.e.*,  $s$  would also increase), so that we would expect something like a bell-shaped function, as discussed by ORR (2000). An extension of this model should take this phenomenon into account by modeling the deleterious effects explicitly.

In this section we have analyzed only the dynamics

that occur during the period where the beneficial mutation has yet to reach fixation in the population. By doing so, we have been able to estimate the probabilities that mutators make a net gain following an environmental shift. This allows us to explain the puzzles with which we started: how the mutators often manage to increase in frequency despite the dual challenges that they are not the first to acquire the beneficial mutation, and that in the expectation, they will undergo a large decline in frequency. We have not treated the dynamics that occur subsequent to fixation of the beneficial allele (on whatever combination of mutator or nonmutator background), as this phase of the dynamics is not central to explaining these puzzles. Nonetheless, the reader may note that if the cost of accelerated mutation rate  $s$  is sufficiently small and the rate of environmental change is not too slow, genotype frequencies will not change appreciably once the beneficial mutation has swept through the

population on any combination of mutator and nonmutator background. In this case, the dynamics simulated in SIMULATION MODEL will be closely approximated by the dynamics treated here, followed by a period of no change in genotype frequencies that extends until the next environmental shift.

## DISCUSSION

**Mutators can succeed:** We have demonstrated that in continually changing environments where the adaptation to any given environmental state is independent of the adaptation to other states, mutator genes can climb to prominence by associated selection (*i.e.*, by hitchhiking) with the linked beneficial mutations they generate. We have shown that even when mutators are extremely rare, there is a substantial probability that they will increase in frequency. We see this nonequilibrium analysis as a reminder that the average behavior of a process is sometimes an inadequate descriptor of the system as a whole. Although mutators are often lost in our model, in reality there would be recurrent mutation to the mutator state, and even if they are occasionally lost, organisms carrying a mutator gene would eventually dominate. “Eventually,” however, could be a very long time so that we should be unsurprised to sometimes observe low mutator frequencies even in systems that do conform to the assumptions of our model.

While our model answers the questions we raised, we recognize its limitations. First, we have not considered a number of phenomena that would also influence the dynamics of this process. Most prominent of these are multiple mutations in a given environment (TADDEI *et al.* 1997; TENAILLON *et al.* 1999), clonal interference (GERRISH and LENSKI 1998), and genetic exchange (JOHNSON 1999b; TENAILLON *et al.* 2000). We discuss some of these issues in more detail below. We expect that our qualitative conclusions about the role of timing of mutations in the success of mutator genes will hold with more general variants of this model.

Second, it is not clear how commonly environmental changes and fitness relationships of the type considered in this model occur in natural populations of bacteria. How often does a mutation augment fitness in one environment while remaining selectively neutral or deleterious in another? Whatever the natural prevalence of such scenarios, we suggest that the all-too-common human intervention of antibiotic treatment and prophylaxis approximates the scenario that we have modeled. The necessary conditions are as follows: (1) the intensity of antibiotic-mediated selection is less than absolute (resistant bacteria have an advantage but sensitive bacteria can also grow); (2) resistance can arise by single mutations; (3) different antibiotics are used sequentially; and (4) resistance to one drug is independent of resistance to other drugs in the sequence. Other environmental conditions may meet these criteria as well. How robust

our results are to deviations from the scenario considered in this particular incarnation of the model is a question that can and should be addressed theoretically. As we consider later, the properties of this model lead to predictions that can be tested empirically.

Third, while we have modeled environmental changes, we have not explored the effect of the rate of these changes. Classical theory (KIMURA 1967; LEVINS 1967) has suggested that in the long run this rate should affect the success of mutators. When the rate is too low, the opportunities for creating innovations vanish, and low mutation rates are favored. When the rate of environmental fluctuation is too high, rare mutators will never have the chance to increase in frequency. According to these models, increased mutation rates are favored when the rate of fluctuation is appropriately matched to the dynamics of fixation. In our model, the times between environmental shifts are long enough to allow roughly one fixation event. The consequences of this assumption should be examined further in a more general model. For further discussion, see TRAVIS and TRAVIS (2002). Again, we do not know to what extent these conditions occur in nature.

**Comparisons with other models:** In this study we have focused on how mutator genes rise in frequency when they are initially rare. An alternative explanation for this has been provided by TADDEI *et al.* (1997) and TENAILLON *et al.* (1999). In their model, mutators ascend in frequency because they increase the probability of producing adaptive characters that require multiple mutations. The way this process works can be seen with a simple thought experiment in which two mutations are acquired in rapid succession. Suppose that, in a population of  $10^{10}$  bacteria, the frequency of mutators is  $10^{-5}$  and the more-fit phenotype requires two independent mutations, each occurring at rate  $10^{-8}$  in the nonmutator subpopulation and  $10^{-5}$  in the mutator subpopulation. Assuming that neither mutation is present in the initial population (an assumption that may be unlikely, depending on the nature of the mutation in question), we see that—despite the  $10^5$ -fold difference in their population sizes—the mutator subpopulation will generate the two adaptive mutations at  $\sim 10$  times the rate of the nonmutator subpopulation:  $(10^5 \times 10^{-5} \times 10^{-5}) / (10^{10} \times 10^{-8} \times 10^{-8}) = 10$ . Here, we emphasize that mutators are able to succeed even in cases where only a single genetic change is required for the adaptive phenotype. The mutator subpopulation does not have to be the *first* to acquire the adaptive phenotype, so long as it acquires this phenotype substantially sooner than the nonmutator subpopulation *relative to its initial population size*.

Although their study was experimental rather than theoretical, MAO *et al.* (1997) also present an alternative model for the ascent of mutator genes when they are initially rare. How that model works can also be seen with a numerical thought experiment. Assume complete



selection for resistance to an antibiotic (no sensitive cells survive) and mutation rates of  $10^{-8}$  and  $10^{-5}$  for nonmutators and mutators, respectively. If a population of  $10^{10}$  cells are plated on agar containing an antibiotic, there would be  $\sim 100$  resistant nonmutator colonies ( $10^{10} \times 10^{-8}$ ) and 1 mutator colony ( $10^5 \times 10^{-5}$ ). Thus, in one passage, if all colonies were of the same size, the frequency of mutator cells would have gone from  $10^{-5}$  to  $10^{-2}$ . In a second passage with another antibiotic with the same total population size and mutation rates for resistance, the community would be dominated by bacteria carrying mutator genes. That is, there would once again be  $\sim 100$  colonies of resistant nonmutators ( $10^{10} \times 10^{-8}$ ) and 1000 ( $10^8 \times 10^{-5}$ ) colonies of resistant mutators. The model considered here is similar (and was in fact inspired by their experiments) except that in the present model, selection for resistance is not absolute.

In their study of bacterial population genetics, GERRISH and LENSKI (1998) examined, among other things, the substitution rate of beneficial mutations in an asexual population. Their model takes into account multiple beneficial mutants that compete with each other and hinder the fixation process—an effect they call clonal interference. Of particular relevance to this investigation is their finding that while the substitution rate increases with mutation rate, this increase slows down due to clonal interference. As a result, mutation rates are not expected to increase indefinitely. A law of diminishing returns holds. The Gerrish and Lenski analysis also considers a distribution of values for the selective advantage of mutants, which we do not address here.

**Frequency-dependence and dependence on initial conditions:** Our theoretical model is very similar to the experimental models of CHAO and COX (1983) and CHAO *et al.* (1983), which consider single adaptive changes. In both systems, the success of mutators depends on the initial frequency of mutators in the population. Moreover, in the data of Chao *et al.* there appears to be a threshold frequency (sometimes called the *Edge of Chaos*'s), above which fixation of mutators will take place, and below which mutators will be lost. This phenomenon should not be confused with frequency-dependent selection in the classical sense, as defined, for example, by LEWONTIN (1958) or LEVIN (1988). Under disruptive frequency-dependent selection, a frequency exists below which fixation is impossible, and above which fixation is certain.

In the case of mutator dynamics, no systematic force produces a threshold of this kind. The apparent threshold observed in the earlier empirical studies can be explained by models such as ours, in which the probability of mutator success as a function of the *logarithm* of initial mutator frequency is sigmoidal. When considering a very wide range of initial frequencies the transition between failure and success may seem sharp when in fact it is not. Another way of putting it is that if we consider the probability of increase *per cell*, then this proba-

bility does not change much over frequency. In fact, it will decrease slightly with increasing mutator frequency, because increasing the mutator frequency also increases the number of able competitors.

**Predictions:** At this juncture we cannot say very much about the generality of this model for the evolution of mutator genes or the prevalence of the necessary conditions in natural populations of bacteria. We can say, however, that this model has the considerable virtue of making specific predictions that can be tested empirically. One way to test the quantitative as well as the qualitative predictions derived from the analysis of this model is with experimental populations of bacteria containing predetermined numbers of mutator and nonmutator cells already well adapted to chemostat or serial transfer culture. Using antibiotics or other bacteriocidal or bacteriostatic chemicals at concentrations below the minimum inhibitory concentrations for a population of mutators and nonmutators, the sensitive bacterial population would be able to grow, and selection would favor resistant mutants. By manipulating the concentration of the antibiotics, the intensity of this selection could be varied. By transferring washed cells into fresh medium with different antibiotics or other chemicals that are detrimental to bacteria, it would be possible to mimic the conditions specified in the model for the ascent of mutator genes. With independent estimates of the mutation rates to resistance to the chemicals and the intensity of selection, one should be able to use the model to predict (1) the probability that the mutator will ascend and (2) the dynamics of its ascent. We could follow the latter using mutator strains with selectable markers.

Another prediction of this model (and other similar models) could be evaluated retrospectively as well as experimentally. If antibiotic-mediated selection of the sort considered in this model is the reason for the ascent of the mutator strain, then there should be a positive association between antibiotic resistance and mutation rates. In fact, the mutator strain is more likely than the nonmutator strain to be multiply resistant. The findings of OLIVER *et al.* (2000) with regard to *Pseudomonas aeruginosa* in cystic fibrosis patients are suggestive of this effect. Other models (TADDEI *et al.* 1997; TENAILLON *et al.* 1999) are also consistent with this observation.

**Why mutation rates are low:** It is possible that a broad array of environmental conditions favoring the evolution of higher mutation rates exists in nature. And, as seen above, it is relatively easy to produce models to account for the evolution of genes that increase the overall mutation rate. The next question that must be addressed is why mutation rates in bacterial populations are as low as empirical studies have found them to be. We see four general hypotheses for the relative rarity of mutator genes in natural populations of bacteria:

1. While their form may vary, almost all of the models for the evolution of increased mutation rates require

continuous adaptation to novel or varying environmental conditions. It may well be that the natural habitats that bacteria face are relatively stable. Novel conditions and the environmental changes may be relatively rare. Defective mismatch repair and other mutator genes do not ascend because of their intrinsic disadvantage.

2. As a consequence of intense bottlenecks, bacteria with high rates of mutation randomly fix deleterious alleles and decline in frequency. For clever experimental demonstrations of this effect in bacteria, see FUNCHAIN *et al.* (2000) and ANDERSSON and HUGHES (1996).
3. For selection to favor mutator genes, it is essential that the mutator and the beneficial mutations it generates maintain their linkage relationship. Consequently, if recombination occurs frequently enough, it would break down these essential linkage relationships and the frequency of the mutator gene will wane (LEIGH 1970; TENAILLON *et al.* 2000).
4. Bacteria switch from habitat to habitat. While populations of bacteria carrying mutator genes may adapt more rapidly to a particular habitat than populations with lower rates of mutation, that evolution may force these mutator populations into a rapid overspecialization in one habitat and a corresponding disadvantage in another (GIRAUD *et al.* 2001).

We thank Sally Otto, Olivier Tenaillon, Toby Johnson, and a reviewer for their astute comments on this work. We also thank Lin Chao, Jeffrey Miller, Paul Sniegowski, Lauren Ancel, Phil Gerrish, Michael Lachmann, Roland Regoes, and François Taddei for helpful discussions and Marc Lipsitch for providing humor with an edge. This work was supported by National Institutes of Health grants AI40662 and GM33682 to B.R.L. and a University of NSW Faculty Research Grant to M.M.T.

#### LITERATURE CITED

- ALTENBERG, L., and M. W. FELDMAN, 1987 Selection, generalized transmission and the evolution of modifier genes. I. The reduction principle. *Genetics* **117**: 559–572.
- ANDERSSON, D. I., and D. HUGHES, 1996 Muller's ratchet decreases fitness of a DNA-based microbe. *Proc. Natl. Acad. Sci. USA* **93**: 906–907.
- BERGSTROM, C. T., M. LIPSITCH and B. R. LEVIN, 2000 Natural selection, infectious transfer and the existence conditions for bacterial plasmids. *Genetics* **155**: 1505–1519.
- BJÖRKHOLM, B., M. SJOLUND, P. G. FALK, O. G. BERG, L. ENGSTRAND *et al.*, 2001 Mutation frequency and biological cost of antibiotic resistance in *Helicobacter pylori*. *Proc. Natl. Acad. Sci. USA* **98**: 14607–14612.
- BUCCI, C., A. LAVITOLA, P. SALVATORE, L. DEL GIUDICE, D. R. MARSARDO *et al.*, 1999 Hypermutation in pathogenic bacteria: frequent phase variation in meningococci is a phenotypic trait of a specialized mutator biotype. *Mol. Cell* **3**: 435–445.
- CHAO, L., and E. C. COX, 1983 Competition between high and low mutating strains of *Escherichia coli*. *Evolution* **37**: 125–134.
- CHAO, L., C. VARGAS, B. B. SPEAR and E. C. COX, 1983 Transposable elements as mutator genes in evolution. *Nature* **303**: 633–635.
- COX, E. C., and T. C. GIBSON, 1974 Selection for high mutation rates in chemostats. *Genetics* **77**: 169–184.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper & Row, New York.
- DE VISSER, J. A. G. M., C. W. ZEYL, P. J. GERRISH, J. L. BLANCHARD and R. E. LENSKI, 1999 Diminishing returns from mutation supply rate in asexual populations. *Science* **283**: 404–406.
- DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH and J. F. CROW, 1998 Rates of spontaneous mutation. *Genetics* **148**: 1667–1686.
- ESHEL, I., 1973a Clone-selection and optimal rates of mutation. *J. Appl. Prob.* **10**: 728–738.
- ESHEL, I., 1973b Clone selection and the evolution of modifying features. *Theor. Popul. Biol.* **4**: 196–208.
- FIELD, D., M. O. MAGNASCO, E. R. MOXON, D. METZGAR, M. M. TANAKA *et al.*, 1999 Contingency loci, mutator alleles, and their interactions: synergistic strategies for microbial evolution and adaptation in pathogenesis. *Ann. NY Acad. Sci.* **870**: 378–382.
- FUNCHAIN, P., A. YEUNG, J. L. STEWART, R. LIN, M. M. SLUPSKA *et al.*, 2000 The consequences of growth of a mutator strain of *Escherichia coli* as measured by loss of function among multiple gene targets and loss of fitness. *Genetics* **154**: 959–970.
- GERRISH, P. J., 1998 Dynamics of mutation and selection in asexual populations. Ph.D. Thesis, Michigan State University, East Lansing, MI.
- GERRISH, P. J., and R. E. LENSKI, 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **103**: 127–144.
- GILLESPIE, J. H., 1981 Evolution of the mutation-rate at a heterotic locus. *Proc. Natl. Acad. Sci. USA* **78**: 2452–2454.
- GIRAUD, A., I. MATIC, O. TENAILLON, A. CLARA, M. RADMAN *et al.*, 2001 Costs and benefits of high mutation rates: adaptive evolution of bacteria in the mouse gut. *Science* **291**: 2606–2608.
- 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.
- JOHNSON, T., 2000 Theoretical studies of the interaction between deleterious and beneficial mutations. Ph.D. Thesis, University of Edinburgh, Edinburgh.
- JOHNSON, T., and P. GERRISH, 2002 The fixation probability of a beneficial allele in a population dividing by binary fission. *Genetica* **115**: 283–287.
- KIMURA, M., 1960 Optimum mutation rate and degree of dominance as determined by the principle of minimum genetic load. *J. Genet.* **57**: 21–34.
- KIMURA, M., 1967 On the evolutionary adjustment of spontaneous mutation rates. *Genet. Res.* **9**: 23–34.
- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- LECLERC, J. E., B. G. LI, W. L. PAYNE and T. A. CEBULA, 1996 High mutation frequencies among *Escherichia coli* and salmonella pathogens. *Science* **274**: 1208–1211.
- LEIGH, E. G., 1970 Natural selection and mutability. *Am. Nat.* **104**: 301–305.
- LEIGH, E. G., 1973 The evolution of mutation rates. *Genetics* **73** (Suppl.): 1–18.
- LEVIN, B. R., 1988 Frequency-dependent selection in bacterial-populations. *Philos. Trans. R. Soc. Lond. Ser. B* **319**: 459–472.
- LEVIN, B. R., and C. T. BERGSTROM, 2000 Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc. Natl. Acad. Sci. USA* **97**: 6981–6985.
- LEVINS, R., 1967 Theory of fitness in a heterogeneous environment. VI. The adaptive significance of mutation. *Genetics* **56**: 163–178.
- LEWONTIN, R. C., 1958 A general method for investigating the equilibrium of gene frequency in a population. *Genetics* **43**: 419–434.
- LIBERMAN, U., and M. W. FELDMAN, 1986 Modifiers of mutation rate: a general reduction principle. *Theor. Popul. Biol.* **30**: 125–142.
- MAO, E. F., L. LANE, J. LEE and J. H. MILLER, 1997 Proliferation of mutators in a cell population. *J. Bacteriol.* **179**: 417–422.
- MASEL, J., and A. BERGMAN, 2003 The evolution of the evolvability properties of the yeast prion [PSI<sup>+</sup>]. *Evolution* **57** (7).
- MATIC, I., M. RADMAN, F. TADDEI, B. PICARD, C. DOIT *et al.*, 1997 Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* **277**: 1833–1834.
- MILLER, J. H., 1998 Mutators in *Escherichia coli*. *Mutat. Res.* **409**: 99–106.
- OCHMAN, H., J. G. LAWRENCE and E. A. GROISMAN, 2000 Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**: 299–304.
- OLIVER, A., R. CANTON, P. CAMPO, F. BAQUERO and J. BLAZQUEZ, 2000

- High frequency of hypermutable *Pseudomonas aeruginosa* in cystic fibrosis lung infection. *Science* **288**: 1251–1253.
- ORR, H. A., 2000 Rate of adaptation in asexuals. *Genetics* **155**: 961–968.
- OTTO, S. P., and N. H. BARTON, 1997 The evolution of recombination: removing the limits to natural selection. *Genetics* **147**: 879–906.
- PAINTER, P. R., 1975 Clone selection and the mutation rate. *Theor. Popul. Biol.* **8**: 74–80.
- SNIEGOWSKI, P. D., P. J. GERRISH and R. E. LENSKI, 1997 Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**: 703–705.
- SNIEGOWSKI, P. D., P. J. GERRISH, T. JOHNSON and A. SHAVER, 2000 The evolution of mutation rates: separating causes from consequences. *Bioessays* **22**: 1057–1066.
- TADDEI, F., M. RADMAN, J. MAYNARD-SMITH, B. TOUPANCE, P. H. GOUYON *et al.*, 1997 Role of mutator alleles in adaptive evolution. *Nature* **387**: 700–702.
- TENAILLON, O., B. TOUPANCE, H. LE NAGARD, F. TADDEI and B. GODELLE, 1999 Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics* **152**: 485–493.
- TENAILLON, O., H. LE NAGARD, B. GODELLE and F. TADDEI, 2000 Mutators and sex in bacteria: conflict between adaptive strategies. *Proc. Natl. Acad. Sci. USA* **97**: 10465–10470.
- TRAVIS, J. M. J., and E. R. TRAVIS, 2002 Mutator dynamics in fluctuating environments. *Proc. R. Soc. Lond. Ser. B* **269**: 591–597.

Communicating editor: S. P. OTTO

## APPENDIX

**Details of the simulation:** The population sizes of the four types at time  $t$  are as follows:  $n(t)$  and  $m(t)$  are the numbers of unevolved nonmutators and unevolved mutators, respectively, while  $N(t)$  and  $M(t)$  are the numbers of evolved nonmutators and evolved mutators, respectively. Let the total population size be constant. Cells do not mutate from the nonmutator to the mutator phenotype, or vice versa.

At each time step (generation), a beneficial mutation may arise in each subpopulation. We calculate the probabilities of these events as follows. In the case of the nonmutators, the per-cell mutation rate per generation is  $\mu_n$ , so the number of mutations approximates a Poisson distribution with parameter  $\mu_n n(t)$ . Similarly, the number of mutations that occurs in the mutator population is approximately Poisson with parameter  $\mu_m m(t)$ . Selection is included as indicated in the main text. Beneficial mutants that newly arise are subject to loss by genetic drift. In each generation, the number of evolved mutators that pass to the next generation is sampled using a Poisson distribution with parameter  $(1 + b - s)M(t)$ . Similarly, the number of evolved nonmutators that form the next generation is found by a Poisson distribution with parameter  $(1 + b)N(t)$ .

The equations describing the simulation are therefore

$$\begin{aligned} Kn(t+1) &= n(t) - V_n(t) \\ KN(t+1) &= N^*(t) + V_n(t) \\ Km(t+1) &= (1 - s)m(t) - V_m(t) \\ KM(t+1) &= M^*(t) + V_m(t), \end{aligned} \quad (\text{A1})$$

where

$$\begin{aligned} N^*(t) &\sim \text{Poisson}[(1 + b) \times N(t)] \\ M^*(t) &\sim \text{Poisson}[(1 + b - s) \times M(t)] \\ V_n(t) &\sim \text{Poisson}[\mu_n n(t)] \\ V_m(t) &\sim \text{Poisson}[\mu_m m(t)], \end{aligned}$$

and where  $K$  is the normalizer given by the sum of the right-hand sides of Equations A1 divided by the total population size ( $n(t) + N(t) + m(t) + M(t)$ ), which we held constant over time. If  $N(t)$  or  $M(t)$  exceed the pre-set value of 100, then we allow them to continue growing without drift. That is, the equations for the evolved lineages become the deterministic expressions:  $N(t + 1) = [(1 + b)N(t) + V_n(t)]/K$  and  $M(t + 1) = [(1 + b - s)M(t) + V_m(t)]/K$ . We note that the equations above are approximate, since as  $N(t)$  approaches the  $n(t)$ , the fitness of  $M(t)$  must shift to be relative to evolved nonmutators rather than to unevolved nonmutators.

When the environment changes,  $n(t)$  in the new environment is assigned the value  $n(t) + N(t)$  from the previous environment, and similarly,  $m(t)$  is changed to  $m(t) + M(t)$  evaluated before the change.  $N(t)$  and  $M(t)$  are then reset to zero. The initial conditions, as given in the main text, are  $N(0) = M(0) = 0$ ,  $m(0) = m_0$ , and  $n(0) = 10^{10} - m_0$ .

**Definition of time cycle:** The *time cycle* used in the analytical model is defined as the period following an environmental shift, during which a beneficial mutation appears in the nonmutator subpopulation and proliferates until it reaches size  $n$ . We narrow our focus on cases in which environmental cycles appear after this event. For this analysis to be interpreted in the long run, we also assume that the next environmental shift occurs soon enough after the end of the time cycle so that the gain made by mutators is not substantially deteriorated through their cost  $s$ .

The length of the time cycle  $t^*$  is the time it takes for a beneficial mutant destined to survive drift to appear in the nonmutator subpopulation ( $W$ ) plus the time it takes for that mutant cell to sweep through the population ( $\tilde{w}$ ). For the purposes of comparing the simulation and the analytical models, let the (varying) number of unevolved nonmutator cells in the simulation be denoted by  $\hat{n}(t)$ , and we continue to denote the constant number of cells in the analytical model by  $n$ . Similarly, let the number of evolved nonmutators in the simulation model be  $\hat{N}(t)$ . Under the assumption that mutators ( $m + M$ ) remain relatively rare during the time cycle, the dynamics of  $\hat{n}(t)$  and  $\hat{N}(t)$  are similar to a diallelic haploid selection model whose dynamics are approximately logistic (CROW and KUMURA 1970, p. 192), leaving aside genetic drift. Thus, with  $n$  as the carrying capacity we can write

$$\hat{N}(t) \approx \frac{n}{1 + ((n - N_0)/N_0)e^{-b(t-W)}} \quad \text{for } t > W, \quad (\text{A2})$$

where  $N_0$  is the initial number of evolved mutators.

*Case 1:* ( $n\mu_n f_n < 1$ ). If the mutant were to grow exponentially (Equation 6), then the time taken for  $N(t)$  to

reach  $n$  from  $N_0 = 1$  is  $\tilde{w} = \ln(n)/b$ . Substituting this quantity for time in (A2) gives  $\hat{N}(t^*) = n^2/(2n - 1) \simeq n/2$ . That is, the end of the time cycle in the analytical model roughly corresponds in the simulation model to the time at which the unevolved nonmutators ( $\hat{n} = n - \hat{N}$ ) decline to  $n/2$ . Because the dynamics are approximately logistic,  $\hat{n}(t)$  spends most of the time period close to  $n$ .

*Case 2:* ( $n\mu_n f_n > 1$ ). To compare Equation 9 with  $\hat{N}(t)$ , we need to choose  $N_0$  such that the two functions are comparable in the long term (as  $t^*$  is approached). We therefore use  $N(t) = n\mu_n f_n(1 + 1/b)e^{bt}$  instead of (9) and set  $N_0 = n\mu_n f_n(1 + 1/b)$ . The time cycle in this case has the deterministic length given by the second part of Equation 11. Substituting this time period in (A2) simplifies to

$$\hat{N}(t^*) = \frac{n}{1 + (1 - N_0/n)(b/(\mu_n f_n + b))}.$$

Since  $\mu_n f_n \ll b/(1 + b)$ ,  $\hat{N}(t^*) \simeq n/2$ . Again, we conclude that during the time cycle,  $\hat{n}(t)$  slowly declines from  $n$  to  $\sim n/2$ .

**The distribution of  $R$ :** We first present the model for cases where  $n\mu_n f_n < 1$ . Consider the two exponentially distributed random variables:  $X \sim \text{Exp}(\lambda_m)$  and  $W \sim \text{Exp}(\lambda_n)$ . To find the probability density function (PDF) of  $U = X - W$ , we perform the following convolution:

$$f_U(u) = \begin{cases} \int_u^\infty f_X(x)f_W(x - u) dx & \text{if } u \geq 0 \\ \int_0^\infty f_X(x)f_W(x - u) dx & \text{if } u < 0. \end{cases} \quad (\text{A3})$$

The integration limits are given by the constraint that the variables  $W$  and  $X$  must both be greater than zero.

We therefore obtain

$$f_U(u) = \begin{cases} f_{U1}(u) = \frac{\lambda_m \lambda_n}{\lambda_m + \lambda_n} e^{-\lambda_m u} & \text{if } u \geq 0 \\ f_{U2}(u) = \frac{\lambda_m \lambda_n}{\lambda_m + \lambda_n} e^{\lambda_n u} & \text{if } u < 0. \end{cases} \quad (\text{A4})$$

From Equation 7,

$$\begin{aligned} \ln(M(t^*)) &= (b - s)(t^* - X) \\ &= (b - s)((t^* - W) - U). \end{aligned} \quad (\text{A5})$$

The term  $t^* - W = \tilde{w}$  is the time taken for a successful beneficial mutation in nonmutators to reach size  $n$ , which equals the constant  $\ln(n)/b$  (from Equation 6). Thus

$$R = -(b - s)U + \frac{b - s}{b} \ln(n) - \ln(m_0), \quad (\text{A6})$$

or

$$U = \frac{R - k_2}{k_1}, \quad (\text{A7})$$

where

$$k_1 = -(b - s) < 0, \quad \text{by assumption,} \quad (\text{A8})$$

and

$$k_2 = \frac{b - s}{b} \ln(n) - \ln(m_0). \quad (\text{A9})$$

We now transform the distribution of  $U$  (A4) using (A6) to produce the PDF of the log gain made by the mutator in a given time cycle ( $R$ ):

$$f_R(r) = \begin{cases} f_{R1}(r) = \frac{\lambda_m \lambda_n}{(-k_1)(\lambda_m + \lambda_n)} e^{-\lambda_m(r - k_2)/k_1} & \text{if } r \leq k_2 \\ f_{R2}(r) = \frac{\lambda_m \lambda_n}{(-k_1)(\lambda_m + \lambda_n)} e^{\lambda_n(r - k_2)/k_1} & \text{if } r > k_2. \end{cases} \quad (\text{A10})$$

The cumulative density function (CDF) of  $R$  is given by

$$\begin{aligned} F_R(r) &= \begin{cases} \int_{-\infty}^r f_{R1}(y) dy & \text{if } r \leq k_2 \\ \int_{-\infty}^{k_2} f_{R1}(y) dy + \int_{k_2}^r f_{R2}(y) dy & \text{if } r > k_2 \end{cases} \\ &= \begin{cases} \frac{\lambda_n}{\lambda_m + \lambda_n} e^{-\lambda_m(\tilde{w} - \tilde{v} - r/(b - s))} & \text{if } r \leq k_2 \\ 1 - \frac{\lambda_m}{\lambda_m + \lambda_n} e^{-\lambda_n(\tilde{w} - \tilde{v} - r/(b - s))} & \text{if } r > k_2. \end{cases} \end{aligned} \quad (\text{A11})$$

Because we are generally interested in rare mutators (small  $m_0$ ), we focus on the first case of this equation ( $r \leq k_2$ ). Equation 13 is based on this first case.

The case where  $n\mu_n f_n > 1$  follows the same reasoning but is simpler to derive because  $W = 0$ . In this case, the length of the time cycle is deterministic and also equal to the duration of the selective sweep:  $t^* = \bar{t} = \tilde{w} = [\ln(1 + (b/\mu_n f_n)) - \ln(1 + b)]/b$ . We next transform the waiting time until the evolved mutator appears to produce the PDF of the log gain:

$$f_R(r) = \frac{\lambda_m}{(b - s)} \exp\left[-\lambda_m\left(\tilde{w} - \frac{\ln(m_0)}{b - s} - \frac{r}{b - s}\right)\right]. \quad (\text{A12})$$

We can again find the CDF and thus write

$$\Pr(R > r) = 1 - e^{-\lambda_m(\tilde{w} - \tilde{v} - r/(b - s))}. \quad (\text{A13})$$