

Analytical Description of Mutational Effects in Competing Asexual Populations

Daniel Pinkel¹

Comprehensive Cancer Center and Department of Laboratory Medicine, University of California, San Francisco, California 94143

Manuscript received May 7, 2007
Accepted for publication October 1, 2007

ABSTRACT

The adaptation of a population to a new environment is a result of selection operating on a suite of stochastically occurring mutations. This article presents an analytical approach to understanding the population dynamics during adaptation, specifically addressing a system in which periods of growth are separated by selection in bottlenecks. The analysis derives simple expressions for the average properties of the evolving population, including a quantitative description of progressive narrowing of the range of selection coefficients of the predominant mutant cells and of the proportion of mutant cells as a function of time. A complete statistical description of the bottlenecks is also presented, leading to a description of the stochastic behavior of the population in terms of effective mutation times. The effective mutation times are related to the actual mutation times by calculable probability distributions, similar to the selection coefficients being highly restricted in their probable values. This analytical approach is used to model recently published experimental data from a bacterial coculture experiment, and the results are compared to those of a numerical model published in conjunction with the data. Finally, experimental designs that may improve measurements of fitness distributions are suggested.

THE adaptation of asexual populations to a new environment is the result of selection operating on the suite of stochastically occurring mutations, each of which may confer a different selective advantage. The mutations occur throughout time, so that multiple clones of mutant cells are present simultaneously. Mathematical modeling of the population dynamics typically employs numerical simulation to calculate a particular instance of the system, sampling probability distributions to include stochastic effects. The calculations may involve detailed tracking of each mutant clone through the history of the population. Running the model many times allows determination of its characteristic behavior as a function of the parameters describing mutation and selection. Estimates of the values of these parameters in a living system are obtained by comparisons of the statistical properties of the model with those of experimental data.

By contrast, this article presents an analytical description of the population dynamics. The analysis begins by establishing the identity of the average behavior of sequential finite cultures separated by bottlenecks with the growth of an exponentially expanding effectively infinite culture. Consideration of the infinite system provides analytical expressions for characteristic properties of the finite populations. The results include

quantitative descriptions of growth of the proportion of mutant cells with time and of the accompanying narrowing of the frequency distribution of their selection coefficients. Next, the stochastic behavior of finite systems is considered, resulting in a comprehensive and convenient description of the selection in the bottlenecks. The stochastic description is then used to develop a model of coculture experiments such as the one recently published by Hegreness and Shores (HS) (HEGRENESS *et al.* 2006). Application of the results obtained for the average behavior is used to simplify the model and facilitate its comparison with experimental data. Finally, the results are discussed and alternate experimental designs that may allow better measurement of the fitness distribution of the mutant cells are suggested. Significant additional information concerning the analysis is presented in the supplemental materials at <http://www.genetics.org/supplemental/>, including an Excel workbook for calculation of the statistical distributions that are developed in this article.

THE EXPERIMENTAL SYSTEM

Consideration of the experimental system of HS (Figure 1a) provides concrete motivation for the analysis presented in the remainder of this article. Approximately 2×10^5 cells were seeded into a culture, one-half labeled with yellow fluorescent protein (YFP) and one-half with cyan fluorescent protein (CFP). Otherwise the

¹Address for correspondence: University of California, Box 0808, San Francisco, CA 94143. E-mail: pinkel@cc.ucsf.edu

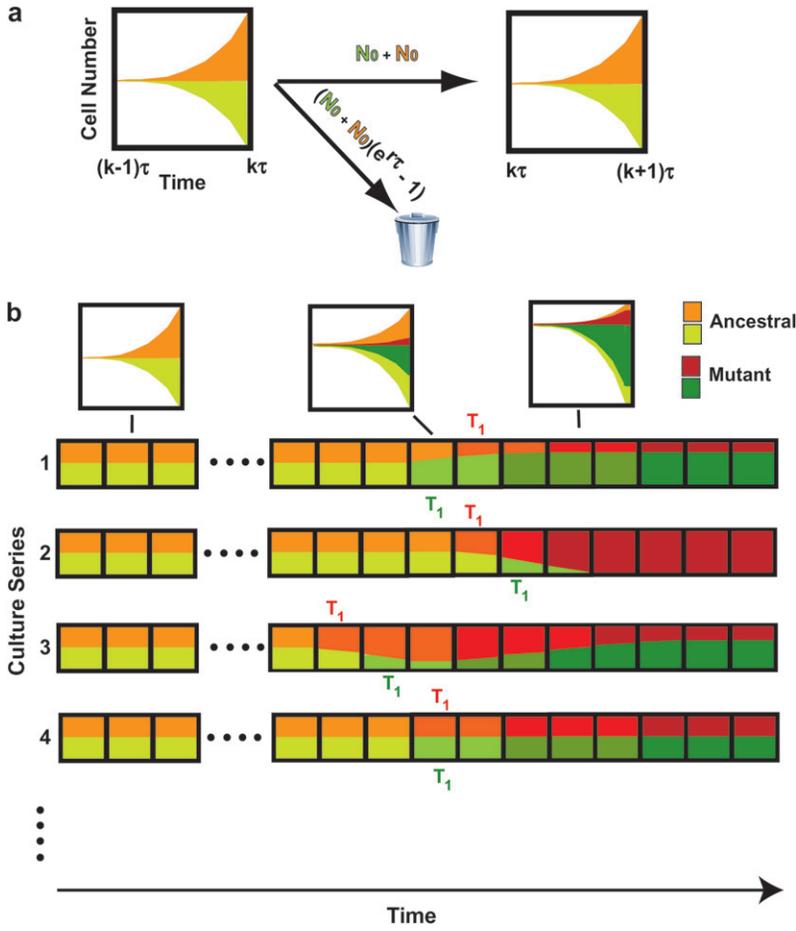


FIGURE 1.—Schematic of the YFP/CFP coculture experiment. (a) The exponential expansion of the YFP and CFP populations (distinguished by green and red colors) during the growth periods before and after the k th bottleneck at a time early in the experiment when mutant cells are present at low abundance. The cultures begin with N_0 of each type of cell and exponentially expand for time τ so that there are $N_0 e^{r\tau}$ of each. At the bottleneck N_0 of each type of cell is removed and transferred to the next culture. The remainder is discarded. In the actual experiment, $\tau = 11.7$ doublings, and the expansion factor is $e^{r\tau} \sim 3300$. (b) Four representative time courses for the behavior of the color ratios. Ancestral cells are indicated by light colors and mutants (of any selection coefficient) by dark colors. The colors are drawn separately for illustrative purposes, but in the actual system the cells are uniformly mixed. Schematic growth curves for the mutant and ancestral populations during three of the growth periods for the first series of cultures are shown in boxes above that series. The behavior of each culture series is described in the text. The red and green T_1 's indicate the times when mutant and ancestral cells are of equal abundance in the respective populations.

ancestral cells were identical. Mutations with selection coefficients s occurred stochastically in both populations of ancestral cells with a probability distribution $\rho(s)$ and an overall frequency μ per generation per cell. The cells grew exponentially for 24 hr, and the culture was then sampled to seed the next passage with $\sim 2 \times 10^5$ cells. After this bottleneck, exponential growth resumed, followed by sampling to seed the next daughter culture after another 24 hr, etc. The full experiment lasted ~ 40 days or ~ 450 doublings for the starting (ancestral) cells. Figure 1a illustrates the growth of the CFP and YFP populations just before and after the k th bottleneck for a time early in the series before the mutant populations have become significant. Each growth phase lasts a time $\tau = 11.7$ population doublings and results in an $e^{r\tau} \sim 3300$ -fold increase in cell number, where $r = \ln(2)$. At the bottleneck, a sampling of $e^{-r\tau} \sim 1/3300$ of the mutant and ancestral cells proceeded to the subsequent daughter culture. The rest of the cells were discarded.

YFP/CFP fluorescence ratios in each of 72 series of such cultures were measured at each bottleneck, providing a measure of the relative numbers of the two cell types as a function of time. Because of the stochastic nature of the mutational process and the passage of mutants through the bottlenecks, one expects that

under some conditions the ratio may change with time, differing in each series of cultures. Figure 1b shows examples of four of the possible time courses for the ratio, using green and red to distinguish the populations. The ancestral cells in both populations are represented as light colors, and the mutants (of any s) as dark colors. This distinction is made for illustrative purposes only since ancestral and mutant cells cannot be distinguished by their fluorescence. While the colors are illustrated as separated, in the actual experiment the two populations are thoroughly mixed.

In the first example, mutant cells are assumed to come to prominence initially in the “green” population. This time is indicated by a green T_1 , defined in our analysis as the time when the proportions of mutant and ancestral cells in the green population are equal. The overall growth rate of the green population detectably increases, and green cells begin to overgrow the “red” ones. Mutant cells with approximately the same s as those in the green population are assumed to come to prominence in the red population at a later time, indicated by the red T_1 . After this time, the green and red populations tend toward the same growth rate, and thus their ratio stabilizes. Mutant cells completely dominate both populations with time, as indicated by the darkening colors. This growth pattern corresponds to

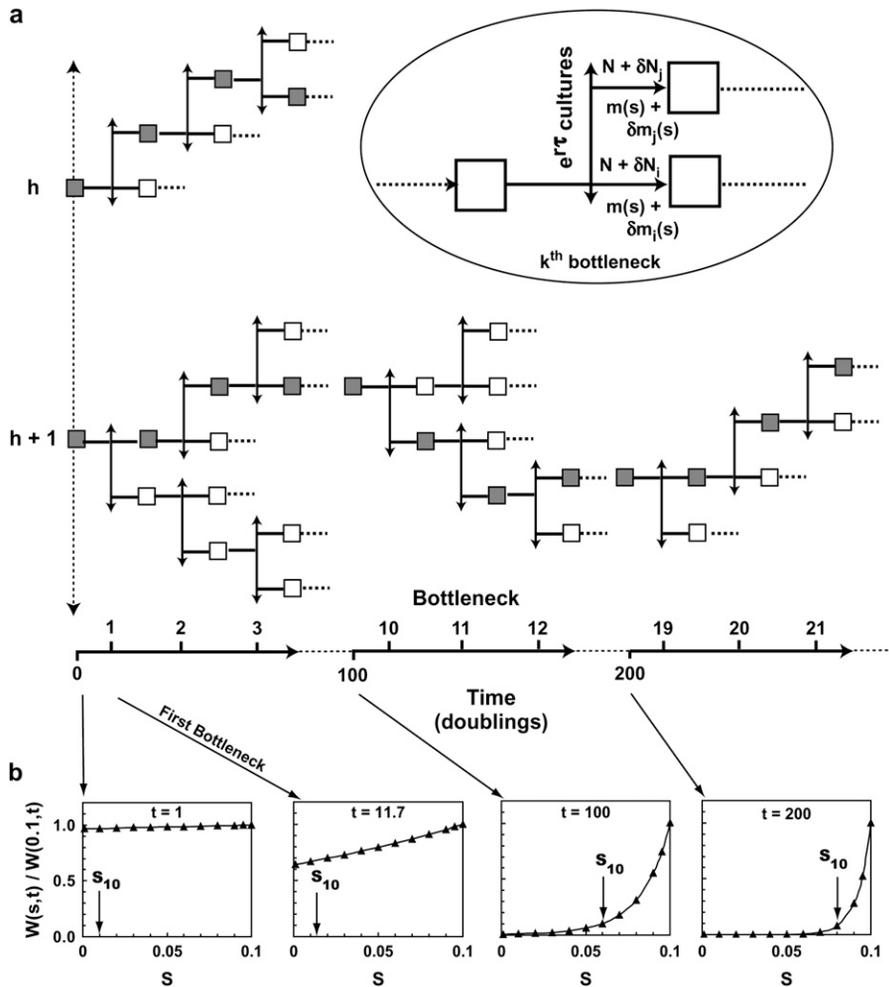


FIGURE 2.—The conceptual experiment. (a) An arbitrarily large number of initial cultures, each containing N_0 cells, are started at $t = 0$. Two such cultures, h and $h + 1$, are shown. After time τ , the time of the first bottleneck, the cells in each initial culture are divided among $e^{t\tau}$ daughter cultures so that each receives N_0 cells. Thus no cells are discarded. After each subsequent time interval τ , each daughter seeds $e^{t\tau}$ next-generation daughters. During each growth phase mutants received from the previous culture expand and *de novo* mutants randomly occur. At each bottleneck there is statistical variation in the number of mutant and ancestral cells received by the daughters. This process is indicated by the inset in the oval where $m(s)$ and $\delta m(s)$ indicate the average number and the fluctuation from the average of the mutants transmitted to the next daughter, and N and δN indicate the same for the ancestral cells. The time (in population doublings) and the number of the bottleneck is indicated along the bottom, assuming 11.7 doublings between bottlenecks. (b) The shape of $W(s, t)$, normalized to its value at $s = 0.1$, is shown as a function of s for representative values of t . As t increases the weight of this function rapidly moves to larger s . The distribution of mutant cells in the pooled culture of a is given by $\rho(s) W(s, t)$. For each of the possible paths through a series of daughters, the selection coefficients of the mutants will differ due to the stochastic variation, but the

powerful shape of $W(s, t)$ ensures that those that become prominent will be confined to a small range of s almost regardless of the shape of $\rho(s)$. If $\rho(s)$ is a constant out to some s_{\max} , then $W(s, t)$ directly gives the expected distribution of the mutants. The values of s , termed s_{10} , above which 90% of the mutant cells occur are indicated on the graphs for this case (see Equation 8).

curves 2 and 3 in Figure 6, a and b. Graphs above series 1 schematically show the growth of the mutant and ancestral populations during three cultures of this series.

In the second example, red mutants completely overtake the culture before green mutants of sufficiently large s come to prominence, although green mutations with low s will have occurred. This corresponds to curve 1 in Figure 6, a and b. In the third example, red mutants are assumed to come to prominence first and the fluorescence ratio begins to change in favor of red. At a later time, mutant cells with larger s come to prominence in the green population, and the ratio changes in favor of green. Still later, mutants that have a selection coefficient equivalent to the green population come to prominence in the red population, and the ratio stabilizes. The initial part of such behavior is shown by curve 5 in Figure 6b. Subsequent variations in the ratio may occur as mutants with ever larger s come to prominence in the two populations. But if $\rho(s)$ has a sufficiently defined maximum s , then the ratio will finally stabilize when both populations are dominated by

mutant cells with this maximal selection coefficient unless one population completely displaces the other as shown in example 2. In the final example, roughly equivalent mutants arise in both populations at about the same times, so that the ratio remains constant as both populations become dominated by mutant cells with selection coefficients tending toward the maximum possible. This is the behavior that always occurs if the size of the cultures is sufficiently large so that many mutations occur.

THE ANALYTICAL FRAMEWORK

Imagine an arbitrarily large ensemble of initial cultures, each starting with N_0 cells (Figure 2a). Only one cell population is shown for clarity. If additional populations are present as in the coculture experiment, their behaviors will be statistically independent. Each initial culture is sampled after growth time τ , but instead of seeding only one next passage culture as in the actual experiment, all of the material from each initial culture

is used to seed $e^{r\tau}$ daughter cultures (3300 in the case of the specific experiment in question). The details of this exhaustive sampling are illustrated in the oval inset in Figure 2a for the k th bottleneck. Each of the $(k + 1)$ st cultures receive $\sim N + \delta N$ ancestral and $m(s) + \delta m(s)$ mutant cells, where the δ 's indicate stochastic fluctuations that affect the individual cultures. The actual experiment is equivalent to selecting 72 of the initial cultures, and for each of these selecting a sequence of daughters, thereby producing 72 series with 40 sequential cultures in each. Two such series are indicated in Figure 2a by the shaded boxes. Examination of Figure 2a shows that the time dependence of the characteristics of the mutant cells averaged over a large number of series of daughters would be identical to the behavior of the total mutant population if all of the daughters were pooled. This pooled culture is just a population in unbounded exponential growth, which can be analyzed with straightforward approaches. Since the structure of this conceptual experiment preserves cell lineages, it also provides the basis for the subsequent calculation of the stochastic properties of the system.

This conceptual experiment is, of course, impossible to implement. Using the parameters of the actual experiment, by the 13th of the 40 days, the volume of culture medium required for the progeny of a single initial culture would fill a ball with a radius about 22 times that of our solar system out to Pluto, and the radius would be increasing faster than the speed of light. Coincidentally, this is about the characteristic time, $\overline{T}_1 \sim 150$ doublings, that mutants became equal in abundance to the ancestral cells in the actual HS experiment (see below).

DESCRIPTION OF THE AVERAGE CHARACTERISTICS

Basic relationships: Consider one of the cell populations in the infinite pooled culture of Figure 2a. Its development is described by standard equations for exponential growth. Let $N_\infty(t)$ and $M_\infty(s, t)$ be the number of ancestral and mutant cells at time t . Then

$$\dot{N}_\infty(t) = (1 - \mu)rN_\infty(t) \cong rN_\infty(t) \quad (1a)$$

$$\dot{M}_\infty(s, t) = \mu\rho(s)N_\infty(t) + r(1 + s)M_\infty(t), \quad (1b)$$

where the dots indicate the derivative with respect to time, μ is the overall mutation rate per generation for the ancestral cells, $\rho(s)$ is the probability distribution for the selection coefficients s of the mutants, and $r = \ln(2)$. Equation 1a describes the increase of ancestral cells due to division and the decrease due to conversion to mutants. Since $\mu \ll 1$, it is neglected in what follows. The coefficient r scales time so that it is measured in units of the doubling time for the ancestral cells. Equation 1b describes the increase in the number of

mutant cells with selection coefficient s through *de novo* mutation and the division of existing mutants with a rate $1 + s$ times that of the ancestral cells. In the real world, singly mutant cells are susceptible to additional mutations. The possibility of multiple mutations raises complex modeling issues that are discussed in section 2 of the supplemental materials at <http://www.genetics.org/supplemental/>. Multiple mutants are neglected in what follows.

This continuous growth model (with bottlenecks introduced below) does not include one important component of the actual experiment. In the real experiment the cultures enter a stationary phase where growth stops prior to seeding the daughter cultures. As mutants become prominent in the population and the overall growth rate increases somewhat, this stationary phase will be reached earlier during each passage. By contrast, the model allows expansion to continue during this stationary period. The error introduced by this simplification is small for the HS experiment. As is shown below, the maximum selection coefficient for the mutants in the experiment is ~ 0.1 . Thus, after cultures become dominated by mutants, in the time equivalent to 11.7 doublings of the ancestral cells the model allows about one additional doubling per passage ($r\tau \sim 1$). Therefore, during the late phases of the experiment the timescale in the model may be somewhat accelerated compared to that of the actual cultures. However, inclusion of this small effect in the analysis is not warranted given the noise in the experimental data with which it will be compared. The presence of the stationary phase raises additional interpretive issues, since as indicated in the DISCUSSION the selective advantage of mutants may be expressed by changes in their behavior as they cease and resume proliferation.

The solution of Equation 1a is $N_\infty(t) = N_\infty e^{rt}$, where N_∞ is the number of ancestral cells at $t = 0$. Inserting $N_\infty(t)$ into Equation 1b allows solving for $M_\infty(s, t)$. It is convenient for the subsequent discussion to calculate $R_m(s, t)$, the ratio of mutant cells with selection coefficient s to the total number of ancestral cells, because this describes how significant the mutant population has become:

$$R_m(s, t) = M_\infty(s, t)/N_\infty(t) \quad \text{or} \quad M_\infty(s, t) = N_\infty(t) R_m(s, t).$$

Inserting this into Equation 1b, one finds

$$R_m(s, t) = \mu\rho(s)(e^{rst} - 1)/rs = \mu\rho(s)W(s, t), \quad (2)$$

where

$$W(s, t) = (e^{rst} - 1)/rs.$$

It is immediately apparent that the distribution of selection coefficients found in the mutant cells at time t is given by the product of $\rho(s)$ with the weight factor $W(s, t)$.

At small values of st , $W(s, t) = t$, and $R_m(s, t) = \mu t\rho(s)$, indicating the buildup of *de novo* mutations linearly with time and with a distribution in s given by $\rho(s)$. As s and/or t increase, $W(s, t)$ increases dramatically due to the relative expansion of mutations that occurred early in the cultures. Figure 2b shows the shape of $W(s, t)$ for $0 < s < 0.1$ and times $t = 1, 11.7$ (the first bottleneck in the actual experiment), 100, and 200 population doublings. For plotting convenience these graphs have been normalized to the values of $W(s, t)$ at $s = 0.1$. As time progresses, the weight of this function becomes concentrated at higher values of s . Thus for almost any shape of $\rho(s)$ that one might choose, the width of the range of selection coefficients that are prominent in the mutant population will become narrower as time progresses. For most shapes of $\rho(s)$ the “effective” selection coefficients will fall within a narrow range near the maximum s that is possible. An alternate derivation of Equation 2, which has the flexibility to address more complex systems, is given in section 1 of the supplemental materials at <http://www.genetics.org/supplemental/>.

Given Equation 2, the ratio of the number of mutant cells with selection coefficients between 0 and some value of s to the total number of ancestral cells at time t , denoted by $P_m(s, t)$, can be calculated. Thus

$$P_m(s, t) = \int_0^s R_m(s', t) ds' = \mu \int_0^s \rho(s') W(s', t) ds'. \quad (3)$$

Integrating over the whole range of s gives $P_m(\infty, t)$, the ratio of the total number of mutant cells to the ancestral cells at time t . The fraction of mutant cells with selection coefficients between any values s_1 and s_2 is then given by

$$Q(s_1, s_2) = [P_m(s_2, t) - P_m(s_1, t)]/P_m(\infty, t). \quad (4)$$

These relationships allow determination of the average behavior of the system for various assumptions concerning μ and $\rho(s)$. The range of effective selection coefficients and the characteristic times T_1 and T_{100} , for which the number of mutant cells are respectively equal to, or 100 times greater than, the ancestral population are now compared for two specific choices $\rho(s)$.

Comparison of uniform and delta-function distributions for $\rho(s)$: Consider first a uniform distribution $\rho(s) = 1/s_{\max}$ for $0 \leq s \leq s_{\max}$ and 0 otherwise. The product $\rho(s) W(s, t)$ is $W(s, t)/s_{\max}$ for $0 \leq s \leq s_{\max}$ and so has the shape of $W(s, t)$ up to s_{\max} , whereupon it drops to 0. Figure 2b shows the shape of this function, indicating that as time increases the predominant mutants have selection coefficients progressively closer to s_{\max} . Applying Equations 3 and 4 allows calculation of the range of s of the effective mutations:

$$\begin{aligned} P_m(s, t) &= \mu \int_0^s \rho(s') W(s', t) ds' = \frac{\mu}{s_{\max}} \int_0^s \left(\frac{e^{rs't} - 1}{rs'} \right) ds' \\ &= \frac{\mu}{rs_{\max}} \int_0^{rst} \left(\frac{e^z - 1}{z} \right) dz = \frac{\mu}{rs_{\max}} I(rst). \end{aligned} \quad (5)$$

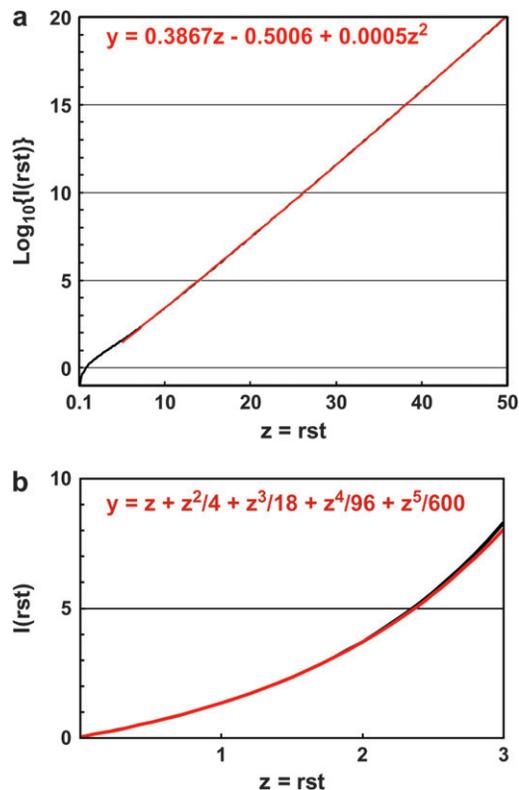


FIGURE 3.—Approximate evaluation of the integral $I(rst)$. (a) The black line shows the numerical evaluation of the integral in Equation 5 for $0.1 < rst < 50$. The red line, coincident with the black line, shows a second-order fit to the values for $5 < rst < 50$. The coefficients of the fit are sensitive to the exact range of the curve that is being fitted, but these are representative. (b) Evaluation of $I(rst)$ for $0 < rst < 3$ (black line). The result of integrating a Taylor series expansion of the integrand is shown in red. The formula for this curve is shown on the graph and used in section 4 of the supplemental materials at <http://www.genetics.org/supplemental/>.

The integral $I(rst)$ can be evaluated numerically in a straightforward manner. Figure 3a shows a graph of $\text{Log}_{10}[I(rst)]$ along with regression fit for the region $5 < rst < 50$. Using this fit as an approximation gives:

$$\begin{aligned} \text{Log}_{10}[P_m(s, t)] &= \text{Log}_{10}[\mu/(rs_{\max})] + \text{Log}_{10}[I(rst)] \\ &\approx \text{Log}_{10}[\mu/(rs_{\max})] + 0.39rst - 0.50 \\ &\quad + 0.0005(rs_{\max})^2 + \dots \\ &\approx \text{Log}_{10}[\mu/rs_{\max}] + 0.27st - 0.34 \end{aligned} \quad (6)$$

or equivalently

$$P_m(s, t) \approx \frac{0.32\mu e^{0.9rst}}{rs_{\max}}$$

The use of the linear approximation for the fit to $\text{Log}_{10}[I(rst)]$ is particularly accurate for $7 < rst < 16$, for which the deviation of the approximate value of $\text{Log}_{10}[P_m(s, t)]$ from the true value is < 0.1 . This corresponds to times on the order of ~ 100 to ~ 230 generations for $s \sim 0.1$. Section 4 of the supplemental materials

at <http://www.genetics.org/supplemental/> and Figure 3b present an approximation to Equation 5 and $I(rst)$ for low values of rst .

The range of s containing 90% of the mutant cells is arbitrarily defined as the range of effective selection coefficients. Since $P_m(s, t)$ is monotonically increasing, the effective range can be obtained by determining the selection coefficient, s_{10} , for which 10% of mutants have lower s . Setting $Q(0, s_{10}) = 0.1$ in Equation 4 yields s_{10} . Since $P_m(0, t) = 0$, one finds

$$\begin{aligned} \text{Log}_{10}Q(0, s_{10}) &= -1 \\ &= \text{Log}_{10}(P_m(s_{10}, t) - \text{Log}_{10}(P_m(s_{\max}, t))) \\ &= 0.27s_{10}t - 0.27s_{\max}t. \end{aligned} \quad (7)$$

Thus within the range of validity of the approximation for $\text{Log}_{10}[I(rst)]$,

$$s_{\max} - s_{10} = \Delta s = \frac{1}{0.27t} \quad \text{or} \quad \frac{\Delta s}{s_{\max}} = \frac{1}{0.27s_{\max}t}. \quad (8)$$

Thus if $s_{\max} = 0.1$, at $t = 100$ doublings, 90% of the mutant cells have s between 0.063 and 0.1, while by $t = 200$ the same proportion will be between 0.083 and 0.1. The values of s_{10} are indicated on the graphs in Figure 2b.

The highly peaked distribution in s of the mutant population at long times even though $\rho(s)$ was assumed to be flat suggests that some aspects of the behavior of cultures of these cells would be similar to a system in which one assumed a delta-function distribution of selection coefficients. Using $\rho(s) = \delta(s - \chi)$ and a mutation rate of μ_δ in Equation 3, one finds analogous to equation 5,

$$P_m(\infty, t) = \frac{\mu_\delta(e^{r\chi t} - 1)}{r\chi} \cong \frac{\mu_\delta e^{r\chi t}}{r\chi} \quad (9)$$

or

$$\begin{aligned} \text{Log}_{10}[P_m(\infty, t)] &= \text{Log}_{10}[\mu_\delta(e^{r\chi t} - 1)/r\chi] \\ &\approx \text{Log}_{10}(\mu_\delta/\chi) + 0.30\chi t + 0.16 \\ &\quad \text{for } r\chi t \gg 1. \end{aligned}$$

This result for the delta function is identical in form to the (approximate) result for the uniform distribution of $\rho(s)$ found in Equation 6. Evaluating Equation 6 at $s = s_{\max}$ so that the entire mutant population is represented, Equations 4 and 9 are quantitatively identical if one sets $\chi = 0.9s_{\max}$ and $\mu_\delta = 0.29\mu$. Thus after the initial period where they are clearly distinct, the total number of mutant cells evolves with approximately the same time course for both of these forms for $\rho(s)$. The delta-function approximation is appropriate for any $\rho(s)$ that results in a highly peaked shape for $\rho(s)W(s, t)$ at long times. The scaling of the overall mutation rates and the “equivalent” selection coefficients will depend on the details of the shape of $\rho(s)$ and may also depend on which aspects of the cultures are being modeled. Some critical aspects of real systems cannot be modeled

by the delta-function approximation, as shown in Figure 6 and the related text.

Characteristic times: There are at least two times in the history of the culture that are interesting to calculate. The most relevant is T_1 , the time when the numbers of mutant and ancestral cells in the culture are equal. While this differs in each series of finite cultures as shown in Figure 1b, it has a well-defined value for the effectively infinite system of Figure 2a. This characteristic time, \overline{T}_1 , is the time when $P_m(\infty, \overline{T}_1) = 1$, which can be determined from Equation 3 for any assumed $\rho(s)$. For the specific case of the uniform fitness distribution, $P_m(\infty, \overline{T}_1) = P_m(s_{\max}, \overline{T}_1) = 1$ or $\text{Log}_{10}[P_m(s_{\max}, \overline{T}_1)] = 0$. Similarly, for a delta-function distribution, $\text{Log}_{10}[P_m(\infty, \overline{T}_{1\delta})] = 0$. Using Equations 6 and 9 one finds

$$\begin{aligned} \overline{T}_1 &= (0.34 - \text{Log}_{10}[\mu/s_{\max}])/(0.27s_{\max}) \quad (\text{uniform distribution}) \\ \overline{T}_{1\delta} &= (-0.16 - \text{Log}_{10}[\mu_\delta/\chi])/(0.30\chi) \quad (\text{delta-function distribution}). \end{aligned} \quad (10)$$

The second relevant time is the time for “fixation” of the mutations, which following HS is defined as the time when mutants are 100 times the abundance of ancestral cells. At this time for the uniform distribution $P_m(s_{\max}, \overline{T}_{100}) = 100$, or $\text{Log}_{10}[P_m(s_{\max}, \overline{T}_{100})] = 2$, while similarly for the delta-function distribution $\text{Log}_{10}[P_m(\infty, \overline{T}_{100\delta})] = 2$. Thus from Equations 6 and 9,

$$\begin{aligned} \overline{T}_{100} &= (2.34 - \text{Log}_{10}[\mu/s_{\max}])/(0.27s_{\max}) \quad (\text{uniform distribution}) \\ \overline{T}_{100\delta} &= (1.84 - \text{Log}_{10}[\mu_\delta/\chi])/(0.30\chi) \quad (\text{delta-function distribution}). \end{aligned} \quad (11)$$

The range of selection coefficients that are prominent in the population at these times for the uniform distribution can be calculated using Equation 8. Using $\mu = 10^{-5}$ and $s_{\max} \sim 0.12$, values used by HS for their simulation of a uniform $\rho(s)$ described in their Figure 1B, Equation 10 yields $\overline{T}_1 = 136$ generations. At that time 90% of the mutant cells are in the range $\Delta s/s_{\max} \sim 0.23$, or $0.089 < s < 0.12$. Similarly, from Equation 11 $\overline{T}_{100} = 198$ generations, and at that time $\Delta s/s_{\max} = 0.16$, or $0.10 < s < 0.12$. Thus at these times 90% of the mutant cells have selection coefficients very near the maximum possible.

STOCHASTIC EFFECTS

Differences in behavior among different series of cultures are due to stochastic variation in the times t_{m_i} new mutants arise, the corresponding selection coefficients s_i of the mutants, and the effects of the selection bottlenecks. This section demonstrates that the stochastic effect of the multiple bottlenecks is equivalent to altering the actual mutation times t_{m_i} to effective times $t_{m_i}^*$, and derives probability distributions for $t_{m_i}^*$, given t_{m_i} , s , and the length of the growth periods, τ . This formalism, coupled with statistical sampling of the mutation times on the basis of the mutation rate and sampling of the selection coefficients on the basis of $\rho(s)$,

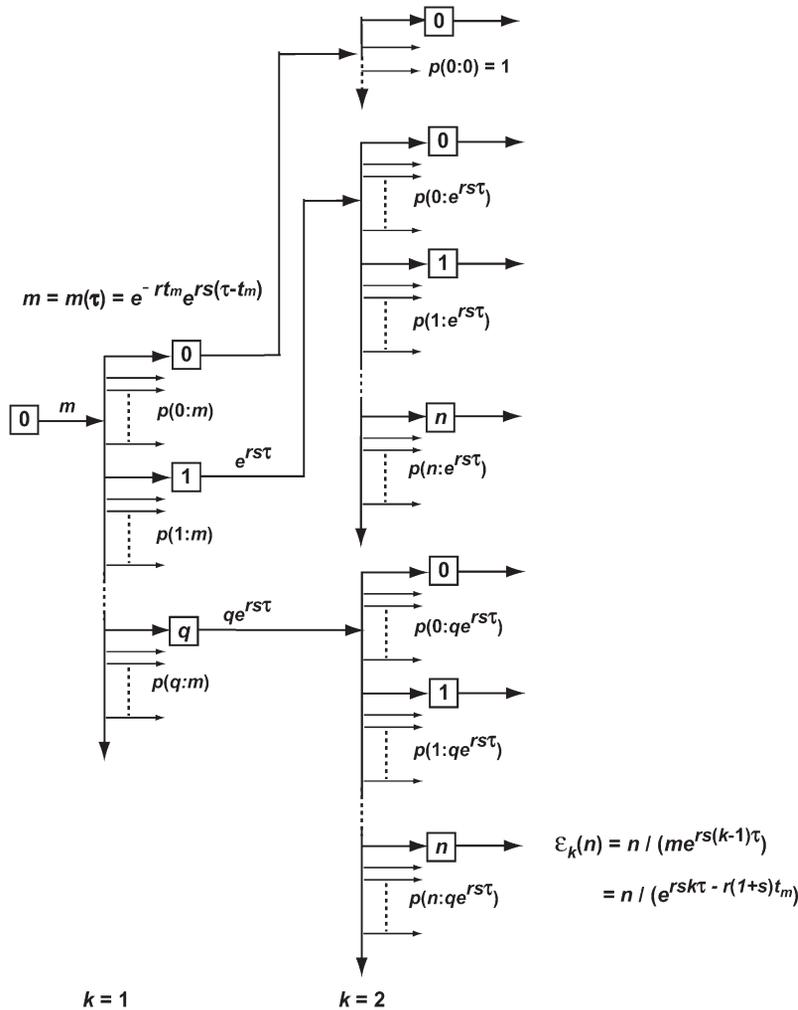


FIGURE 4.—Schematic of the stochastic transmission of the progeny of a single mutation through two bottlenecks. The boxes indicate individual cultures that are related to each other according to the exhaustive sampling scheme described in Figure 2a. A mutation is assumed to occur at some time during the initial culture period and on average m mutant progeny are passed to each of its $e^{r\tau}$ daughters. The daughters are arranged in order of the number of mutant cells they receive, which is indicated within the boxes. The small arrows indicate the additional daughter cultures that received the same number of mutant cells. The proportion of the first-generation daughters receiving an actual number, q , is given by a suitable probability distribution such as Poisson. Between the first and the second bottlenecks these mutants expand relative to the ancestral cells by a factor $e^{rs\tau}$. These will again be probabilistically distributed to their daughters. From the second bottleneck onward, daughters with a particular number of cells, n , descend from multiple predecessor cultures. The probability of having n mutants after bottleneck k , $\Gamma_k^*(n)$, is given by summing over these possibilities, as shown in Equation 13. The ratio of the actual number of mutant cells in an individual culture after the k th bottleneck to the average number for that time is ε_k (Equation 14).

allows a complete description of the stochastic behavior of the system. The analysis proceeds by first calculating the statistics of the behavior of a single-mutant clone with a defined mutation time and selection coefficient and then combining multiple mutants to describe the system completely.

Stochastic behavior of a single-mutant clone: Figure 4 shows the details of the behavior of the progeny of a mutant cell expanding through a series of daughter cultures such as described in the conceptual experiment Figure 2a. No cells are discarded. Assume that the mutation occurred at time $0 \leq t_m \leq \tau$, where τ is the length of each of the culture periods. At the k th bottleneck, $t = k\tau$, and the average number of mutants, $m(k\tau)$, transferred to each daughter is the total number of progeny from this mutation divided by the total number of daughter cultures:

$$m(k\tau) = \frac{e^{r(1+s)(k\tau-t_m)}}{e^{rk\tau}} = e^{rsk\tau - r(1+s)t_m} = me^{rs(k-1)\tau}$$

$$m = e^{-rt_m} e^{rs(\tau-t_m)}. \tag{12}$$

m is the average number of mutant progeny that are transferred to each daughter at the first bottleneck.

Due to statistical fluctuations in the bottlenecks, the actual number of cells each daughter receives is distributed around this average. As illustrated in Figure 4, after the second and subsequent bottlenecks, daughters with the same number of mutants can descend from different predecessors, so that the probability, $\Gamma_k^*(n)$, of a daughter receiving n mutant cells after the k th bottleneck requires summing over these multiple possibilities. Suppose that q mutants were transferred into a daughter after the $(k - 1)$ st bottleneck. At the next bottleneck, the k th, these have expanded so that on average $qe^{rs\tau}$ mutant cells will be distributed to each of its $e^{r\tau}$ daughters. The statistical distribution in the number n received by these daughters is given by a probability distribution $p(n:qe^{rs\tau})$. But the probability of having a culture with q initial cells is given by $\Gamma_{k-1}^*(q)$. Thus,

$$\Gamma_k^*(n) = \sum_{q=0}^{\infty} p(n : qe^{rs\tau}) \Gamma_{k-1}^*(q)$$

$$= \sum_{q=0}^{\infty} \frac{(q\lambda)^n e^{-q\lambda}}{n!} \Gamma_{k-1}^*(q), \tag{13}$$

where $\lambda = e^{rs\tau}$, $k \geq 2$, and the Poisson probability distribution has been used because it is appropriate,

at least initially, when the number of mutant cells is substantially smaller than the total number of cells. As shown below, this condition is met for the entire time period during which stochastic variation is important. As can be seen from examination of Figure 4, $\Gamma_1^*(n) = p(n:m(\tau))$, where from equation 12, $m(\tau) = e^{r\tau - r(1+s)t_m} = \lambda e^{-r(1+s)t_m}$. Evaluation of Equation 13 proceeds by using $\Gamma_1^*(n)$ to calculate $\Gamma_2^*(n)$, etc. The supplemental materials at <http://www.genetics.org/supplemental/> contains an Excel workbook that performs this calculation, as well as the calculations of the related statistical distributions discussed below. Note that $\Gamma_k^*(n)$ depends on τ , s , and t_m through m and λ .

The basic behavior of Equation 13 is easily understood. Beginning with a Poisson distribution after the first bottleneck, it progressively broadens after subsequent bottlenecks. Figure 5a shows plots of $\Gamma_k^*(n)$ for $k = 1-7$ when $t_m = 1$ and $s = 0.09$. Note that $\Gamma_k^*(n)$ varies smoothly and progressively more slowly as k increases, except between $n = 0$ and 1 where there is a pronounced discontinuity that increases with increasing k . Initially the broadening of $\Gamma_k^*(n)$ is due to the combination of stochastic events in the bottlenecks and the expansion of the mutant clone. As the number of mutant cells increases with time, stochastic variation in the number transferred from a particular culture to its immediate daughters eventually becomes insignificant compared to the mean number transferred since $\Delta n/n \sim n^{-1/2}$. Thus it is expected that the important stochastic variation induced by the bottlenecks occurs early in the experiment. This expectation can be quantitatively described in the following manner.

Let $\varepsilon_k(n)$ be the ratio of the actual number of mutant cells n in a particular culture to the average number (Equation 12) after the k th bottleneck. Then

$$n = \varepsilon_k(n)m(k\tau) = \varepsilon_k(n)e^{rsk\tau - r(1+s)t_m}. \quad (14)$$

The values of $\varepsilon_k(n)$ for daughters receiving different numbers of mutants after the second bottleneck are illustrated in Figure 4. The average value of $\varepsilon_k(n) = 1$. Although n is rigorously an integer so that only specific values of $\varepsilon_k(n)$ can occur, the slow variation of $\Gamma_k^*(n)$ with n , Figure 5a, allows an accurate description of the system to be obtained by treating n as a continuous variable and defining $\Gamma_k(n)$ as a continuous probability density that has the values $\Gamma_k^*(n)$ for integer values of n . For noninteger n , $\Gamma_k(n)$ can be obtained by interpolation. This approximation is valid for $n \geq 1$, but not for $n = 0$ due to the substantial discontinuity in $\Gamma_k^*(n)$ between $n = 0$ and 1. Thus in the calculations that follow, probabilities corresponding to $n = 0$ will be given by $\Gamma_k^*(0)$, while those corresponding to other values of n will be based on calculations that treat parameters as continuous. Using this approximation, which is increasingly accurate as k increases, the probability distribution,

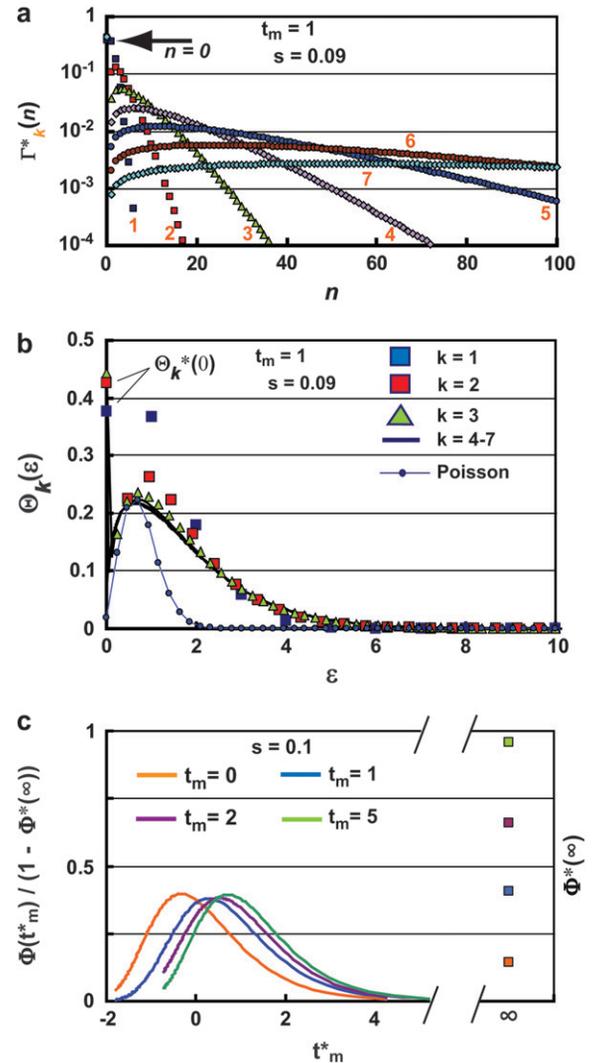


FIGURE 5.—Statistical distributions for n , ε , and the effective mutation time t_m^* . (a) The probability $\Gamma_k^*(n)$ of having n progeny of a mutant after k bottlenecks for $s = 0.09$, $\tau = 11.7$, and $t_m = 1$. The value of k for each curve is shown by the orange numbers. Note that the distribution broadens to higher values of n with increasing k , and its maximum magnitude for $n > 0$ correspondingly decreases. For the chosen parameters, the probability that a culture will have 0 progeny of the mutant (indicated by the arrow) is basically established at the first bottleneck and subsequently increases only slightly. (b) Probability distribution for ε_k after each bottleneck. Distributions are shown for $k = 1-7$, for $t_m = 1$, $\tau = 11.7$, and $s = 0.09$. Symbols corresponding to each n are shown for the first three bottlenecks; for larger k they are so closely spaced that solid curves are shown. Note that differences become imperceptible for $k > 4$ and that the width is significantly broader than a Poisson distribution, which is shown in blue for comparison. The probability of a culture receiving 0 cells hardly increases after the first bottleneck for these parameters. These values are given by $\Theta_k^*(0) = \Gamma_k^*(0)$. (c) Dependence of the probability distribution $\Phi(t_m^*)$ of the effective mutation times, t_m^* , on the actual mutation time, t_m , for $\tau = 11.7$ and $s = 0.1$. As a mutation occurs at later times, the probability that a daughter culture will receive 0 progeny increases, as shown by the boxes at $t_m^* = \infty$. The curves for finite t_m are plotted after normalization by $1 - \Phi^*(\infty)$, where $\Phi^*(\infty)$ is the probability of having $n = 0$ progeny of the mutation. Note that the curves move to slightly higher values of t_m^* as the mutation time increases, but still overlap 0.

$\Theta_k(\varepsilon_k)$, for ε_k , after each bottleneck can be calculated using Equations 12–14,

$$\Gamma_k(n)dn = \Gamma_k(n)\frac{dn}{d\varepsilon_k}d\varepsilon_k \equiv \Theta_k(\varepsilon_k)d\varepsilon_k \Rightarrow \Theta_k(\varepsilon_k) = \Gamma_k(n(\varepsilon_k))m(k\tau), \tag{15}$$

where $m(k\tau)$ is given by Equation 12. Equation 15 holds for $\varepsilon_k > 0$. The probability of having $\varepsilon_k = 0$, which corresponds to $n = 0$, is $\Theta_k^*(0) = \Gamma_k^*(0)$.

Figure 5b shows behavior of $\Theta_k(\varepsilon_k)$ and $\Theta_k^*(0)$ for $k = 1-7$, $s = 0.09$, $\tau = 11.7$, and $t_m = 1$. (The distributions for other parameter values can be calculated using the Excel sheet in the supplemental materials at <http://www.genetics.org/supplemental/>.) Note that the shape changes substantially for the first few bottlenecks, but for $k \geq 4$ it becomes constant. This is the quantitative description of the previous statement that once the number of progeny of the mutated cell becomes large, no significant additional variation is introduced by the subsequent bottlenecks. Thus a single distribution, $\Theta(\varepsilon)$, calculable from first principles, can be used to describe the behavior of the mutants after sufficient time. The distribution for each mutant clone depends on the actual time the mutation occurred, the selection coefficient, and the length of the growth periods. In this example calculation, the stable distribution is reached by the analysis of cultures containing up to 100 mutant cells, 1000-fold less than the number of ancestral cells present in the actual experiment. Thus statistical stability is reached well before the time T_1 . For comparison, Figure 5b also shows a Poisson distribution scaled so that its maximum is located at the same position as the maximum of the limiting mutant distribution. The enhanced broadening due to the series of bottlenecks is evident.

The fact that a stable, readily calculable probability distribution can be used to describe the progeny of each mutation after several bottlenecks allows a particularly simple general description of the stochastic behavior of the entire system. Consider the ratio, $\overline{R}(t)$, of the average number of mutant cells to the number of ancestral cells per culture:

$$\overline{R}(t) = \frac{m(t)}{N_0 e^{rt}} = \frac{e^{r(1+s)(t-t_m)}}{N_0} = \frac{e^{rst-r(1+s)t_m}}{N_0}. \tag{16}$$

The actual ratio in the daughter cultures, $R(t)$, will differ from this ratio by the factor ε that has developed due to stochastic variation in the earlier bottlenecks. Therefore, for times long enough after the mutation for statistical stability to be established one has

$$R(t) = \varepsilon \overline{R}(t). \tag{17}$$

Therefore,

$$R(t) = \varepsilon \frac{e^{rst-r(1+s)t_m}}{N_0} = \frac{e^{rst-r(1+s)t_m+\ln\varepsilon}}{N_0} = \frac{e^{rs(t-t_m^*)}}{N_0 e^{rt_m^*}}, \tag{18}$$

where $t_m^* = t_m + \Delta t_m$ is the effective time the mutation occurred, adjusted from the actual t_m by $\Delta t_m = -\ln(\varepsilon)/r(1+s)$ due to the stochastic effects in the bottlenecks, and $N_0 e^{rt_m^*}$ is the size of the ancestral population at the effective mutation time. Negative values of Δt_m indicate a stochastic fluctuation that increases the abundance of a mutant clone relative to its average value; *e.g.*, it appears that the mutation occurred earlier than it actually did, while positive values indicate a stochastic decrease relative to the average since the effective occurrence time was later than the actual time. If a mutant is lost from a series of cultures, $\Delta t_m = \infty$. Note that Equation 18 assumes that $0 \leq t_m \leq \tau$.

Calculation of the probability distribution, $\Phi(t_m^*)$, for t_m^* allows a statistical description of the behavior of the mutant cells. Using the continuous variable approximation discussed prior to Equation 15,

$$\Gamma_k(n)dn = \Gamma_k(n)\frac{dn}{dt_m^*}dt_m^* \equiv \Phi_k(t_m^*)dt_m^* \Rightarrow \Phi_k(t_m^*) = \Gamma_k(n)nr(1+s) \tag{19a}$$

or

$$\Phi_k(t_m^*) = \Gamma_k(n)\frac{n}{m(k\tau)}m(k\tau)r(1+s) = \Theta_k(\varepsilon_k)\varepsilon_k r(1+s) \rightarrow \Theta(\varepsilon)\varepsilon r(1+s), \tag{19b}$$

where the relationships among the various parameters required to evaluate Equation 19 are $t_m^* = t_m - \ln(\varepsilon)/r(1+s) = t_m - \ln(n/m(k\tau))/r(1+s)$, and Equation 15 was used to relate $\Phi_k(t_m^*)$ to $\Theta_k(\varepsilon_k)$. Equation 19b demonstrates that $\Phi_k(t_m^*)$ becomes independent of k after several bottlenecks since $\Theta_k(\varepsilon_k)$ becomes independent of k . As a practical matter, $\Phi_k(t_m^*)$ is calculated by choosing n and $m(k\tau)$ for any bottleneck after stabilization of the distributions, with $m(k\tau)$ given by Equation 12. Equation 19 is valid for finite t_m^* . For $t_m^* = \infty$, which corresponds to having $n = 0$ progeny of the mutant, $\Phi^*(\infty) = \Gamma_k^*(0)$. The stabilization of the distribution for t_m^* after several bottlenecks means that the t_m^* of a mutation in a particular daughter culture remains the same for all of its descendant cultures. Thus the effective mutation times provide a suitable basis for describing the stochastic character of the long-term evolution of the system.

Figure 5c shows the behavior of $\Phi(t_m^*)$ and $\Phi^*(\infty)$ for $t_m = 0, 1, 2$, and 5 , with $s = 0.1$ and $\tau = 11.7$. Given the assumed parameters, if a mutation actually occurs at $t_m = 0$, then on average ~ 2.25 mutant cells from this clone will be transferred to each daughter at the first bottleneck, so that most will receive at least one cell. For larger t_m , the proportion of daughter lineages that receive no progeny of the mutant (*e.g.*, have $t_m^* = \infty$) increases, correspondingly reducing the magnitude of

$\Phi(t_m^*)$ for finite t_m^* . Therefore, to conveniently visualize the behavior of $\Phi(t_m^*)$ for different t_m in one figure, $\Phi(t_m^*)/(1 - \Phi^*(\infty))$ has been plotted. Note that the distributions of effective mutation times for those daughter cultures that have at least one cell with this mutation include $t_m^* = 0$ and have widths of several doubling times. Thus if the progeny of a mutation are not lost in a series of cultures, they behave as if their founding mutation occurred near the beginning of the culture period in which they originated. As t_m increases the distributions move somewhat toward higher values t_m^* , but eventually stabilize except for a decrease in overall magnitude that has been normalized in this plot. The stabilization comes about since as t_m increases eventually all daughter cultures receive either 0 or 1 cell at the first bottleneck, and those that receive a cell subsequently develop with similar statistical behavior.

Description of multiple mutations: Equations 18 and 19 allow a complete description of the multiple mutations that occur during growth in the cultures. The ratio of the total number of mutant to ancestral cells in daughter culture, $\hat{P}_m(t)$, is obtained by summing over all mutations. Using Equation 18,

$$\hat{P}_m(t) = \sum_{k=0}^K \sum_{i=1}^{I_k} \frac{e^{rs_{ki}(t-t_{m_{ki}}^*)}}{N_0 e^{r(t_{m_{ki}}^* - k\tau)}} \quad (20a)$$

$$\hat{P}_m(t) = \frac{1}{N_0} \sum_{k=0}^K \sum_{i=1}^{I_k} \frac{e^{-rs_{ki}k\tau} e^{rs_{ki}(t-t_{m_{ki}}^*)}}{e^{r(\hat{t}_{m_{ki}}^*)}}, \quad (20b)$$

where the k th mutation occurs during the growth period after the k th bottleneck at time $t_{m_{ki}}$ and with selection coefficient s_{ki} and effective mutation time of $t_{m_{ki}}^*$, and the times in the denominator of Equation 20a are adjusted to account for the fact that after each bottleneck the culture returns to having N_0 ancestral cells. To simplify interpretation of Equation 20a and calculation of the effective mutation times, let $\hat{t}_{m_{ki}} = t_{m_{ki}} - k\tau$ and correspondingly $\hat{t}_{m_{ki}}^* = t_{m_{ki}}^* - k\tau$, where $\hat{t}_{m_{ki}}$ is the actual mutation time and $\hat{t}_{m_{ki}}^*$ is the effective mutation time measured relative to the time of the k th bottleneck. Thus the distributions for the $\hat{t}_{m_{ki}}^*$ can be calculated using Equation 19, employing the adjusted mutation times $\hat{t}_{m_{ki}}$.

The evaluation of Equation 20 formally requires random sampling from the mutation rate distribution times the instantaneous population of ancestral cells to obtain the mutation times, assigning a selection coefficient to each mutation by sampling from the distribution $\rho(s)$, and finally assigning the corresponding effective mutation times by sampling from the appropriate $\Phi(t_m^*)$ calculated from Equation 19. However, many aspects of its general behavior can be understood much more simply, as discussed in section 3 of the supplemental materials at <http://www.genetics.org/supplemental/>.

Note that $\hat{P}_m(t)$ is the finite-system analog of the previously derived $P_m(\infty, t)$ from Equation 3. $P_m(\infty, t)$ is the average value of Equation 20 over multiple finite cultures.

ANALYSIS OF THE YFP/CFP COMPETITION EXPERIMENT

Calculation of the fluorescence ratio: Equation 20 allows a full description of the stochastic behavior of the YFP/CFP cocultures. In the coculture the differentially labeled populations grow independently and are independently subject to the statistics of mutation formation and bottlenecks. Thus, if N_Y and N_C are the total numbers of YFP and CFP cells, respectively, and N_{YA} and N_{CA} are the corresponding numbers of ancestral cells, then

$$\begin{aligned} N_Y &= N_{YA}[1 + \hat{P}_{mY}(t)] \\ N_C &= N_{CA}[1 + \hat{P}_{mC}(t)] \\ \text{Log}_{10}(N_Y/N_C) &= \text{Log}_{10} \left[\frac{1 + \hat{P}_{mY}(t)}{1 + \hat{P}_{mC}(t)} \right] + \text{Log}_{10}(N_{YA}/N_{CA}). \end{aligned} \quad (21)$$

At the beginning of an experiment $\text{Log}_{10}(N_Y/N_C)$ remains constant (equal to 0 if the initial population sizes are equal) until a sufficient number of mutant cells arise such that $\hat{P}_{mY}(t)$ and/or $\hat{P}_{mC}(t) \sim 1$. Around this time, whose characteristic value is \bar{T}_1 , variations in $\text{Log}_{10}(N_Y/N_C)$ may develop. The behavior depends on the experimental design and mutational properties of the YFP and CFP populations, which are incorporated into evaluation of Equation 20.

While Equation 20 appears complex, it can be substantially simplified in a manner that preserves its quantitative accuracy and facilitates understanding of the essential factors that affect the behavior of the cultures. The simplification results from recognizing that only a small subset of the terms is significant and that key parameters are restricted in their values. First, most of the terms in Equation 20 are equal to 0 because the effective mutation time for most mutants is infinite, as indicated by the approach of $\Phi^*(\infty)$ to 1 as $t_{m_{ki}}$ increases during a culture period (Figure 5c) (*e.g.*, most are lost due to “drift” in the bottlenecks). Additionally, the non-zero terms have effective mutation times clustered near 0, as also indicated by Figure 5c. Moreover, Equation 2 shows that as time progresses the cells that constitute an appreciable proportion of the mutant population will have selection coefficients in the range where $\rho(s)W(s, t)$ has significant magnitude, which is typically very narrow. Finally, a mutant clone originating at an effective time $\hat{t}_{m_{ki}}^*$ contributes substantially to Equation 20 only if its selection coefficient is larger than any of the selection coefficients of mutations with smaller $\hat{t}_{m_{ki}}^*$. Therefore the most significant contributors to the mutant population

come from the ordered subset of mutations $[t_{m_{ki}}^*, s_{ki}]$, where both monotonically increase and the s_{ki} are restricted to a narrow range just below the maximum available selection coefficient, while the $t_{m_{ki}}^*$ are near 0.

Thus as time passes clonal succession occurs, with the population being dominated by mutants with selection coefficients tending toward the largest available. Section 4 of the supplemental materials at <http://www.genetics.org/supplemental/> calculates the average number of newly arising mutant cells transferred from a culture to its daughters, elucidating the buildup of the total mutant population.

Consider the YFP/CFP coculture at the time period around \overline{T}_1 , the time when $\hat{P}_{mY}(t)$ and $\hat{P}_{mC}(t) \sim 1$ on average. The behavior of the population ratio depends on the stochastically generated differences in $\hat{P}_{mY}(t)$ and $\hat{P}_{mC}(t)$, which are related to how densely the ranges of effective mutation times and selections coefficients are sampled in the two mutant populations. Imagine that N_0 for both the YFP and the CFP populations is very large, so that at \overline{T}_1 , which is independent of N_0 , many terms are required to make $\hat{P}_m(t) \sim 1$. Note that $\hat{P}_m(t)$ contains the factor $1/N_0$ (Equation 20) so that increasingly more terms are required as N_0 increases. Biologically the increase in the number of terms comes from the proportional increase in the number of mutations that occur due to the larger population size. But if there are many significant terms, then the intervals of both the effective mutation times and selection coefficients between sequential terms in the ordered subset $[t_{m_{ki}}^*, s_{ki}]$ must be small due to their limited ranges of the effective values. Although the specific values in $[t_{m_{ki}}^*, s_{ki}]$ will differ for the YFP and CFP populations due to the stochastic effects, the behavior of the $\hat{P}_m(t)$'s is insensitive to these differences and for both populations approaches that of $P_m(\infty, t)$ given by Equation 3. Thus the YFP and CFP populations evolve indistinguishably in terms of cell numbers, and the population ratio remains constant. All cocultures will appear to behave the same, and no indication of the internal stochastic differences will be measurable.

As N_0 is decreased, fewer terms are required in Equation 20 to result in $\hat{P}_m(t) \sim 1$ at $t \sim \overline{T}_1$, so the intervals between sequential terms in $[t_{m_{ki}}^*, s_{ki}]$ increase. The stochastic differences between the exact terms included in $[t_{m_{ki}}^*, s_{ki}]$ for the two populations lead to increasing possibilities for differential behavior between $\hat{P}_{mY}(t)$ and $\hat{P}_{mC}(t)$, so that greater variation in the population ratio will occur among a group of "identical" cocultures. The easily measurable effects will include increased variability in the T_1 's, increased magnitude of the ratio variations that develop, and a larger proportion of cultures in which the ratio variations are so large that one population effectively displaces the other. If the initial YFP and CFP populations differ substantially in size, for example, one is "small" and the other is "large," then their stochastic behaviors may be quite different in

character. Section 5 of the supplemental materials at <http://www.genetics.org/supplemental/> presents a rough method of estimating the number of significant terms in Equation 20 or, conversely, estimating the population size above which ratio variations are not expected.

The effect of population size on the stochastic variations in the population ratio can be appreciated by examination of the experimental data of HS (reproduced in section 6 of the supplemental materials at <http://www.genetics.org/supplemental/>). If their initial population size were increased by a factor of 10, the expected ratio behavior could be estimated by averaging randomly selected sets of 10 of their experimental curves, after first properly transforming the data prior to averaging. Clearly the ratio deviations would begin at approximately the same time, indicating the constancy of \overline{T}_1 , but would be of substantially reduced magnitude.

Equation 22 shows Equation 21 with the first few nonzero terms explicitly displayed,

$$\begin{aligned} & \text{Log}_{10}(N_Y/N_C) \\ &= \text{Log}_{10} \left[\frac{1 + (1/N_{0Y})(e^{-r_{Y1}h_{Y1}\tau} e^{r_{Y1}(t-\overline{t}_{Y1})} / e^{r(\overline{t}_{Y1})} + e^{-r_{Y2}h_{Y2}\tau} e^{r_{Y2}(t-\overline{t}_{Y2})} / e^{r(\overline{t}_{Y2})} + \dots)}{1 + (1/N_{0C})(e^{-r_{C1}h_{C1}\tau} e^{r_{C1}(t-\overline{t}_{C1})} / e^{r(\overline{t}_{C1})} + e^{-r_{C2}h_{C2}\tau} e^{r_{C2}(t-\overline{t}_{C2})} / e^{r(\overline{t}_{C2})} + \dots)} \right], \end{aligned} \quad (22)$$

where the offset term due to unequal populations has been neglected. As just discussed, in the regime where substantial population variations occur only a few terms are important. Figure 6a shows the behavior of Equation 22 under the assumptions that the initial population sizes are equal, $\rho(s)$ is a delta function, and only one mutant clone is contributing significant progeny in each population. Since all the selection coefficients are equal, if ratio fluctuations occur due to differences in the effective mutation times, the ratio will stabilize at some constant value after mutant cells dominate both populations. Mathematically the function always reaches a stable value, curve 2. However, this value may be so extreme that it represents the extinction of one of the populations, as illustrated by curve 1. The parameters for the curves are shown at the bottom of Figure 6.

The appropriateness of Equation 22 for describing real experiments can be assessed by comparison of its behavior to the experimental data of HS. The overall shapes of the curves in the time period where they begin to depart from $\text{Log}_{10}(N_Y/N_C) = 0$ and the rough magnitudes of the ranges of variation of the curves for the model and the experimental data are qualitatively similar. Thus a model containing single significant mutant clones in the two populations, and employing effective mutation times consistent with the range defined by Figure 5c, reasonably describes these aspects of the data. However, Figure 6a does not properly describe the long-term behavior of the ratio.

The experimental data clearly show that as mutant cells become dominant in both populations, the slope of $\text{Log}_{10}(N_Y/N_C)$ becomes small, but it is not typically 0. This is a clear indication of differences in the selection

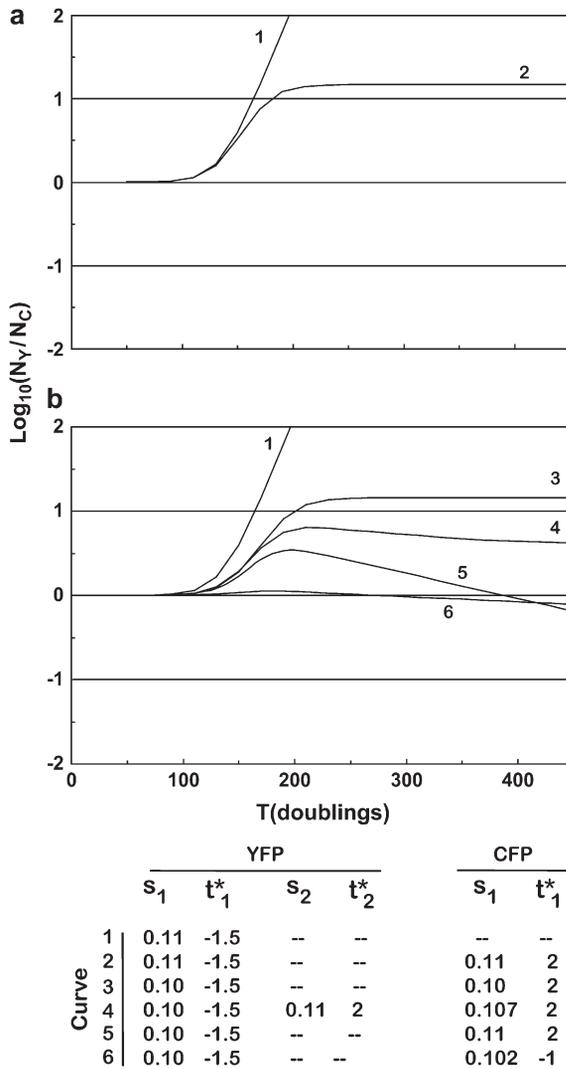


FIGURE 6.—Characteristic behavior of $\text{Log}_{10}(N_Y/N_C)$ as a function of time. (a) If all mutants have the same selection coefficient, $\rho(s)$ equals a delta function, and two basic behaviors are possible for $\text{Log}_{10}(N_Y/N_C)$. The ratio either departs from 0 and continues to increase (or decrease) until one population displaces the other (curve 1) or stabilizes at a constant value when mutants come to dominate both populations (curve 2). The effective mutation times t_m^* and the selection coefficients for the mutant clones for both a and b are shown at the bottom. (b) If the distribution of selection coefficients has finite width, then the behavior is more complex. The curves show the behavior of Equation 22 with either one or two (curve 4 only) mutant clones dominating $\hat{P}_m(t)$ for each population. In curve 1, a mutant occurs first in the YFP population, and it completely dominates the culture. Curve 3 shows the behavior if mutants with the same selection coefficient occur in both populations but at different effective times. This is unlikely to be stable as time progresses unless both selection coefficients are at the maximum available value, since a clone with higher s will eventually become dominant. Curve 4 shows two significant mutant clones in the YFP population with slightly different values of s and effective mutation times and a single significant clone in the CFP population with a value of s between those in the YFP mutants. In the YFP population the mutant with larger s must occur at a later effective time or the lower- s mutant would never affect the behavior of the culture. Note that the curve departs from 0 in

coefficients of the YFP and CFP mutants that are most prevalent in that time period. This behavior is modeled very well by employing differences in both the selection coefficients and the effective mutation times in Equation 22, as shown in Figure 6b. All but one of the curves in Figure 6b use one exponential term (one mutant clone) to describe the mutants in each population. The slopes of the curves after the mutants have overgrown the ancestral cells are directly related to selection coefficients of the dominant mutant clones. Taking the derivative of Equation 22 and assuming only one exponential term in the numerator and the denominator, one finds

$$s_Y - s_C = \frac{\Delta \text{Log}_{10}(N_Y/N_C)}{r(\Delta t) \text{Log}_{10}(e)} = 3.32 \frac{\Delta \text{Log}_{10}(N_Y/N_C)}{\Delta t}. \quad (23)$$

If more than one clone has significant prevalence in these populations, then the effective selection coefficients are weighted averages of the several clones. However, if more than one clone is prominent, then one expects the slope to progressively change with time as the one with the highest selection coefficient increases in prevalence relative to the others. This is illustrated by curve 4 in Figure 6b. Note that on the timescale of the actual experiment, the change in slope due to clonal succession is very subtle, since the selection coefficients are necessarily closely spaced.

In summary, if the population sizes are sufficiently large, multiple mutations will contribute significant proportions of mutant cells at times on the order of \bar{T}_1 . This corresponds to having many significant terms in Equation 20, and ratio variations will be of low amplitude. If the population sizes are sufficiently small, then Equation 22 with only one exponential term in the numerator and the denominator describes the experimental system, and one expects substantial ratio variations. Moreover, the smaller the population sizes are the larger the proportion of culture series that will be entirely overtaken by either the YFP or the CFP cells, e.g., curve 1 in Figure 6, a and b.

Measurement of bacterial characteristics from the experimental data: As indicated in the discussion following Equation 22, the experimental data from HS are qualitatively described by a very simple model containing only one prominent mutant clone in the YFP and CFP populations. Quantitative determination of biological parameters of the bacteria based on the model is

the positive direction and bends over to develop a negative slope, which slowly become less negative as the larger- s clone become more dominant in the YFP population. Curve 5 shows the behavior with single prominent mutant clones in both populations. Finally, curve 6 shows the effect of single mutants in each population with nearly identical selection coefficients and effective mutation times. The curve hardly departs from 0.

straightforward in this case. Once these parameters are obtained, they can be checked to determine if they are quantitatively consistent with the single-mutant clone description.

Estimation of $\overline{T_1}$: $\overline{T_1}$ can be determined by examining the timing of the initial departures of $\text{Log}_{10}(N_Y/N_C)$ from 0. T_1 is the time at which $\hat{P}_m(t) = 1$ for a particular population. If this happens in the YFP population significantly prior to the CFP population, then T_1 is just the time when $\text{Log}_{10}(N_Y/N_C) = 0.3$ (or -0.3 if the reverse occurs). More generally, both $\hat{P}_{mY}(t)$ and $\hat{P}_{mC}(t)$ have some significant value, so that $|\text{Log}_{10}(N_Y/N_C)|$ will be < 0.3 at $t \sim \overline{T_1}$. Thus a reasonable estimate for $\overline{T_1}$ is the time when many of the experimentally measured traces of $\text{Log}_{10}(N_Y/N_C)$ from a collection of culture series have significant departures from 0, but do not fully reach ± 0.3 . For the experimental data of HS, $\overline{T_1} \sim 150$ generations. A more complex statistical fitting of Equation 21 to the data would produce a better estimate for $\overline{T_1}$. The value of such a procedure depends on the noise level of the data.

Estimation of the maximum effective selection coefficient for the mutations: The maximum value of s can be estimated by examining the population-ratio curves for all of the experimental cultures to find the maximum slope of $\text{Log}_{10}(N_Y/N_C)$ after its initial departure from 0. This typically occurs in a culture series where one population completely and rapidly overtakes the other, presumably because a mutation with a near-maximum selection coefficient had a very early effective occurrence time (or perhaps was preexisting in the population). This approach leads to an estimate of $s_{\max} \sim 0.11$, using Equation 23 with one selection coefficient set equal to 0. To assist with the analysis, lines with various slopes have been added to the data figures of HS that are reproduced in the supplemental materials at <http://www.genetics.org/supplemental/>.

The analysis assuming a single-mutant clone is self consistent. Using the procedure in section 5 of the supplemental materials at <http://www.genetics.org/supplemental/>, Equation S8 estimates that there are typically 1.1 significant mutant clones in both populations. Of course, some cultures will by chance have more than one significant mutant clone in one or both populations, but as mentioned before, these cultures will typically have relatively low population-ratio deviations, and initial slopes of the curves will be relatively small compared to the others.

Estimation of the width of the effective selection coefficients: Examination of the slopes of $\text{Log}_{10}(N_Y/N_C)$ for the experimental data at times after mutants dominate both populations finds ~ 32 cultures series with $0 < |s_Y - s_C| < 0.02$ and only 5 with $0.02 < |s_Y - s_C| < 0.04$. (Given the large number of curves contained in the HS figures, and the measurement noise, it is difficult to be precise in these estimates.) Therefore most of the mutants must have nearly identical selection coefficients, and these

must be very close to s_{\max} given the large slope of the curves during their initial departure from 0. Thus one would estimate that most of the selection coefficients of the mutants that are significant in these experiments fall into the range of 0.09–0.11. This is just the behavior expected for a system with an effective mutation distribution described by $\rho(s)W(s, t)$ from Equation 2. Given the strong dependence of $W(s, t)$ on s , and the noise in the data, it is very difficult to determine much about the actual shape of $\rho(s)$ for the bacteria. Alternate experimental designs that may reveal more details of $\rho(s)$ are proposed in the DISCUSSION.

Estimation of the mutation rate: Finally, assuming a uniform distribution for the selection coefficients for new mutations, Equation 10a finds that $\mu \sim 10^{-5}$ if $\overline{T_1} \sim 150$ generations. The numerical value obtained for μ depends on the assumed form for $\rho(s)$. The inferred mutation rate would be much lower if a larger proportion of the weight of $\rho(s)$ were at higher selection coefficients and would be much higher if $\rho(s)$ were preferentially weighted toward 0. For comparison, if $\rho(s)$ were a delta function, the estimate from Equation 10b for the mutation rate would be $\sim 8.5 \times 10^{-7}$. Since all shapes of $\rho(s)$ lead to a very narrow distribution of effective selection coefficients, the value of the mutation rate for the delta-function distribution is an estimate of an effective mutation rate for the system—the other mutations that may occur will have limited discernible effect on the cultures. Therefore it represents a “lower bound” estimate of the actual mutation rate for these cells.

DISCUSSION

This article has presented an analytical description of competing asexual populations that highlights the essential processes underlying their behavior. Beginning with a qualitative consideration of the competing populations (Figure 1), it introduced a hypothetical construction that established the relationship of a single large culture in unbounded exponential growth to a set of embedded sequential finite cultures that are separated by bottlenecks (Figure 2). The finite cultures are subject to stochastic variation due to the mutational processes and the random assortment of mutants at the bottlenecks, but their average behavior is identical to the “infinite” culture. Since the infinite culture is described by the standard equations of exponential growth, average characteristics of the stochastic processes could be readily calculated.

Consideration of the infinite, continuously expanding system immediately yielded the result that the distribution of selection coefficients in the mutant cells is given by $\rho(s)W(s, t)$ (Equation 2). As time increases, $W(s, t)$ increases more strongly with increasing s (Figure 2b) so that the distribution of effective selection coefficients

becomes progressively more confined to a narrow range near the maximum s permitted by $\rho(s)$. Since this must also be true on average for the embedded finite cultures, Equation 2 is the analytical description of the “equivalence principle” that HS found through extensive numerical simulation.

The qualitative description of the behavior of a series of finite cultures (Figure 1b) highlighted the importance of the times T_1 , when the mutant and ancestral populations are of equal magnitude, for determining when discernible variations in the population ratios may begin to occur in competition experiments. Analysis of the infinite culture also allowed determination of the characteristic value for this time, \overline{T}_1 . Thus the distribution of selection coefficients that are prominent in the mutant populations, and the characteristic times at which ratio variations might occur in competition experiments, could be calculated without consideration of the details of the stochastic processes.

The stochastic aspects of the cultures were addressed by analyzing the fate of the progeny of a single mutation, again considering a hypothetical experiment in which all of the cells in a culture are transferred to daughter cultures (Figure 4). This analysis demonstrated that after a sufficient number of bottlenecks, the progeny of a mutated cell expand in a lineage of daughter cultures as a simple exponential, but with an apparent mutation time that is shifted from the actual mutation time according to readily calculable statistical distributions (Equations 15 and 19 and Figure 5). Summing over multiple mutants provided a complete description of the system (Equation 20).

Full evaluation of Equation 20 requires substantial effort. However, the results of the average value calculations, specifically those for \overline{T}_1 and the demonstration of the narrow range of effective selection coefficients, and the restricted range of effective mutation times derived from the statistical analysis, permitted a qualitative understanding of the basic behavior of Equation 20. These considerations lead to a dramatically simplified single-clone (or several-clone) description of a system of competing populations (Equation 22), which is applicable to the experimental regime where variations in the population ratios are prominent. Thus the interplay between a detailed statistical analysis and simple calculations of average values of various parameters for infinite populations leads to a tractable analytical description of competition experiments.

The analysis of experimental data to extract quantitative estimates of parameters of the mutational process is straightforward if the population sizes are small enough so that only a single exponential term is required in Equation 22 to represent the mutant cells. The maximum slope that the population ratio achieves after its initial departure from horizontal gives an estimate of the maximum value of the selection coefficient available to the mutant cells, and the average time that nonzero

slopes appear allows an estimate of \overline{T}_1 . If the population sizes are intermediate, so that ratio changes are observable but progeny of multiple mutations contribute, then the maximum deviations in the ratios are reduced since mutant clones become significant in both populations at about the same time and with about the same values of selection coefficients. In this case extraction of the biological parameters from the data would require a more complex statistical fitting process.

Application of the single-mutant clone version of Equation 22 to the experimental data of HS estimates that the maximum selection coefficient of the mutations is $s \sim 0.11$ and that the mutation rate is $\mu = 10^{-5}$ or 8.5×10^{-7} , depending on whether one assumes that $\rho(s)$ is uniform or a delta function. By contrast, HS found $s \sim 0.054$ and $\mu \sim 10^{-6.7} = 2 \times 10^{-7}$ when using their procedures to fit a delta-function model to their data (their Figure 3B). The reasons for the considerable differences in the estimates produced by the two modeling approaches require elucidation.

One of the benefits of the analytical approach taken here is that it reveals general relationships that may be obscured in more complex computational analyses. For example, exponential fitness distributions, $\rho(s) = (1/k)e^{-ks}$, have been postulated as being relevant to real biological systems (ORR 2003). However, if the expected behavior of a culture with such a distribution is calculated using Equation 3, one finds that at times $t > k/r$ the product $\rho(s)W(s, t)$ is unbounded with increasing s . Thus the fraction of mutant cells becomes infinite at finite times. This is a physically unreasonable result and indicates that in real systems the fitness distribution must go to 0 more rapidly than exponentially. An exponential truncated at some maximum s has the desired properties, and indeed truncation is assumed in the distributions proposed by ORR (2003).

The information concerning the shape of $\rho(s)$ that can be obtained from the HS measurements is limited because the experimental behavior is so strongly dominated by the limited range of selection coefficients defined by $\rho(s)W(s, t)$. Altering aspects of the experimental design may allow one to obtain more information. For example, one could substantially reduce the population sizes in the experiment and correspondingly increase the number of culture series. Such a change would not affect the characteristic time, \overline{T}_1 , at which the mutant populations will become significant, but would dramatically increase the variation in the T_1 's for the YFP and CFP populations in each culture series. This would increase the fraction of culture series in which one population completely displaces the other, *e.g.*, the curves labeled “1” in Figure 6, a and b. These displacements are predominantly due to single-mutant clones, and determination of the frequency with which different slopes are obtained during the displacement process provides a direct measure of the shape of $\rho(s)$. This design should allow measurement of selection

coefficients that are significantly below the maximum since the small population size would give these mutants a greater chance of becoming dominant prior to appearance of another mutant clone with larger s . The smaller the population size is, the lower the values of s that could be addressed in such an experiment, but the larger the number of series of cultures that would be needed.

Another possible approach would be to perform single-cell comparative growth experiments on a massive scale. This could be done by first growing several series of large cultures with all bacteria containing the same label, say YFP, for a time on the order of \overline{T}_1 . Approximately half of the cells in these mass cultures would be mutant, with the mutant population distributed in s according to $\rho(s)W(s, t)$. Growing the cultures for different times would provide different proportions of mutant and ancestral cells and somewhat different frequency distributions given the change in $W(s, t)$ with time. Shorter growth times would yield a broader distribution of selection coefficients for the mutants but a greater proportion of the cells would be normal, while longer times would produce a greater proportion of mutant cells but a narrower frequency distribution for s . Selection of a multitude of single YFP cells from the mass cultures and individually comparing their growth rates to normal CFP-labeled cells would allow measurement of the shape of $\rho(s)W(s, T_m)$, where T_m is the length of the initial mass culture. Division by $W(s, T_m)$ would provide a determination of $\rho(s)$. The range of s for which a reasonable estimate could be obtained would depend on the number of mutants that are measured and the accuracy of the measurements.

The better the experimental means of measuring the two cell populations is, the shorter the culture time required for the measurement and the more accurate the result. Thus high-sensitivity fluorescence monitoring of growth in microtiter plates, or single-cell analyt-

ical procedures such as flow cytometry or fluorescence microscopy, or perhaps highly parallel microfluidic measuring systems, would be beneficial. Ideally one might add one YFP cell to one ancestral CFP cell and determine the relative numbers at several time points during a single growth phase without the introduction of any bottlenecks. However, it is important to note that such an approach would measure the selection coefficients under conditions different from those of the original selection. Specifically, if a mutation provided its advantage during the initial mass culture by increasing the ability of cells to divide as the culture enters and/or leaves stationary phase, then that advantage will not be properly quantified.

In conclusion, the analytical approach presented here allows a qualitative and quantitative interpretation of the behavior of many aspects of competing asexual populations without the need to perform detailed numerical simulations. Analytical expressions for the narrowing of the range of effective selection coefficients and for various characteristic times during the adaptation of a population to new selective conditions are presented. The stochastic behavior of the population is accurately described, and the critical aspects of the description are highlighted to allow a simplified quantitative interpretation of experimental data. This analysis may be useful in considering alternate experimental designs that enhance probing of specific aspects of the adaptation process.

LITERATURE CITED

- HEGRENESS, M., N. SHORESH, D. HARTL and R. KISHONY, 2006 An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**: 1615–1617.
 ORR, H. A., 2003 The distribution of fitness effects among beneficial mutations. *Genetics* **163**: 1519–1526.

Communicating editor: A. D. LONG