Evaluating the Impact of Population Bottlenecks in Experimental Evolution

Lindi M. Wahl,*[,1] Philip J. Gerrish[†,2] and Ivan Saika-Voivod*

**Applied Mathematics, University of Western Ontario, London, Ontario N6A 5B7, Canada and*
[†]Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Manuscript received January 7, 2002
Accepted for publication July 23, 2002

ABSTRACT

Experimental evolution involves severe, periodic reductions in population size when fresh media are inoculated during serial transfer. These bottlenecks affect the dynamics of evolution, reducing the probability that a beneficial mutation will reach fixation. We quantify the impact of these bottlenecks on the evolutionary dynamics, for populations that grow exponentially between transfers and for populations in which growth is curbed by a resource-limited environment. We find that in both cases, mutations that survive bottlenecks are equally likely to occur, per unit time, at all times during the growth phase. We estimate the total fraction of beneficial mutations that are lost due to bottlenecks during experimental evolution protocols and derive the "optimal" dilution ratio, the ratio that maximizes the number of surviving beneficial mutations. Although more severe dilution ratios are often used in the literature, we find that a ratio of 0.1–0.2 minimizes the chances that rare beneficial mutations are lost. Finally, we provide a number of useful approximate results and illustrate our approach with applications to experimental evolution protocols in the literature.

RAPIDLY evolving organisms such as bacteria, viruses, and protozoa will adapt to laboratory conditions on short, experimentally feasible timescales. In a single controlled experiment, major evolutionary change may occur in these populations, while both phenotypic and genotypic differences can be monitored (Lenski *et al.* 1991; Lenski and Travisano 1994; Rosenzweig *et al.* 1994; Bell and Reboud 1997; Bull *et al.* 1997; Sniegowski *et al.* 1997; Rainey and Travisano 1998; Treves *et al.* 1998; Papadopoulos *et al.* 1999; Wichman *et al.* 1999). Understandably, these experiments are generating enormous interest in evolutionary biology (Appenzeller 1999).

The usual elements of the experimental protocol are as follows. In experiments involving bacteria, these are grown at a constant temperature in a sugar-rich broth. After a phase of population growth, a small set of the founder population is typically sampled and reintroduced into an identical but unpopulated environment. This procedure is repeated over many generations (serial passaging). When viruses are studied, a host species, commonly a bacterium, is maintained alongside the phage in culture. In a two-stage chemostat, samples from the phage tube are extracted and used to reinoculate the system when tubes are changed, which occurs on a regular basis.

[1]*Corresponding author:* Applied Mathematics, University of Western Ontario, London, Ontario N6A 5B7, Canada.
E-mail: lwahl@uwo.ca

[2]*Present address:* Programa de Investigacion en Matemáticas Aplicadas, Instituto Mexicano del Petroleo, Eje Central Lázaro Cárdenas No. 152, Colonia San Bartolo Atepehuacán, México, D.F. 07730.

For most experiments in the field, therefore, the population "life cycle" as described above has an important distinguishing feature: population bottlenecks. These severe, regular bottlenecks profoundly affect the dynamics of evolution, reducing the probability that a beneficial mutation will reach fixation. Before interpreting results obtained by experimental evolution, we would therefore like to understand the impact of these bottlenecks on the evolving population. We need to answer the following intriguing questions: What fraction of beneficial mutations are lost due to population bottlenecks? Which mutations are preferentially lost? And how do bottlenecks ultimately affect the variability of the evolutionary trajectory? These questions are important not only for experimental populations, but also for natural infection cycles: Bacteria or viruses, for example, may colonize hosts from an initial inoculum that represents a population bottleneck.

The answer to each of these questions relies fundamentally on our understanding of the fixation probability: the probability that a rare beneficial mutation will ultimately fix in a population. For a population of constant size, this question was first addressed by the "great trinity" (Crow 1994) of population genetics, Fisher (1922), Haldane (1927), and Wright (1931). Kimura (1957, 1962) was able to extend this classic work in a number of directions using a diffusion approximation, although populations of constant size were still assumed. For cyclic population sizes, Ewens (1967) first computed the fixation probability using an iterative solution to a branching Poisson model. This work was extended to treat a number of interesting cases in a recent article

by Otto and Whitlock (1997). In all of the above solutions, the offspring distribution is assumed to be Poisson; the implications of relaxing this assumption are explored by Pollak (2000).

In a previous article, we describe two derivations for the *extinction* probability in populations with periodic bottlenecks (Wahl and Gerrish 2001). We use extinction probability, the probability that a rare mutation is lost due to bottlenecks, rather than fixation probability because factors other than bottlenecks may influence the fixation probability in asexual populations; clonal interference (Gerrish and Lenski 1998) is one example. In the sections that follow, we extend our approach numerically to treat populations in resource-limited environments. We then explore the implications of these results for experimental evolution.

Although many of these implications follow more-or-less directly from previous work (Ewens 1967; Otto and Whitlock 1997; Wahl and Gerrish 2001), their impact for experimental evolution has not been elucidated. We find, for example, that bottlenecks, like genetic drift, filter our view of beneficial mutations: The selective advantage of mutations that eventually survive bottlenecks is about twice as large as the mean selective advantage of all beneficial mutations that occur. We also find that even in populations that grow exponentially between bottlenecks, and therefore produce many more mutations toward the end of the growth phase, successful mutations are equally likely to occur at all times during population growth. Finally, we are able to determine the dilution ratio—∼0.1–0.2—which allows the largest number of beneficial mutations to survive. For the approximation we use for the survival probability, this optimal dilution ratio is completely independent of such factors as population size and growth rate.

EXTINCTION PROBABILITY IN EXPERIMENTAL EVOLUTION

To model a serial passaging protocol, we consider a population of initial size N_0, which grows to a final size N_f during time interval $[0, \tau]$. At time τ, the population is sampled with dilution ratio D, such that $DN_f = N_0$. This cycle of growth and sampling is repeated many times. We are interested in the fate of a rare beneficial mutation with selective advantage s, which might occur for the first time at time t during the growth phase.

Exponential growth: In a previous article (Wahl and Gerrish 2001), we derive the extinction probability, $V(t, s)$, for such a mutation in a population that grows exponentially during the growth phase at rate r. $V(t, s)$ reflects the probability that a single mutation with advantage s occurring at time t during a growth phase will leave no descendants in the distant future. We found the following analytical approximation to the extinction probability when s is small:

$$V(t, s) \approx 1 - 2se^{-rt}r\tau. \quad (1)$$

Growth in a limited resource: We want to test the validity of the exponential model of population growth by comparing it with extinction probabilities for growth in a limited resource environment. A range of models are available for resource-limited growth (Edelstein-Keshet 1988); we have chosen the following set of differential equations (Levin *et al.* 2000) because parameters are available for a specific serial transfer protocol. In a later section of the article, we consider another, simpler model that is based on the environmental carrying capacity.

The dynamics of population growth in the "resource concentration" model are given by

$$\begin{aligned} \dot{n} &= \psi(R)\, n \\ \dot{m} &= \psi(R)(1 + s)\, m \\ \dot{R} &= -\psi(R)E(n + m(1 + s)) \end{aligned} \quad (2)$$

with $\psi(R) = WR/(K + R)$ (Stewart and Levin 1973). This system models a volume, ν, of the serial transfer culture, where n is the density, or concentration of "wild-type" individuals in the culture, m is the concentration of individuals carrying the gene of interest, and R is the resource concentration remaining in the environment. $\psi(R)$ gives the growth rate, per hour, for the population at a given resource concentration and is determined by Michealis-Menten kinetics with maximum growth rate W and half-maximal concentration K. The conversion parameter E gives the amount of resource required to produce a single new individual in the population.

Integrating these equations numerically, we determine the population density during a growth phase in a serial passaging protocol. The frequency of the gene of interest at the end of the growth phase, z, is simply calculated as $z = m(\tau)/(m(\tau) + n(\tau))$.

The sampling step is modeled as a binomial process. For gene frequency z and sample size N_0, the probability that j copies of the gene are in the sample is given by

$$p_j = \binom{N_0}{j} z^j (1 - z)^{N_0 - j}.$$

The output from the sampling process is a distribution of possible values for j and their respective probabilities. Each of these values of j is then treated as an input to the next phase of growth and sampling ($m(0) = j/\nu$), and the outputs are weighted by the appropriate probabilities and summed.

The ultimate probability of extinction computed by these numerical methods is denoted $V_n(t, s)$. To estimate this value we examine the total probability that zero copies of the original mutation are present in the population after each bottleneck and continue computations until this probability changes negligibly from bottleneck to bottleneck.

The parameter values provided by Levin *et al.* (2000)

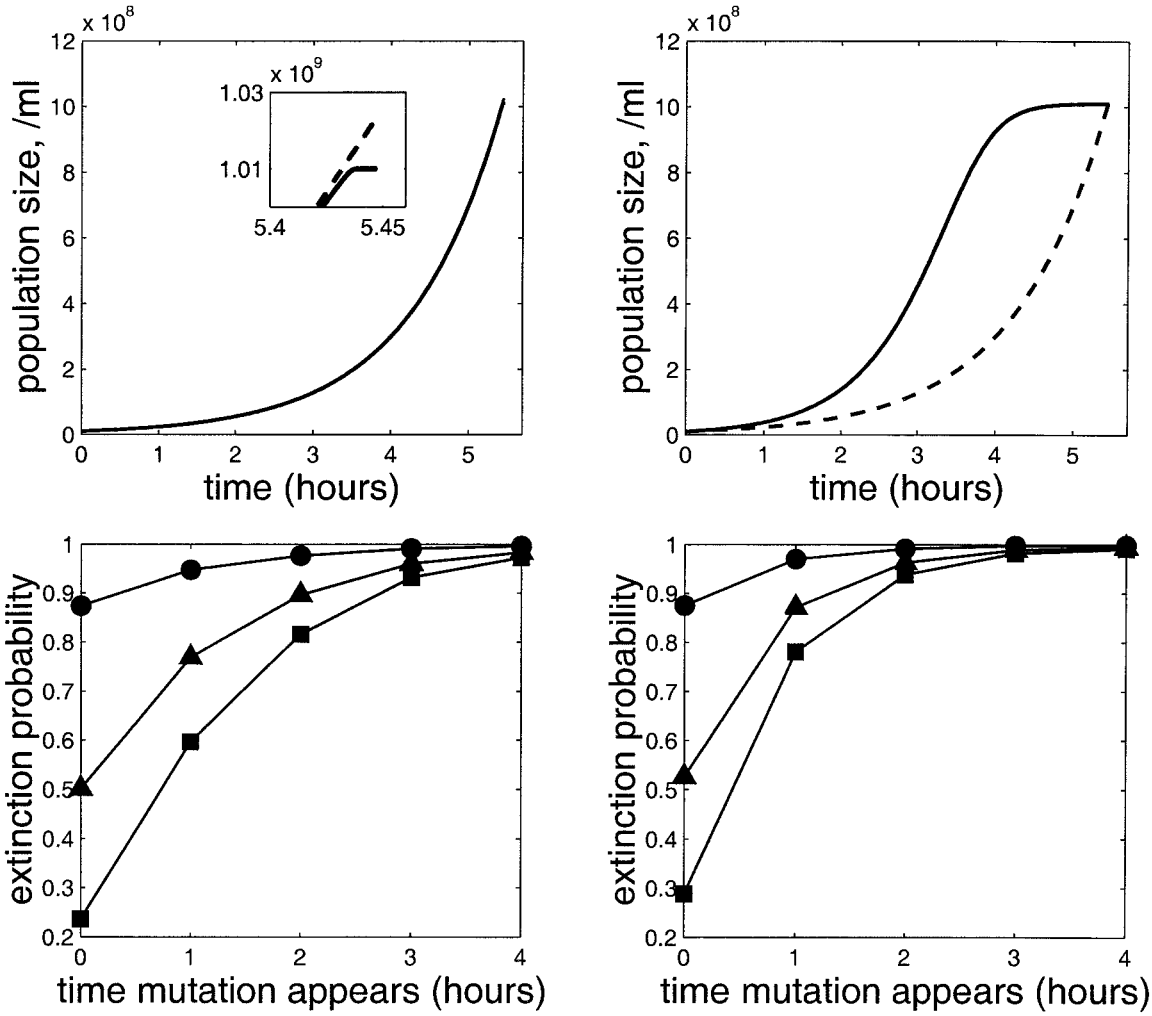

FIGURE 1.—Extinction probabilities, resource-limited growth. An experimental protocol in which resource-limited growth occurs in a 10-ml volume with an initial population density of 10^7 bacteria per ml is modeled. After the resource has been consumed and growth has stopped (~5 hr), the population is sampled with a dilution factor $D = 0.01$ and the process is repeated. The top left shows the total population size *vs.* time (solid line). The dashed line shows for comparison the time course of population growth when growth is exponential; differences are visible in the inset only. At time t during the growth phase, a single copy of a beneficial mutation is introduced. Symbols in the bottom left show numerical estimates for the probability that the mutation is ultimately eliminated by the bottlenecks, given a selective advantage $s = 0.01$ (circles), 0.05 (triangles), and 0.1 (squares). Parameters used were as per LEVIN *et al.* (2000): $W = 0.85$, $K = 0.25$, $R_0 = 500$, $E = 5 \times 10^{-7}$. The right top and bottom show the same results for a parameter set that gives a slower turnover of population growth; parameters that differ are $V = 2$, $K = 250$. The dashed line again plots the population size under exponential growth for comparison.

describe the growth of *Escherichia coli* in Davis minimal medium, supplemented with 500 μg/ml glucose as the sole and limiting carbon source. Cultures (10 ml) were initiated with $\approx 10^7$ bacteria/ml and grew to final densities of $\approx 10^9$ bacteria/ml in just over 5 hr. Figure 1, top left, shows the total population size as a function of time for these parameters (solid line). Integration was stopped when the population ceased to grow, that is, when the resource concentration reached zero. For comparison, we created a second parameter set that allowed the population to grow to the same final size in the same time, but for which the "turnover" was more gradual (Figure 1, top right, solid line). The figure also shows the population size for strictly exponential growth at the initial growth rate (dashed line). This curve is indistinguishable from the experimental growth curve, differing only during the final few minutes of the growth phase (inset). Thus for the experimentally determined parameter values, near-exponential growth is maintained in the resource-limited system until the resource is severely depleted.

In Figure 1, bottom, the ultimate extinction probability is plotted for mutations with selective advantages $s =$ 0.01, 0.05, and 0.1 in a resource-limited environment. For the parameter values from the literature (Figure 1, bottom left), these extinction probabilities were slightly lower than those calculated for exponential growth, differing in the third decimal place (data not shown). For

these parameter values, the only significant effect of resource-limited growth is to limit the time over which exponential growth can be maintained. Extinction probabilities, and by extension other aspects of the evolutionary dynamics, are affected very little. When the shoulder of the growth curve occurs earlier and more gradually (Figure 1, right), extinction probability at each time is increased.

TIME DISTRIBUTION OF SUCCESSFUL MUTATIONS

We want to know when "successful" mutations occur, that is, mutations that survive not only the first bottleneck they face, but all subsequent bottlenecks. The expected number of mutations that occur at time t and survive bottlenecks, $\lambda(t)$, is given by the following integral:

$$\lambda(t) = \int_{s=0}^{\infty} \dot{N}\mu\alpha e^{-\alpha s}(1 - V(t, s))\, ds. \qquad (3)$$

Here $\dot{N}$ is the time derivative of N, that is, the number of replications at time t, and μ is the beneficial mutation rate per replication. Our implicit assumption is that population growth is a "pure birth process"; we assume that the death rate of individuals in the population is negligible compared to the growth rate during the growth phase. Note that this assumption would not hold, generally, in chemostat populations. (If a population has a significant death rate, Equation 3 gives an upper bound on the expected number of mutations.) For exponential growth, where $N(t) = N_0 e^{rt}$, the number of replications $\dot{N}$ is simply $rN(t) = N_0 re^{rt}$. For resource-limited growth, we evaluate system 2 numerically, and the expected number of replications is given by $\dot{n} = \psi(R)n$. (Also note that μ is the rate at which beneficial mutations occur per new individual contributed to the population. In bacteria, for example, this mutation rate is twice the usual mutation rate per genome per replication, because a new strand is synthesized in each of two daughter cells after bacterial fission.)

The factor $1 - V(t, s)$ is the survival probability of a mutation with selective advantage s, but we need to know the probability distribution of s to complete Equation 3. To determine this distribution, note that we require only the distribution of *advantageous* mutations, drawn from a very large number of mutational neighbors, M, of the replicating genome. If the replicating genome has fitness W_0, for example, the fitness values of all possible daughter genomes (W_j, $j = 1 \ldots M$) are members of some unknown fitness distribution. Define the set B to be all j such that $W_j > W_0$; *i.e.*, B is the set of indices of beneficial mutations. If the right tail of the parent fitness distribution approaches zero exponentially, then the *fitness difference*, $S_i = W_i - W_0$, where $i \in B$ has an exponential distribution that is independent of the wild-type fitness (for large M). If the right tail is heavier than exponential such that it "varies regularly" (see Feller 1971 for definition), the *selection coefficient*, $S_i = W_i/W_0 - 1$, $i \in B$, has an exponential distribution that is independent of the wild-type fitness (H. A. Orr and P. J. Gerrish, personal communication). S_i has an invariant exponential distribution because W_0 and the W_i are extreme values of the unknown parent distribution of fitnesses. Our results can thus be tailored to either assumption about tail behavior by defining the S_i appropriately. We therefore use an exponential function, $\alpha e^{-\alpha s}$, to model the distribution of S, defined appropriately, for beneficial mutations.

Equation 3 illustrates a catch-22 for beneficial mutations in bottlenecked populations: While mutations that arise early in the growth phase have a significant probability of survival, the expected number of mutations at these times is very small; conversely late mutations are likely to occur and unlikely to survive.

As an example of how these two factors interact, we plot the probability that a beneficial mutation occurs at time t and ultimately survives bottlenecks in Figure 2, for exponential growth (solid line) and resource-limited growth (dashed line). In both cases, the distribution is relatively flat for all times throughout the growth phase. This implies that although most replications occur toward the end of the growth phase, mutations that are ultimately successful occur at all times during growth. Note that by "successful" we mean mutations that survive the direct effects of the bottleneck—beneficial mutations may also be lost due to drift (Crow and Kimura 1970) or competition (Gerrish and Lenski 1998).

FITNESS DISTRIBUTION OF SUCCESSFUL MUTATIONS

In analogy to the previous section, we can also determine the number of mutations with selective advantage s that occur during one growth phase and ultimately survive bottlenecks, $\Lambda(s)$. In this case we evaluate the following integral:

$$\Lambda(s) = \int_{t=0}^{\tau} \dot{N}\mu\alpha e^{-\alpha s}(1 - V(t, s))\, dt. \qquad (4)$$

Here we find another catch-22: Mutations with large s are most likely to survive a bottleneck, while mutations with small s are most likely to occur.

Typical distributions for $\Lambda(s)$ are illustrated in Figure 3. The solid lines plot the distribution of s for mutations that ultimately survive the bottleneck protocol, for three different values of the dilution ratio ($D = 0.1$, 0.01, and 0.001 from top to bottom, respectively). These curves are plotted again in the inset for comparison with the original distribution of s (dashed line). We find that bottlenecks severely affect the distribution of beneficial mutations, effectively eliminating mutations with very small selective advantage and vastly reducing the frequency of mutations with moderate benefit.

Note, however, that the dotted lines in Figure 3 correspond to a dilution ratio of 0.5. Diluting at 0.5 implies

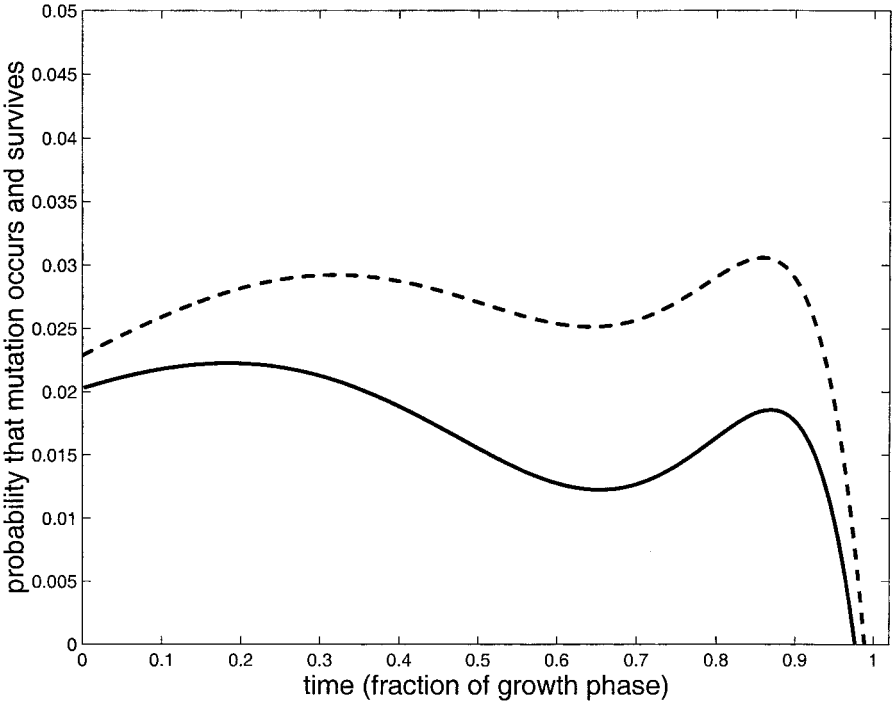

FIGURE 2.—Time distribution for successful mutations. The probability that a mutation occurs at time t and ultimately survives bottlenecks is plotted for exponential (dashed line) and resource-limited (solid line) growth. Note that the distribution is fairly flat; mutations that are ultimately successful are equally likely to occur at all times. Time is normalized by the length of the growth phase, τ. Parameters for exponential growth are $N_0 = 1 \times 10^5$, $r = \ln 2$, $\tau = 7$, $\mu = 5 \times 10^{-5}$. Parameters for resource-limited growth are as given in Figure 1, with $\tau = 5.45$ and $\mu = 4 \times 10^{-9}$. In both cases we have assumed that the mutation has selective advantage $s = 0.1$.

that the population is allowed to double for one generation, but only one-half of these offspring survive; $D = 0.5$ is formally equivalent to classical models of a population of constant size experiencing genetic drift. For comparison, we plot results for two constant population sizes: a population size of N_0 and a population size of N_f (top and bottom dotted lines, respectively). Perhaps counterintuitively, the total number of successful mutations is *lower* in both of these cases than it would be for a population experiencing bottlenecks with $D = 0.1$. Although bottlenecks reduce the fixation *probability* of any mutation (see Figure 4), many more mutations occur when the population is allowed to grow exponentially for several generations, as opposed to just a single doubling, between bottlenecks. This implies that the total substitution *rate* of mutations can be larger in bottlenecked populations than in populations of constant size, although the effect is small if the constant population can be maintained at N_f. This point will be taken up again in the DISCUSSION.

SOME USEFUL APPROXIMATIONS

Time distribution of successful mutations: Figure 2 demonstrated that the distribution across time for mutations that are ultimately successful is remarkably flat. We formalize this intriguing result as follows. Using Equation 1 as an approximation for the extinction probability in Equation 3, the expected number of successful mutations at each time t is roughly

$$\lambda(t) \approx \mu r N_0 e^{rt} \int_0^\infty \alpha e^{-\alpha s}(2e^{-rt} rs\tau)\, ds = \frac{2N_0 \mu r^2 \tau}{\alpha}.$$

Note that, as predicted, the expected number of successful mutations does not depend on t; mutations that are ultimately successful occur at all times during the growth phase with equal probability.

Distribution of s for successful mutations: Similarly, we can approximate the distribution of the selective advantage for successful mutations using Equations 1 and 4. Here we find that the expected number of successful mutations during one growth phase with advantage s is

$$\Lambda(s) \approx \alpha e^{-\alpha s} \int_0^\tau \mu r N_0 e^{rt}(2e^{-rt} rs\tau)\, dt = 2N_0 \mu (r\tau)^2 \alpha s e^{-\alpha s}.$$

This allows us to compute two interesting results. First, since the total number of successful mutations is proportional to $se^{-\alpha s}$, the probability distribution for successful mutations must be given by $\alpha^2 se^{-\alpha s}$. The mean of this distribution is $\bar{s}_o = 2/\alpha$. Thus the mean value of the selective advantage for successful mutations, *i.e.*, for those mutations that would actually be observed during experimental evolution, is twice the mean value of the underlying distribution of the selective advantage. This result is not surprising; the same is true for mutations that survive drift.

Second, for a mutation with selective advantage s, the approximate probability of fixation can be written as a function of the dilution ratio:

$$U(s) \approx 2sD(\ln D)^2.$$

This result is obtained by dividing the number of successful mutations with advantage s by the total number of mutations that occur with advantage s. Once again

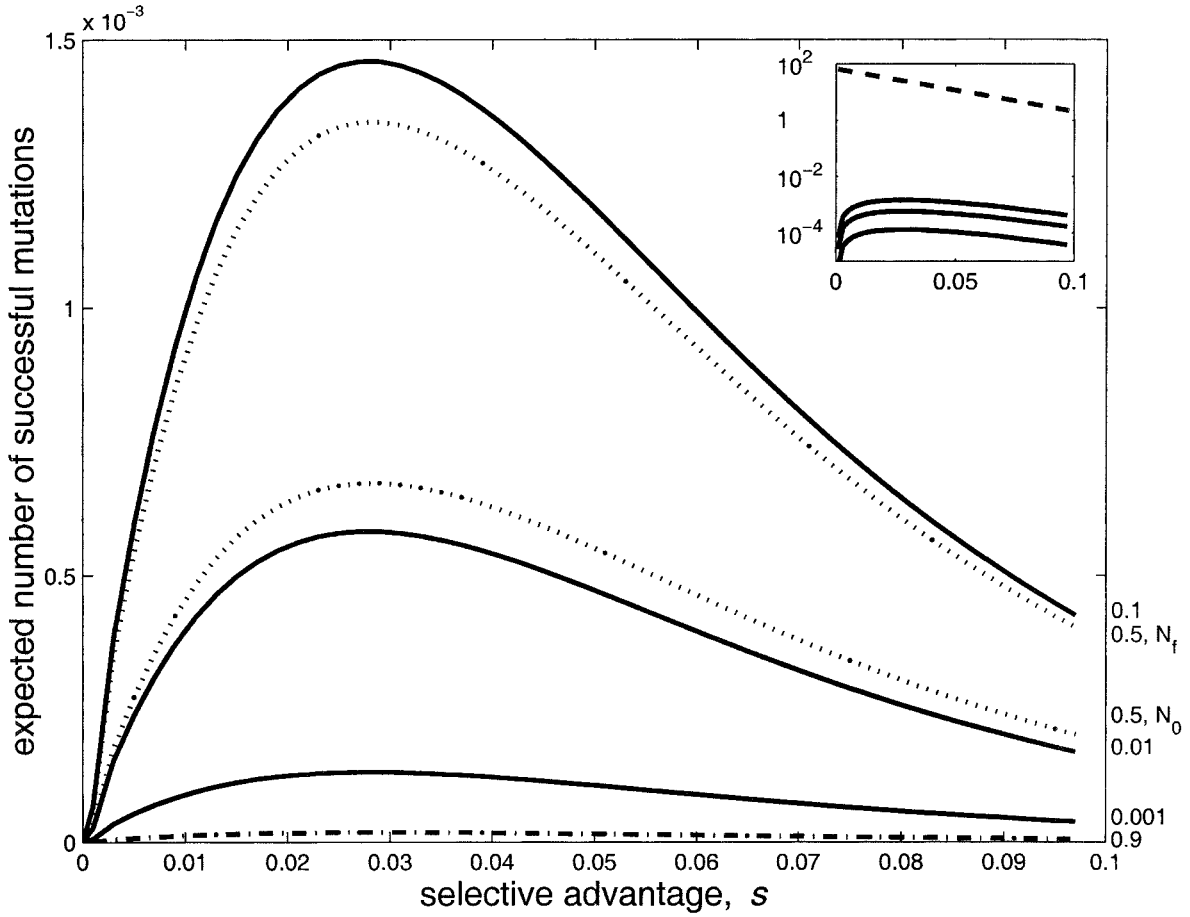

FIGURE 3.—Distribution of s for successful mutations. The expected number of successful mutations with selective advantage s is plotted against s. Results are shown as solid lines for dilution factors of 0.1, 0.01, and 0.001, from top to bottom, respectively. For comparison, results for dilution factors of 0.9 (dot-dashed line) and 0.5 (dotted lines) are also shown. Note that the dilution rate of 0.5 corresponds to a constant population size subject to genetic drift (see text for details). We plot two cases, a constant population size of N_f and N_0 (top and bottom dotted lines, respectively). The distributions for $D = 0.1$, 0.01, and 0.001 are redisplayed on a semilog plot in the inset (solid lines) for comparison with the distribution of all mutations that are expected to occur (dashed line). Note that bottlenecks, like drift, reduce the number of mutations by many orders of magnitude. Exponential growth was assumed with $N_0 = 1 \times 10^9 D$, $N_f = 1 \times 10^9$, $r = \ln 2$, $\tau = -\ln D/r$, $\mu = 2 \times 10^9$. For the population held constant at N_f, $N_0 = 1 \times 10^9$ and $N_f = 2 \times 10^9$.

we find a classic result: The probability that a mutation ultimately survives varies as $2s$ (HALDANE 1927). For bottlenecks, however, this probability is reduced by the factor $D(\ln D)^2$. Figure 4 shows this reduction in the classical fixation probability as a function of the dilution factor. We observe, for example, that fixation probability is ~75% of the classical prediction for dilutions of 1:10, but falls to only 1% of the classical value for dilutions of $1:10^4$. More importantly, we see that the fixation probability is maximized when $D = e^{-2} \approx 0.135$. This suggests that for any experimental evolution protocol with repeated bottlenecks, a dilution ratio of ~0.135 will minimize the probability that beneficial mutations are lost during bottlenecks. This intriguing result is explained further in the following subsection.

Total number of beneficial mutations that occur and survive: We can also substitute Equation 1 into a double integral over t and s to estimate the number of mutations that occur during one growth phase and ultimately survive. We find

$$L \approx \int_0^\infty \int_0^\tau \alpha e^{-\alpha s} \mu r N_0 e^{rt} (2e^{-rs\tau})\, dt ds$$

$$\approx \frac{2}{\alpha} \mu N_0 (\ln D)^2$$

$$= \frac{2}{\alpha} \mu N_f D (\ln D)^2. \quad (5)$$

From the central line of this equation, we see that L is maximized as we might expect for high mutation rates and a distribution of s that has a large mean value. L is also maximized by having a large initial population, N_0. If N_0 is fixed, L is maximized by having an infinitely small dilution ratio. This means that for a fixed N_0 (and variable N_f), the greatest number of successful mutations

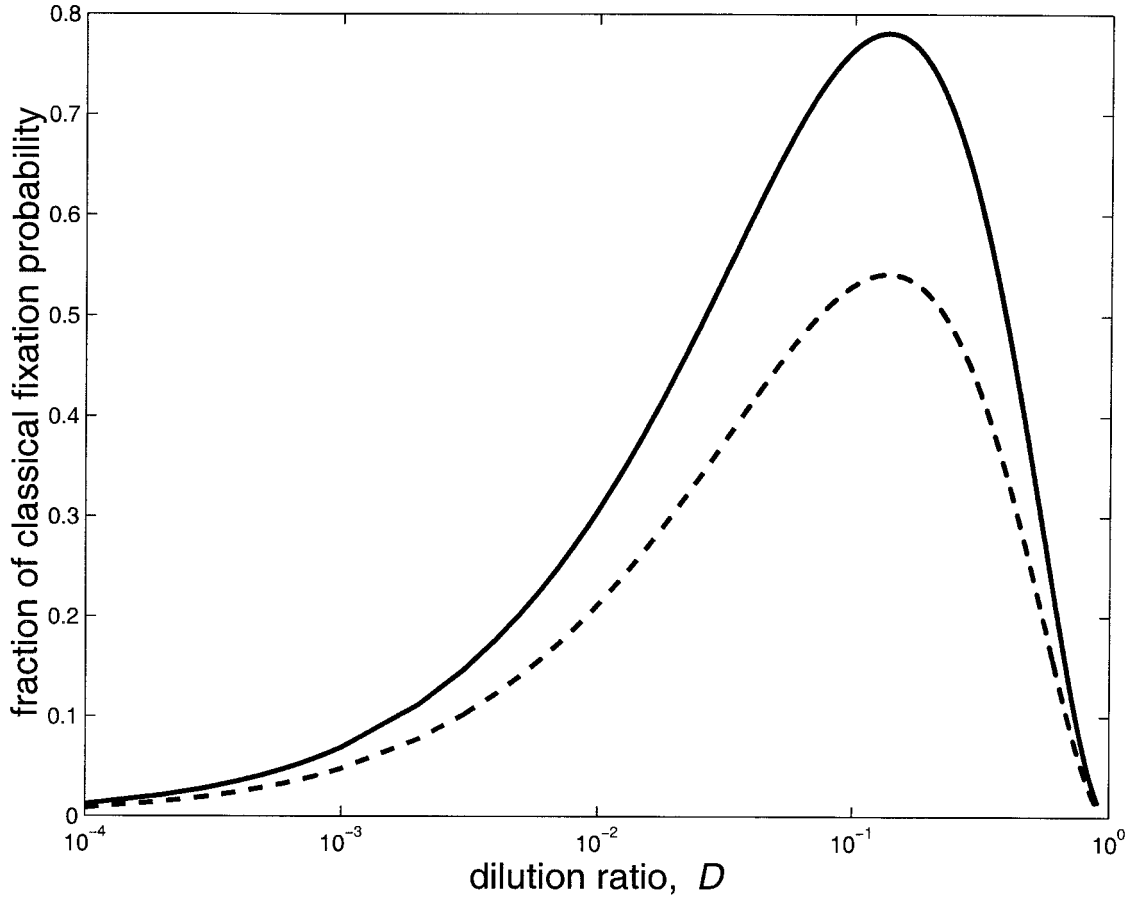

FIGURE 4.—Reduction in fixation probability due to bottlenecks. The probability that a mutation with selective advantage s survives drift is classically estimated as $U(s) \approx 2s$. The probability that such a mutation survives population bottlenecks is given by $2s\,D(\ln D)^2/r$; that is, $U(s)$ is reduced by a factor $D(\ln D)^2/r$. This factor is plotted against s for $r = \ln 2$ (solid line) and $r = 1$ (dashed line). For severe dilution factors, the probability of survival is greatly reduced by population bottlenecks. Note that the optimal dilution ratio occurs at $D = e^{-2} \approx 0.135$.

is produced by using an infinitely long growth phase; *i.e.*, one should never actually dilute the culture. From the last line of Equation 5, we find the more realistic case where N_f is constrained. In this situation the largest number of successful mutations is produced when N_f is constrained to be as large as possible, and $D = e^{-2} \approx 0.135$.

Note that each of the expressions in this section relies on the approximation to $V(t, s)$ given in Equation 1, which relies on the assumptions that s is small and growth is exponential. We reexamine our derivation of the optimal dilution ratio below, for resource-limited growth.

MAXIMIZING THE RATE OF EVOLUTION

When tracking phenotypic or genotypic change over the course of an experiment, we might want to minimize the probability that a beneficial mutation is eliminated by chance. This does not imply that we wish to alter the selective pressures on specific mutations; instead, we wish to reduce the overall probability that beneficial mutations are eliminated by population bottlenecks in the experimental protocol.

To address this question with greater accuracy, we consider population growth in a limited-resource environment. Unfortunately the model described in system 2 is somewhat unwieldy for our purposes, and so in this section we have chosen to use a simpler model of population growth:

$$\dot{x} = r\left(1 - \frac{x}{K}\right)x. \tag{6}$$

Here x is the population density, r is the growth rate (per unit time), and K is the carrying capacity of the environment (test tube).

The solution to this equation is

$$x = \frac{Ae^{rt}}{1 + (A/K)e^{rt}}, \tag{7}$$

where $A = N_0/(1 - N_0/K)$, and N_0 is the initial population size. Recall that the dilution ratio, D, is defined as the ratio N_0/N_f, where N_f is the population size at the

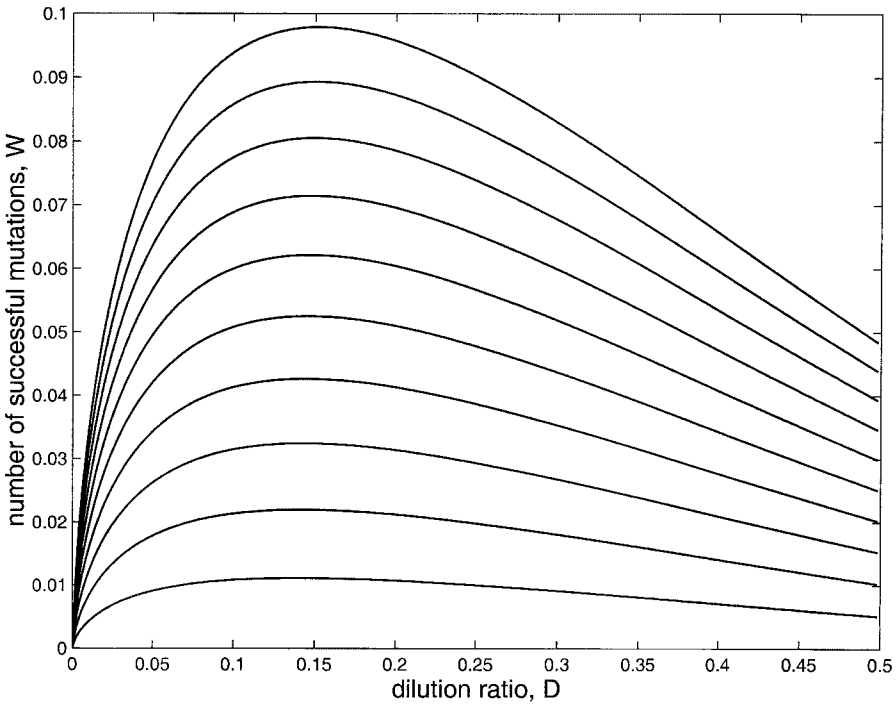

FIGURE 5.—The optimal dilution ratio. *W*, the expected number of surviving mutations, is plotted against the dilution ratio for $s = 0.01, 0.02, 0.03, \ldots 0.1$. We find that the number of surviving beneficial mutations is maximized in all cases for $D \approx 0.15$. The other parameters used in the model are $r = 0.85$/hr, $K = 10^9$, and $\mu = 10^{-9}$. Results were determined numerically.

end of one growth phase of duration τ. We note that typically $N_f \approx K$ and so $D \approx N_0/K$.

Let L be the total number of mutations that are expected to occur during one growth phase and that will ultimately survive bottlenecks. L is given by the double integral,

$$L = \int_0^\infty \alpha e^{-\alpha s} W(s)\, ds, \tag{8}$$

where $W(s)$, the expected number of successful mutations with selective advantage s, is given by

$$W(s) = \int_0^\tau \mu \dot{x}(t)(1 - V(t, s))\, dt. \tag{9}$$

Here $V(t, s)$ is the extinction probability as previously defined, and we again assume that the number of mutants arising in the population is proportional to the total number of replications, $\mu\dot{x}(t)$; *i.e.*, the death rate in the population during the growth phase is negligible.

W is a function of s, N_0 (or alternately D), and τ. Suppose we fix τ; that is, we wish to sample our population every 24 hr, for example. We can then solve for the dilution ratio, D, which maximizes W for a particular value of s. Figure 5 plots W as a function of the dilution ratio for several values of s; these results were obtained numerically. We find that the dilution ratio that maximizes W is only weakly dependent on s. Thus the total number of successful mutations, L, will simply be maximized when W is maximized.

Given the parameters describing the growth rate and carrying capacity of the medium, we are thus able to find an optimal value of the dilution ratio, a value that minimizes the number of beneficial mutations that are lost during the bottleneck. For the parameters used in this example, a dilution ratio of $D \approx 0.15$ is clearly optimal. This result is interesting since more severe dilution ratios are often used in the literature. The parameters we used were chosen, once again, to correspond to the model parameters provided for a specific serial passaging regime for *E. coli* (LEVIN *et al.* 2000); in numerical exploration of the surrounding parameter space, we found this result to be surprisingly robust. We hypothesize that this value is largely determined by the underlying survival probability, $U(s)$, as approximated in the previous section. Further investigation of these intriguing results is clearly necessary, although beyond the scope of this article.

APPLICATIONS

We conclude our article with three examples, illustrating the possible application of our results and further predictions of the model.

Survival probability for a specific mutation: In some experimental protocols, the fate of a known mutation is of interest. An example here is the study of adaptive evolution in *E. coli*. LEVIN *et al.* (2000) compare rates of reversion and compensatory mutations for streptomycin-resistant *E. coli* evolving in the absence of the antibiotic (SCHRAG and PERROT 1996). One purpose of the study was to determine how often serial cultures will be dominated by fitness-compensated mutants, as opposed to reversion or resistant mutants. Through simulation studies, LEVIN *et al.* (2000) found that 56.8% of cultures are dominated by fitness-compensated mutants after 50

serial transfers, an important result that we can also derive analytically.

Given a beneficial mutation with a known selective advantage s that arises at rate μ per replication, we use Equation 4 to determine the probability that the mutation in question occurs during a single growth phase and ultimately survives bottlenecks. The fitness advantage of the compensatory mutant over the resistant mutant was determined by pairwise competition in experimental culture; this procedure gave an estimate of s as $\sim 0.92/0.8 - 1 \approx 0.15$. Using appropriate experimental parameters for resource-limited growth, we determined that the probability of occurrence and survival for a mutation with a similar fitness advantage is $\Lambda(s) \approx 0.0168$ per transfer. Thus, the probability that the mutation does not occur and survive in 50 transfers is $(1 - 0.0168)^{50}$, or 42.8%. This gives an analytical estimate that fitness-compensated mutants will dominate 57.2% of cultures after 50 transfers, in excellent agreement with the published results.

Estimating the mutation rate for beneficial mutations: The number of mutations that fix during experimental evolution may be used to estimate the fraction of all mutations that are beneficial. As an example, consider a recent study of "big-benefit" mutations in the adaptation of the bacteriophage ΦX174 to heat (Bull *et al.* 2000). As part of this study, "flask-adapted" phage were evolved from ancestor isolates; the evolving population was grown through 66 serial passages of 10-ml cultures with a dilution ratio of 10^{-4}. An isolate from the 66th passage differed from one ancestor phage at four nucleotide positions.

Bull *et al.* (2000) report that the 10,000-fold increase during each growth phase (from 10^5/ml to 10^9/ml) required 45 min initially, but this time dropped to 35 min over the course of the experiment. From these numbers, we find that $N_0 = 10^6$, $D = 10^{-4}$, the initial growth rate $r = 12.28$/hr, and the dilution time, τ, is $\sim$40 min. The growth rate after the 66th passage was 15.79/hr, and therefore the mean observed fitness increase per substitution is $\bar{s}_o = \ln(1.29)/4 \approx 0.063$ (Wahl and Krakauer 2000). As we have shown in a previous section, $\bar{s}_o = 2/\alpha$, and so we have as a rough estimate that $\alpha \approx 32$.

L denotes the expected number of mutations that occur during one growth phase and ultimately survive. If L is small, it gives the probability that a single, ultimately successful mutation occurs during one growth phase. The probability that four beneficial mutations occur during 66 passages and ultimately survive is therefore

$$P_4 = \binom{66}{4} L^4 (1 - L)^{62}.$$

By maximum likelihood, we find that four successful mutations are most likely to occur when $L \sim 0.061$. Substituting this value into Equation 5, we find that μ, the rate at which beneficial mutations occur per replication, is $\sim 1.44 \times 10^{-8}$. Since the overall mutation rate per site per replication is $\sim 10^{-6}$ in these phage, and the ΦX174 genome has 5386 bases, this result implies that $\sim$1 in 1 million mutations is beneficial during the "flask adaptation" of this phage.

Fitness gains as a function of bottleneck size: Burch and Chao (1999) studied the evolution of the RNA virus ϕ6, examining the effect of bottleneck size on the total number and size of adaptive "steps" taken during recovery from a deleterious mutation. We have replicated a similar set of experiments, assuming phage populations that expand from 1 to 8×10^9 phage in five generations and are then subject to seven bottleneck sizes (10, 33, 100, 333, 1000, 2500, and 10,000). Note that in our formalism, these bottleneck sizes correspond to variations in N_0; the dilution ratio is constant at $1/(8 \times 10^9)$.

For the results described below, we used $\alpha = 5$, consistent with estimates in another RNA phage (Wichman *et al.* 1999; Wahl and Krakauer 2000), and assumed a mutation rate to beneficial mutations of $\mu = 10^{-7}$. The latter value was estimated for a genome of 10^4 bases, a per base mutation rate of 10^{-5}, and a 1 in 1 million chance that a given mutation is beneficial. Our results are highly sensitive to these *ad hoc* parameter values and are intended as an illustration, not a quantitative prediction for ϕ6 evolution.

Using these parameters, we applied our model and numerically estimated the total number of mutations that are expected to occur during one growth phase and ultimately reach fixation. This result was multiplied by the total number of growth phases (20), to estimate the total number of adaptive steps during an evolutionary trajectory. As shown in the top of Figure 6, our model predicts that *the number of steps should increase with the size of the bottleneck.* (The two highest points on the graph have been truncated for clarity.) While this trend was not examined experimentally, our predictions agree well with the experimental results for bottlenecks <1000 (see Figure 3 in Burch and Chao 1999). Note, however, that Burch and Chao propagated each population either for 20 bottlenecks (100 generations), or until fitness was recovered to the original level. Thus for larger bottleneck sizes, propagation was discontinued after only 5 or 10 bottlenecks. When examined on a log-log plot (not shown), our data suggest that the number of adaptive steps increases exponentially with the logarithm of the bottleneck size.

We then multiplied the expected number of adaptive steps by the mean step size, $\log_{10}(1 + \bar{s})$, to obtain the expected total gain in fitness after 20 bottlenecks. In excellent agreement with experimental results, the fitness recovery was less than but approached one for bottleneck sizes <333. Thereafter, however, *our model predicts exponentially increasing gains in fitness with log(bottleneck size).* Once again, this was not tested experimen-

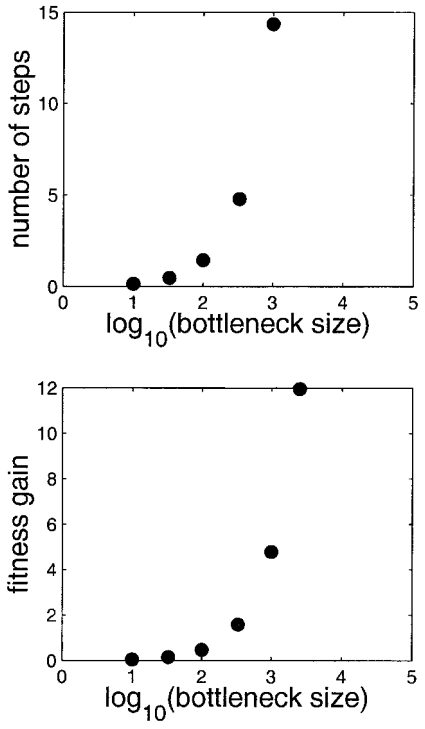

FIGURE 6.—Adaptive steps and fitness gain *vs.* bottleneck size. In the top, the total expected number of mutations that would occur and reach fixation in 100 generations (20 bottlenecks) is plotted against the size of the inoculum, using parameters that mimic experiments by BURCH and CHAO (1999). These values give the number of adaptive steps expected in experimental fitness trajectories. The bottom plots the total fitness gain expected in the same populations after 100 generations. For n expected adaptive steps with mean step size $\bar{s}$, the total fitness gain was calculated as $n \log_{10}(1 + \bar{s})$. Inoculum sizes, N_0, were 10, 33, 100, 333, 1000, 2500, and 10,000; populations grew for $T = 5$ generations at $r = 4.56$, corresponding to each phage expanding to 8×10^9 phage within a plaque. Other parameters were $D = 1/(8 \times 10^9)$, $\alpha = 5$, $\mu = 10^{-7}$. Three points were truncated from the graph for clarity: The numbers of steps were 36 and 144 at $N_0 = 2500$ and 10,000; fitness gain was 48 at 10,000.

tally because none of the populations >333 were propagated for 20 bottlenecks.

The major discrepancy between the predictions of our model and these experimental results involves one of the key findings of BURCH and CHAO (1999). As discussed previously, our model predicts that the mean selective advantage of surviving mutations, $\bar{s}$, is determined only by the underlying distribution of s and is given by $2/\alpha$. Thus the step size, or fitness difference, expected for the first adaptive sweep in a population should not depend on the bottleneck size. We find that the balance between (1) the low probability that a mutation has a large effect and (2) the high probability that such a mutation will fix is independent of population size. Our model, however, considers only the survival of the mutation with respect to bottlenecks; this discrepancy may be resolved when clonal interference, the competition between beneficial mutations arising at similar times in the population, is taken into account as well (GERRISH and LENSKI 1998).

DISCUSSION

For the bottlenecks modeled here, a large, randomly chosen fraction of the population is instantaneously eliminated at the end of τ generations. This is reminiscent of classic models of populations of fixed size, in which one-half of the offspring are eliminated, at random, after each generation. In fact, the population bottlenecks in serial passaging, as modeled here, are formally equivalent to many classic models of fixed population size: In serial passaging, however, the bottlenecks are *less* frequent (once every τ generations, rather than once per generation) and more severe ($D = 1/100$, for example, rather than $D = 1/2$). In experimental evolution, one could argue that the periods of sustained exponential growth between bottlenecks, not the bottlenecks themselves, are the most distinguishing feature of the dynamics.

With this in mind, many of the results worked out in the previous sections are not surprising. We find that the survival probability of a rare mutation is proportional to $2s$, and thus that the distribution of mutations that might be observed during experimental evolution has a mean that is twice the mean of the underlying distribution of possible mutational effects. We also find that using a more complex model that includes resource-limited growth has little effect on the fate of beneficial mutations.

Perhaps less intuitive, however, is the finding that successful mutations are equally likely to occur at all times during the growth phase. This is because the tendency for mutations to occur at late times is roughly balanced by the tendency for mutations to survive if they occur early in the growth phase.

Another intriguing result is our derivation of an optimal dilution ratio. When dilution occurs at $\sim D = e^{-2} \approx 0.135$, we find that the number of beneficial mutations lost during bottlenecks is minimized. In fact, at this optimal ratio, the total number of ultimately successful mutations, or the substitution rate, is *larger* than it would be in a constant population size experiencing genetic drift (see Figure 3). For populations held constant at the inoculum size, N_0, this effect is pronounced; the result still holds, however, if the population size is held constant at N_f. Thus although bottlenecks reduce fixation *probability* compared to constant populations, the overall fixation *rate* may be increased because of sustained periods of exponential growth between bottlenecks.

We solved for the optimal dilution ratio using first-order approximations and assuming s is small; for resource-limited growth we solved for the optimal D numerically for values of s between 0.01 and 0.1. The selective advantage of beneficial mutations in experimental evolution can be quite large, however; values as

high as 13.8 have been reported in the literature (Bull *et al.* 2000). To determine the extent to which this result holds for values of $s > 0.1$ or 0.2, we first need a derivation of the extinction probability that does not assume weak selection (Heffernan and Wahl 2002). The derivation of an accurate closed-form solution to the optimal dilution ratio for both strong and weak selection is thus a clear avenue for future work.

As a coherent mathematical framework describing experimental evolution becomes available (Wahl 2001), it is our hope that a large number of quantitative questions may be addressed. The applications section above provides a small sample of such questions. The development of this framework will allow researchers to take full advantage of the rich experimental data produced in this exciting field.

We are indebted to J. J. Bull and an anonymous reviewer for a number of helpful suggestions. This work was supported by the Natural Sciences and Engineering Research Council of Canada and by a National Science Foundation grant to R. E. Lenski in support of P. Gerrish.

LITERATURE CITED

Appenzeller, T., 1999 Test tube evolution catches time in a bottle. Science **284:** 2108–2110.

Bell, G., and X. Reboud, 1997 Experimental evolution in *Chlamydomonas* II. Genetic variation in strongly contrasted environments. Heredity **78:** 498–506.

Bull, J. J., M. R. Badgett, H. A. Wichman, J. P. Huelsenbeck, D. M. Hillis *et al.*, 1997 Exceptional convergent evolution in a virus. Genetics **147:** 1497–1507.

Bull, J. J., M. R. Badgett and H. A. Wichman, 2000 Big-benefit mutations in a bacteriophage inhibited with heat. Mol. Biol. Evol. **17:** 942–950.

Burch, C. L., and L. Chao, 1999 Evolution by small steps and rugged landscapes in the RNA virus ϕ6. Genetics **151:** 921–927.

Crow, J. F., 1994 "Foreward," p. xiii–xv in *Population Genetics, Molecular Evolution and the Neutral Theory: Selected Papers*, edited by M. Kimura. University of Chicago Press, Chicago.

Crow, J. F., and M. Kimura, 1970 *An Introduction to Population Genetics Theory.* Harper & Row, New York.

Edelstein-Keshet, L., 1988 *Mathematical Models in Biology.* McGraw-Hill, New York.

Ewens, W. J., 1967 The probability of survival of a new mutant in a fluctuating environment. Heredity **43:** 438–443.

Feller, W., 1971 *An Introduction to Probability and Its Applications*, Vol. II, pp. 268–280. John Wiley & Sons, New York.

Fisher, R. A., 1922 On the dominance ratio. Proc. R. Soc. Edinb. **50:** 204–219.

Gerrish, P. J., and R. E. Lenski, 1998 The fate of competing beneficial mutations in an asexual population. Genetica **102/103:** 127–144.

Haldane, J. B. S., 1927 The mathematical theory of natural and artificial selection. Proc. Camb. Philos. Soc. **23:** 838–844.

Heffernan, J. M., and L. M. Wahl, 2002 The effects of genetic drift in experimental evolution. Theor. Pop. Biol. (in press).

Kimura, M., 1957 Some problems of stochastic processes in genetics. Ann. Math. Stat. **28:** 882–901.

Kimura, M., 1962 On the probability of fixation of mutant genes in a population. Biometrics **19:** 1–17.

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

Lenski, R. E., M. R. Rose, S. C. Simpson and S. C. Tadler, 1991 Long-term experimental evolution in *Escherichia coli.* I. Adaptation and divergence during 2000 generations. Am. Nat. **138:** 1315–1341.

Levin, B. R., V. Perrot and N. Walker, 2000 Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics **154:** 985–997.

Otto, S. P., and M. C. Whitlock, 1997 The probability of fixation in populations of changing size. Genetics **146:** 723–733.

Papadopoulos, D., D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski *et al.*, 1999 Genomic evolution during a 10,000-generation experiment with bacteria. Proc. Natl. Acad. Sci. USA **96:** 3807–3812.

Pollak, E., 2000 Fixation probabilities when the population size undergoes cyclic fluctuations. Theor. Popul. Biol. **57:** 51–58.

Rainey, P. B., and M. Travisano, 1998 Adaptive radiation in a heterogenous environment. Nature **394:** 69–72.

Rosenzweig, R. F., R. R. Sharp, D. S. Treves and J. Adams, 1994 Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli.* Genetics **137:** 903–917.

Schrag, S., and V. Perrot, 1996 Reducing antibiotic resistance. Nature **381:** 120–121.

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.

Stewart, F. M., and B. R. Levin, 1973 Resource partitioning and the outcome of interspecific competition: a model and some general considerations. Am. Nat. **107:** 171–198.

Treves, D. S., S. Manning and J. Adams, 1998 Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli.* Mol. Biol. Evol. **15:** 789–797.

Wahl, L. M., 2001 Experimental evolution: analytical approaches and the need for a specific, integrated theory. Comm. Theor. Biol. **6:** 189–204.

Wahl, L. M., and P. J. Gerrish, 2001 Extinction probability in populations with periodic bottlenecks. Evolution **55**(12): 2606–2610.

Wahl, L. M., and D. C. Krakauer, 2000 Models of experimental evolution: the role of genetic chance and selective necessity. Genetics **156:** 1437–1448.

Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne and J. J. Bull, 1999 Different trajectories of parallel evolution during viral adaptation. Science **285:** 422–424.

Wright, S., 1931 Evolution in Mendelian populations. Genetics **16:** 97–159.

Communicating editor: M. W. Feldman