DOI: 10.1534/genetics.106.062406

The Distribution of Beneficial Mutant Effects Under Strong Selection

Rowan D. H. Barrett,[1] Leithen K. M'Gonigle and Sarah P. Otto

Zoology Department, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Manuscript received June 22, 2006
Accepted for publication September 23, 2006

ABSTRACT

For a general theory of adaptation, it is essential to know the distribution of fitness effects of beneficial mutations. Recent theoretical and empirical studies have made considerable progress in determining the characteristics of this distribution. To date, the experiments have largely verified the theoretical predictions. Despite the fact that the theoretical work has assumed small selection coefficients, strong selection has been observed in some experiments, especially those involving novel environments. Here, we derive the distribution of fitness effects among fixed beneficial mutants without the restriction of low selection coefficients. The fate of strongly favored alleles is less affected by stochastic drift while rare, causing the distribution of fitness effects among fixed beneficial mutations to reflect more closely the distribution among all newly arising beneficial mutations. We also find that when many alleles compete for fixation within an asexual population (clonal interference), the beneficial effects of a newly fixed mutant cannot be well estimated because of the high number of subsequent mutations that arise within the genome, regardless of whether selection is strong or weak.

POPULATIONS adapt to their environments through the appearance and subsequent spread of random beneficial mutations. In a sexual population, recombination can bring together beneficial mutations that arise in different lineages. In asexual populations, however, mutations can fix only sequentially (Novick and Szilard 1950; Atwood *et al.* 1951; Crow and Kimura 1965). Distinct genotypes cannot recombine and instead must compete with each other, a phenomenon known as "clonal interference" (Gerrish and Lenski 1998; Miralles *et al.* 1999; de Visser and Rozen 2006). Thus, for a mutation to contribute to adaptation it not only must escape sampling error (drift), but also must fix before being eliminated by the occurrence and more rapid sweep of a superior mutation.

Theory for determining the probability that selection will fix a new favorable mutation was first formulated 80 years ago by Fisher (1922, 1930) and Haldane (1927), who focused on the fate of a single isolated mutation. In a recent flurry of articles, researchers have explored the distribution of fitness effects expected among the array of possible beneficial mutations that might arise within a population (Gillespie 1983, 1984, 1991; Orr 2002, 2003). This distribution can be seen as the starting point for progress toward a general theory of adaptive evolution. From this first distribution, we can determine the distribution of those mutations not lost by drift, referred to as "contending mutations" (Gerrish and Lenski 1998; Rozen *et al.* 2002). The distribution of contending mutations can subsequently be used to determine the distribution of mutations that outcompete other genotypes and fix in a population ("fixed mutations"), thus contributing to adaptation (Orr 1998; Rozen *et al.* 2002).

In addition to theory, recent empirical studies have examined the steps from newly arisen mutations, to contending mutations, to fixed mutations. The underlying distribution of fitness effects of new beneficial mutations inferred from experiments is generally consistent with an exponential distribution (Imhof and Schlotterer 2001; Rozen *et al.* 2002; Sanjuan *et al.* 2004; Kassen and Bataillon 2006), while the final distribution of fixed beneficial mutants appears roughly bell-shaped (Rozen *et al.* 2002; Rokyta *et al.* 2005; Barrett *et al.* 2006). These empirical results must be interpreted cautiously, however, as there is typically little power to reject other distributions (*e.g.*, more L-shaped or more bell-shaped distributions, see Kassen and Bataillon 2006).

An important caveat to the theoretical side of this work, however, is that it has assumed weak selection. R. A. Fisher (Fisher 1930) first justified this assumption using the analogy of movement from the outer surface of a sphere (representing phenotype space) to an optimum at the center; Fisher argued that mutations of small size have a 50% chance of bringing the population closer to the optimum, while larger mutations have a rapidly diminishing probability of being advantageous. This argument led to a "gradualist" view of adaptation, in which evolutionary change overwhelmingly proceeds through the selection of very slightly beneficial alleles

[1]*Corresponding author:* Zoology Department, University of British Columbia, 6270 University Blvd., Vancouver, BC V6T 1Z4, Canada. E-mail: rbarrett@zoology.ubc.ca

(Orr 2005). Assumptions of weak selection common to all diffusion equations and many other theoretical approaches make it difficult to predict the dynamics of strongly beneficial mutations (Morjan and Rieseberg 2004). Simulations have demonstrated that although fixation times for strongly advantageous alleles are accurately predicted by diffusion, the probability of fixation is underestimated as the strength of selection increases (Whitlock 2003).

In recent empirical studies, researchers have observed mutations with very high selection coefficients (Bull *et al.* 2000; Barrett *et al.* 2006), especially when organisms face novel environments. Strong selection has an important impact on the theory of adaptation. In particular, the fixation probability of mutations and the number of competing mutations will be highly dependent on the fitness effect of the mutation in question (Rozen *et al.* 2002). When selection is assumed to be weak, new mutations remain at low frequency for a considerable period of time before reaching fixation (Gerrish and Lenski 1998). This provides ample opportunity for beneficial mutations to be lost and for competing mutations to arise (Gerrish 2001). In contrast, when selection is strong, the probability of fixation approaches its maximum value of one and the time to fixation is relatively short, reducing the number of competing mutations and the importance of clonal interference. Yet previous predictions cannot be applied to the case of strong selection, because of the pervasive theoretical assumption that selection is weak.

Here, we derive population genetic theory to describe the impact of drift and clonal interference on the fitness distribution of fixed beneficial alleles without the assumption of weak selection. We derive these distributions for a wide range of selection coefficients and test our analytical theory against numerical simulations.

THE MODEL

In the following, we describe the probability density functions (pdfs) of the selection coefficient, s, among newly arising beneficial mutations, contending beneficial mutations, and fixed beneficial mutations. Following Rozen *et al.* (2002), we denote these pdfs as $f(s)$, $g(s)$, and $h(s)$, respectively. Where needed, we use uppercase letters to refer to the corresponding cumulative density functions (cdfs) [$F(s)$, $G(s)$, and $H(s)$, respectively]. Throughout, we assume that the population size is large, haploid, and asexual. The results may be applied to asexual diploids by replacing s with $h\,s$, where h is the dominance coefficient and mutations are assumed to fix in the heterozygous condition. An extension to sexual diploids is straightforward (at least numerically), but it requires that the joint distribution of h and s be specified.

The distribution of beneficial mutations: It is generally assumed that the wild type has very high fitness and almost all mutations are deleterious (Gillespie 1983, 1984; Orr 1998). It follows that beneficial mutations will lie in the extreme right tail of a distribution of all mutant fitness effects. This inference justifies the application of extreme value theory (Gumbel 1958) to describe the distribution of beneficial mutant effects. Extreme value theory suggests that the distribution of mutant effects, restricted to beneficial mutations, will be nearly exponential (Gillespie 1983, 1984). This requires, however, that only a tiny minority of mutations are beneficial. In very harsh or novel environments, mutations that were previously deleterious may become beneficial, thus increasing the size of the beneficial mutant class. In these situations, extreme value theory may not hold and an exponential distribution might not be an adequate description of the selective effects of new mutants.

While empirical and theoretical studies indicate that the exponential distribution is a plausible distribution describing the fitness effects of new beneficial mutations, distributions with other shapes cannot be rejected and might be more appropriate under certain circumstances. Thus, to allow greater flexibility, we assume that the selection coefficients of new beneficial mutants, s, follow a gamma distribution with mean selection coefficient, σ, and coefficient of variation, cv,

$$f(s) = \frac{e^{-s/(\mathrm{cv}^2\sigma)}(\mathrm{cv}^2\sigma)^{-1/\mathrm{cv}^2}s^{-1+1/\mathrm{cv}^2}}{\Gamma[1/\mathrm{cv}^2]}, \qquad (1)$$

where $\Gamma[a]$ is Euler's gamma function. The shape of the gamma distribution varies from L-shaped (high cv) to bell-shaped (low cv), allowing a broader range of distributions to be described. The exponential distribution represents a specific case of the gamma where the coefficient of variation equals one.

The distribution of contending beneficial mutations: We begin by deriving the distribution of fitness effects among those contending mutations that survive stochastic loss while rare. Haldane (1927) used a branching process to show that P, the probability of fixation, satisfies $1 - P = e^{-(1+s)P}$ in populations of constant size when the number of offspring per parent is Poisson distributed. Using a diffusion approximation, Kimura (1957, 1964, 1983) extended this theory for populations of finite size, N, showing that $P \approx (1 - e^{-2s})/1 - e^{-2Ns})$. For weak selection ($1/N \ll s \ll 1$), both of these equations yield the same approximate fixation probability: $P \approx 2s$. When strong selection is possible, however, a more accurate approximation for both equations is: $P = 1 - e^{-2s}$, obtained by letting N get large in the diffusion result (Figure 1). These results assume that a mutant allele is either lost or fixed before other mutations arise. When the mutation rate is sufficiently high, however, a mutation may survive loss while rare, but eventually be outcompeted by mutations arising in the future. In this case, P describes the probability of surviving stochastic

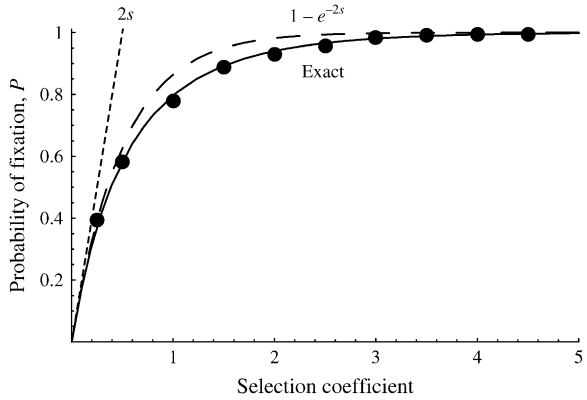

FIGURE 1.—The probability of fixation for a beneficial mutation as a function of its selection coefficient. The dots show the probability of fixation of a single mutant with a given selection coefficient in Wright–Fisher simulations of a large population ($N = 10^6$; based on 1000 replicates). The solid curve shows the exact fixation probability derived by HALDANE (1927) using a branching process. The short-dashed line shows the common weak selection approximation $P \approx 2s$. The long-dashed curve shows the approximation used in this article, $P = 1 - e^{-2s}$, which is indistinguishable from Kimura's diffusion result for these parameters.

loss while rare, not the ultimate fixation probability, which depends on the nature of future mutations.

Because newly arisen mutations have a probability of surviving stochastic loss while rare of $\sim 1 - e^{-2s}$ in populations of large size, the distribution of selection coefficients among contending mutations becomes

$$g(s) = \frac{f(s)(1 - e^{-2s})}{\int f(s)(1 - e^{-2s})ds}. \quad (2)$$

The denominator represents the probability of surviving drift averaged across the distribution of new mutational effects, Π:

$$\Pi = \int f(s)(1 - e^{-2s})ds = 1 - (1 + 2\,\mathrm{cv}^2\sigma)^{-1/\mathrm{cv}^2}. \quad (3)$$

Using this result the pdf of the selection coefficients among contending mutations is

$$g(s) = \frac{e^{-s/(\mathrm{cv}^2\sigma)}(\mathrm{cv}^2\sigma)^{-1/\mathrm{cv}^2}s^{-1+1/\mathrm{cv}^2}(1 - e^{-2s})}{\Gamma[1/\mathrm{cv}^2](1 - [1 + 2\,\mathrm{cv}^2\sigma]^{-1/\mathrm{cv}^2})}. \quad (4)$$

Because the probability of surviving loss while rare, $P = 1 - e^{-2s}$, asymptotes at one (Figure 1), the distribution $g(s)$ is similar to the prior distribution $f(s)$ for mutations of large effect.

The distribution of fixed beneficial mutations: ROZEN *et al.* (2002) defined the expected number of contending mutations arising within other genetic backgrounds before the fixation of a focal mutation as

$$\lambda(s) = PN\mu\frac{T}{2}, \quad (5)$$

where P is the average probability of surviving loss while rare, N is the population size, μ is the beneficial mutation rate, and T is the average amount of time until fixation (the $\frac{1}{2}$ reflects the fact that, by symmetry, half of the population will not carry the focal mutation when averaged over the period of time during which the focal mutation rises from a single copy to fixation; see supplemental Figure 1 at http://www.genetics.org/supplemental/). ROZEN *et al.* (2002) then calculated $\lambda(s)$ under the assumption of weak selection. To relax this assumption, we use the general solution for the deterministic haploid model, $q(t)/p(t) = (1+s)^t q(0)/p(0)$, where $p(t)$ and $q(t)$ are the frequencies of nonmutant and mutant individuals, to solve for the time taken for an allele initially at frequency $1/N$ to reach a frequency of $1 - 1/N$. This gives $T = 2\ln(N-1)/\ln(1+s) \approx 2\ln(N)/\ln(1+s)$, which can be used along with $P = 1 - e^{-2s}$ in (5) to estimate the number of contending mutations:

$$\lambda(s) \approx (1 - e^{-2s})N\mu\frac{\ln(N)}{\ln(1+s)}. \quad (6)$$

When there are n contending mutations and one focal mutation, each with a selection coefficient drawn from a cdf given by $G(s)$, the cumulative density function of the highest of the selection coefficients is $G(s)^{n+1}$ (RICE 1988). Assuming that the number of contending mutations that appear during the spread of a focal mutation follows a Poisson distribution, with mean $\lambda(s)$, the cdf for the selection coefficient of the most advantageous of the contending and focal mutations is

$$H(s) = N^{-N\mu\kappa/\ln(1+s)}\left(1 - \frac{\kappa}{\Pi}\right), \quad (7)$$

where Π is the average probability of fixation (Equation 3) and

$$\kappa = \frac{\Gamma[1/\mathrm{cv}^2,\ s/(\sigma\,\mathrm{cv}^2)]}{\Gamma[1/\mathrm{cv}^2]} - (1 - \Pi)\frac{\Gamma[1/\mathrm{cv}^2,\ s(1-\Pi)^{-\mathrm{cv}^2}/(\sigma\,\mathrm{cv}^2)]}{\Gamma[1/\mathrm{cv}^2]}. \quad (8)$$

If newly arising beneficial mutations follow an exponential distribution ($\mathrm{cv} = 1$), these coefficients simplify to

$$\Pi = \frac{2\sigma}{1 + 2\sigma} \quad \text{and} \quad \kappa = e^{-s/\sigma} - \frac{e^{-2s-s/\sigma}}{1 + 2\sigma}.$$

This equation accounts only for contending mutations that arise after the focal mutation, as no improvement in fit was observed when accounting for prior mutations (data not shown). (Essentially, we consider the first contending mutation to be the focal mutation.) The corresponding probability density function for fixed mutations is then $h(s) = dH(s)/ds$.

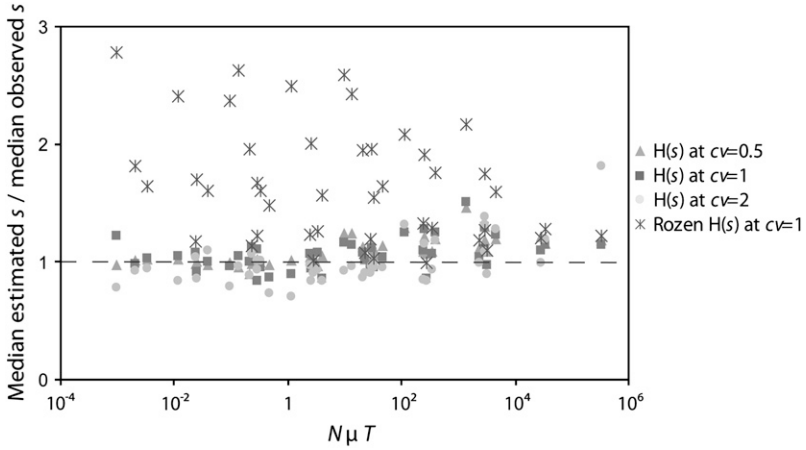

FIGURE 2.—Median selection coefficient of fixed beneficial mutations estimated from analytical results *vs.* numerical simulations. Median estimated s is given from the results of either our generalized selection model, given by $H(s)$ in Equation 7, or a model assuming weak selection and a cv of 1 (Equation 2 in ROZEN *et al.* 2002). The horizontal axis measures the degree of clonal interference and equals the number of mutations that appear within the population over the time to fixation, T, for a new mutation with selection coefficient given by the median observed s. Results are based on 200 replicate simulations per parameter combination.

NUMERICAL SIMULATIONS

We compare the above analysis to explicit numerical simulations using a Wright–Fisher model (EWENS 1979). We tracked all beneficial mutants segregating in an asexual haploid population of constant size N until a fixation event. Each generation, the number of new mutations appearing within the population was drawn at random from a Poisson distribution with mean $N\mu$. Each mutation was then randomly assigned a selection coefficient drawn from a gamma distribution with mean σ and coefficient of variation cv and assigned a unique identifier. Multiple mutations had independent effects on fitness (no epistasis on a multiplicative scale). Offspring were then sampled with replacement according to a multinomial distribution from the parental distribution of genotypes, weighted by the fitness of these genotypes. A fixation event was defined as the first point in time when all individuals in the population shared a common mutation (with the same identifier). Similar results were obtained when we recorded data for the fifth mutation rather than the first mutation to fix (data not shown). At this point, the process was stopped and the selection coefficient of the fixed mutant was recorded. This selection coefficient was defined as the fitness effect of the fixed mutant when placed in the ancestral background; the average fitness within the population at the time of fixation was also recorded.

To evaluate the robustness of the analytical results, we ran simulations with every combination of the following average selection coefficients ($\sigma =$ 0.01, 0.1, 1, 2, and 10), beneficial mutation rates ($\mu = 10^{-5}$, 10^{-7}, and 10^{-9}), population sizes ($N = 10^5$, 10^6, and 10^7), and coefficients of variation (cv = 0.5, 1, and 2). Simulations were carried out in Mathematica (WOLFRAM RESEARCH 2005; available upon request).

NUMERICAL RESULTS

The numerical simulations closely match the predictions from our model across the parameter range explored (Figure 2). In contrast, analyses based on weak selection (ROZEN *et al.* 2002) consistently overestimate the selection coefficient among fixed mutations (Figure 2). Assuming a fixation probability of $2s$ gives unrealistically high fixation probabilities for mutants with large selection coefficients, which inflates the proportion of contending and fixed mutations of large effect.

Both analytical and numerical results predict bell-shaped distributions for the selection coefficients among fixed beneficial mutations (Figure 3). The distribution of contending mutations, $g(s)$ (thin solid curves), is always bell-shaped because weakly selected mutations are likely to be lost while rare (KIMURA 1983). With a low mutation rate (Figure 3, A and B), the distribution of fixed mutations (histogram) is very nearly equal to the distribution of contending mutations, $g(s)$, and clonal interference has little effect. With a higher mutation rate (Figure 3, C and D), clonal interference becomes more important, and only the most fit of the contending mutations fixes within the population, shifting the distribution of fixed mutations to the right. When selection is, on average, stronger (Figure 3, B and D), mutations are less likely to be lost through stochastic drift while rare, causing the distribution of fixed beneficial mutations to be more similar to the distribution of selection coefficients among newly arising mutations, $f(s)$ (dotted curves); consequently, the mean and coefficient of variation among mutations that fixed in the simulations are more similar to the original mean, σ, and cv (inset boxes). The shape of the contending and fixed distributions is also influenced by the shape of the distribution of underlying beneficial mutations (supplemental Figure 2 at http://www.genetics.org/supplemental/). Increasing the coefficient of variation of beneficial mutations results in more contending mutations of large and small effect, increasing the variation observed among fixed mutations.

If there is a high input of new mutations, it becomes more likely for several beneficial mutations to coexist in a population. This leads to clonal interference, as beneficial mutations in different genetic backgrounds

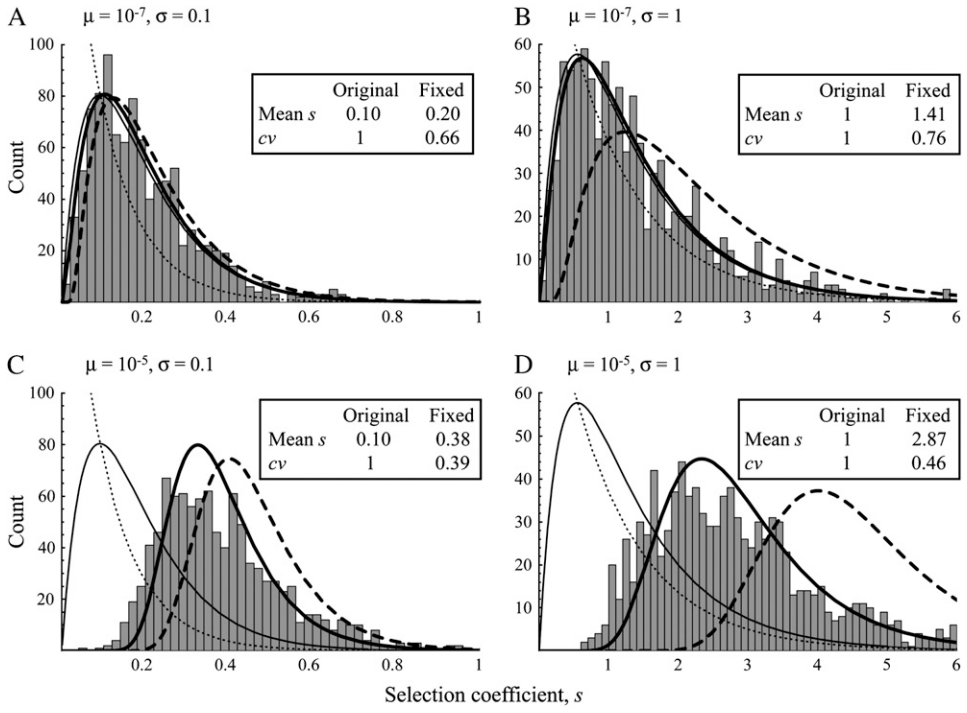

FIGURE 3.—Distribution of fitness effects. The histogram was generated from the fitness effects of fixed beneficial mutations among 1000 replicate simulations; fixed mutations have a higher average selection coefficient and exhibit less variation than newly arising mutations (inset boxes). New mutations were distributed according to an exponential distribution, $f(s)$ (thin dotted curve). Contending mutations that survived loss while rare (thin solid curve) are shifted to the right [$g(s)$ from Equation 4]. Fixed mutations (thick solid curve) that survived clonal interference are even further shifted to the right [$h(s)$ from the derivative of Equation 7] but not as much as predicted using the weak selection approximation given by Equation 2 of ROZEN *et al.* (2002) (thick dashed curve). Clonal interference is more important in C and D, where mutation rates are higher (A and B, $\mu = 10^{-7}$; C and D, $\mu = 10^{-5}$). Selection is weaker, on average, in A and C ($\sigma = 0.1$) than in B and D ($\sigma = 1$). Remaining parameters: cv = 1 and $N = 10^5$.

compete with one another. In addition, however, a high mutation rate also makes it more likely that multiple beneficial mutations will arise in the same background as previous beneficial mutations. In this case, several beneficial mutations can assist each other's spread to fixation, and the combined fitness advantage from these mutations will be higher than the fitness advantage conferred by the single original mutation. To assess the importance of assisted fixation, we measured the average fitness advantage in the population at the time a beneficial mutation fixed (relative to the nonmutant ancestor). The average fitness advantage was greater than the fitness advantage conferred by the mutation alone (Figure 4), often by orders of magnitude when the mutation rate was high enough to expect clonal interference ($N\mu > 1$). As expected, this discrepancy was caused by the effects of additional beneficial mutations segregating within the population at the time of fixation.

We have so far assumed a constant population size, but many experiments designed to detect beneficial mutations involve repeated bottlenecks and a fluctuating population size. Such fluctuations dramatically increase the chance of loss of beneficial mutations, so that only the most favorable alleles are likely to fix. In the APPENDIX we modify the theory developed above to describe how fluctuating population size alters the fixation probability and the time to fixation. We then estimate the number of competing mutations, $\lambda(s)$, and the distribution of fixed mutations, $h(s)$.

DISCUSSION

On the basis of his geometric model of the adaptive process, FISHER (1930) argued that mutations of very small effect have a nearly 50% chance of pointing toward an optimum, while mutations of very large effect rarely will. This reasoning underlies the common assumption in population genetics that adaptation consists of fine tuning the phenotype with mutations of relatively small effect. What constitutes a large mutation in Fisher's model depends, however, on the fitness of the original population. If a population is initially poorly adapted (*e.g.*, following a recent change in the environment), even major mutations with a substantial effect on phenotype have a nearly 50% chance of pointing toward the optimum. Thus, strongly selected mutations may very well contribute to the process of adaptation, especially during the early stages of adaptation to a novel environment. Furthermore, by virtue of their size, large-effect mutations will have a disproportionate influence on the process of adaptation. Data from genetic analyses of quantitative trait differences (BRADSHAW *et al.* 1998; WANG *et al.* 1999; COLOSIMO *et al.* 2005) and from experimental evolution studies (BULL *et al.* 2000; BARRETT *et al.* 2006) confirm that mutations with large phenotypic and fitness effects can occur and contribute to the process of adaptation. In this article, we have generalized existing theory about the distribution of fitness effects among fixed beneficial mutations so that it can be applied to situations with strong selection.

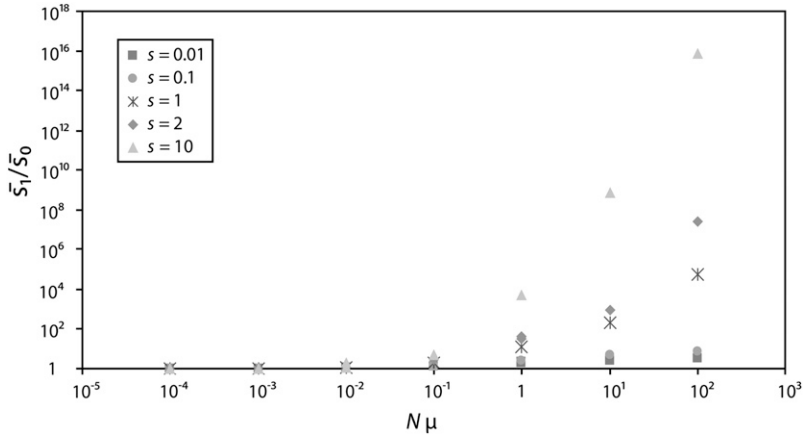

FIGURE 4.—Error caused by measuring the selection coefficient using the average fitness at the time of fixation. The vertical axis gives the amount by which the average fitness in the population has improved over the ancestor when the focal mutation fixes, $\bar{s}_1$, relative to the selection coefficient of that mutation alone, $\bar{s}_0$. The horizontal axis gives the input rate of new mutations, $N\mu$. Results are based on 200 replicate simulations.

The distribution of fitness effects among fixed beneficial mutations is generally derived from the distribution of fitness effects among all possible beneficial mutations, about which little is known. Several theoretical studies have suggested that new beneficial mutations should be exponentially distributed (MUKAI *et al.* 1972; ORR 1998; ROZEN *et al.* 2002; WILKE 2004), on the basis of the fact that beneficial mutations represent the tail of the distribution of potential mutant effects (GILLESPIE 1983, 1984). In a novel environment, however, more mutations are likely to be beneficial and the applicability of such extreme value theory is uncertain. We have thus employed a gamma distribution to describe the fitness effects of possible beneficial mutations. Because the gamma distribution has two parameters (described by the mean selection coefficient, σ, and the coefficient of variation, cv), we can explore a broader range of possible distributions of mutational effects. We find that the shape parameter of the gamma distribution among newly arising mutations influences the distribution of mutations that survive stochastic loss while rare (contending mutations) and the distribution of mutations that survive clonal interference to become fixed (fixed mutations), especially when selection is strong (supplemental Figure 2 at http://www.genetics.org/supplemental/). This result appears to contradict a recent study, which reported that the distribution of mutational sizes for fixed mutations is virtually independent of the underlying distribution of beneficial mutations (ORR 1998; HEGRENESS *et al.* 2006). The simulations run by HEGRENESS *et al.* (2006) cover only a range of parameters, within which clonal interference is severe (the population size was set to 2×10^6 and the mutation rate was 10^{-5}). Indeed, using their combination of parameter values in Equation 7 indicates that the shape of the distribution of fixed beneficial mutations is nearly independent of the shape of the distribution of newly arising beneficial mutations. Furthermore, using their parameters, most fixed mutations have similar selection coefficients, as pointed out by HEGRENESS *et al.* (2006), unlike the fairly broad distributions observed in Figure 3.

Beneficial mutations that survive stochastic loss while rare tend, on average, to have a larger fitness benefit, and their distribution tends to have a lower coefficient of variation (more bell-shaped), because very weakly selected alleles are unlikely to fix (KIMURA 1983; GERRISH and LENSKI 1998; ORR 2000; OTTO and JONES 2000; ROZEN *et al.* 2002; WILKE 2004). While this is generally true, the effect is less pronounced when selection is strong. That is, the distribution of fixed beneficial mutations is more similar to the distribution of newly arising mutations (Figure 3). Consequently, for empirical data involving high selection coefficients, using theory that assumes weak selection will tend to underestimate the mean selection coefficient among newly arising beneficial mutations.

One of the major impediments to theoretical studies of the distribution of fitness effects of fixed beneficial mutations has been a lack of knowledge of realistic parameter values. This is largely because the low frequency of fixed beneficial mutations has prevented empirical work with statistical power. In recent years, however, the use of microbial microcosms has provided a way to increase the number of beneficial mutations likely to arise and fix during an experiment. Three experiments have characterized the distribution of fixed beneficial mutant effects (ROZEN *et al.* 2002; ROKYTA *et al.* 2005; BARRETT *et al.* 2006). All used roughly the same experimental protocol: a number of replicate bacterial or viral lines were introduced into a novel environment and evolution proceeded through the substitution of novel beneficial mutations. By comparing the fitness of an evolved genotype sampled from around the time a mutation fixed within the evolved population to the fitness of the ancestral genotype, these studies claimed to measure the fitness advantage conferred by the single beneficial mutation carried by each evolved genotype. Our simulations indicate, however, that the selection coefficient estimated from the average fitness of individuals at the time of fixation is a very poor measure of the fitness effect of the actual mutation that has just fixed whenever $N\mu > 1$ (Figure 4). Whenever multiple mutations arise during the spread of a focal mutation, so that clonal interference occurs, individuals are likely

to carry multiple mutations by the time that the focal mutation has fixed, and therefore selection coefficients measured will overestimate the true effects of a single mutation. Consequently, experiments aimed at estimating the distribution of beneficial selective effects should avoid large population size to keep $N\mu < 1$.

The studies by Rozen *et al.* (2002) and Barrett *et al.* (2006) serve as a good comparison of how the distribution of fitness effects of fixed beneficial mutations shifts with increasing selection. Both experiments were conducted with similar organisms and transfer protocols and therefore share fairly comparable parameter values, except for the average selection coefficient, which differed by an order of magnitude. This difference is perhaps unsurprising as the ancestral strain in the Rozen *et al.* (2002) experiment was fairly well adapted to consuming the sole carbon source (*Escherichia coli* with glucose), whereas the ancestral strain in the Barrett *et al.* (2006) experiment initially had very poor growth (*Pseudomonas fluorescens* with serine). Despite the difference in average selection coefficients, both studies reported bell-shaped distributions for fixed beneficial mutations. Both sets of authors suggest that this shape is the result of drift and clonal interference transforming an exponential distribution of beneficial mutations. However, since $N_e\mu < 1$, the bell-shaped distributions are unlikely to be strongly influenced by clonal interference. Indeed, the bell-shaped distributions observed in these studies can be accounted for entirely by the stochastic loss while rare of mutations drawn from an exponential distribution (Equation 4), without considering competing mutations.

In conclusion, allowing for strong selection has altered our theoretical understanding of the distribution of fitness effects in the following ways. By correctly accounting for the fact that the fixation probability cannot rise above one, the distribution of fixed beneficial mutations more closely matches the distribution of newly arising beneficial mutations when selection is strong (Figure 3). Although the distribution of surviving mutations is always more bell-shaped, the difference from the distribution of newly arising beneficial mutations is largely confined to regions where selection is weak. Importantly, our results (Figure 4) also demonstrate that data on the selection coefficients of fixed mutations must be treated with caution whenever clonal interference is present, as multiple mutations are likely to be segregating at the time of fixation, causing selection coefficients to be greatly exaggerated.

We thank the OLM discussion group for helpful comments on the manuscript. This work was supported by funding from the Natural Science and Engineering Research Council of Canada to R.D.H.B. (PGS–D grant), L.K.M. (CGS–M grant), and S.P.O. (discovery grant).

LITERATURE CITED

Atwood, K. C., L. K. Schneider and J. F. Ryan, 1951 Periodic selection in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **37:** 146–155.

Barrett, R. D. H., R. C. MacLean and G. Bell, 2006 Mutations of intermediate effect are responsible for adaptation in evolving *Pseudomonas* populations. Biol. Lett. **2:** 236–238.

Bradshaw, H. D., K. G. Otto, B. E. Frewen, J. K. McKay and D. W. Schemske, 1998 Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (Mimulus). Genetics **149:** 367–382.

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.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, M. Dickson *et al.*, 2005 Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science **307:** 1928–1933.

Crow, J. F., and M. Kimura, 1965 Evolution in sexual and asexual populations. Am. Nat. **99:** 439–450.

de Visser, J. A. G. M., and D. E. Rozen, 2006 Clonal interference and periodic selection of new beneficial mutations in *Escherichia coli*. Genetics **172:** 2093–2100.

Ewens, W. J., 1967 The probability of survival of a new mutant in a fluctuating environment. Heredity **22:** 307–310.

Ewens, W. J., 1979 *Mathematical Population Genetics*. Springer, Berlin.

Fisher, R. A., 1922 On the dominance ratio. Proc. R. Soc. Edinb. **42:** 321–341.

Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Oxford University Press, Oxford.

Gerrish, P., 2001 The rhythm of microbial adaptation. Nature **413:** 299–302.

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

Gillespie, J. H., 1983 A simple stochastic gene substitution model. Theor. Popul. Biol. **23:** 202–215.

Gillespie, J. H., 1984 Molecular evolution over the mutational landscape. Evolution **38:** 1116–1129.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, New York.

Gumbel, E. J., 1958 *Statistics of Extremes*. Columbia University Press, New York.

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

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.

Imhof, M., and C. Schlotterer, 2001 Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. Proc. Natl. Acad. Sci. USA **98:** 1113–1117.

Kassen, R., and T. Bataillon, 2006 Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. Nat. Genet. **38:** 484–488.

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

Kimura, M., 1964 Diffusion models in population genetics. J. Appl. Probab. **1:** 177–232.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.

Kimura, M., and T. Ohta, 1974 Probability of gene fixation in an expanding finite population. Proc. Natl. Acad. Sci. USA **71:** 3377–3379.

Miralles, R., P. J. Gerrish, A. Moya and S. F. Elena, 1999 Clonal interference and the evolution of RNA viruses. Science **285:** 1745–1747.

Morjan, C. L., and L. H. Rieseberg, 2004 How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. **13:** 1341–1356.

Mukai, T., S. I. Chigusa, L. E. Mettler and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics **72:** 333–355.

Novick, A., and L. Szilard, 1950 Experiments with the chemostat on spontaneous mutation of bacteria. Proc. Natl. Acad. Sci. USA **36:** 708–720.

Orr, H. A., 1998 The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution **52:** 935–949.

Orr, H. A., 2000 The rate of adaptation in asexuals. Genetics **155:** 961–968.

Orr, H. A., 2002 The population genetics of adaptation: the adaptation of DNA sequences. Evolution **56:** 1317–1330.
Orr, H. A., 2003 The distribution of fitness effects among beneficial mutations. Genetics **163:** 1519–1526.
Orr, H. A., 2005 The genetic theory of adaptation. Nat. Genet. **6:** 119–127.
Otto, S. P., and C. D. Jones, 2000 Detecting the undetected: estimating the total number of loci underlying a quantitative trait. Genetics **156:** 2093–2107.
Otto, S. P., and M. C. Whitlock, 1997 The probability of fixation in populations of changing size. Genetics **146:** 723–733.
Rice, J. A., 1988 *Mathematical Statistics and Data Analysis*. Wadsworth, Belmont, CA.
Rokyta, D. R., P. Joyce, S. Brian Caudle and H. A. Wichman, 2005 An empirical test of the mutational landscape model of adaptation using a single-stranded DNA virus. Nat. Genet. **37:** 441–444.
Rozen, D. E., J. A. G. M. de Visser and P. J. Gerrish, 2002 Fitness effects of fixed beneficial mutations in microbial populations. Curr. Biol. **12:** 1040–1045.
Sanjuan, R., A. Moya and E. Santiago, 2004 The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. Proc. Natl. Acad. Sci. USA **101:** 8396–8401.
Wahl, M. L., P. J. Gerrish and I. Saika-Voivoda, 2002 Evaluating the impact of population bottlenecks in experimental evolution. Genetics **162:** 961–971.
Wang, R. L., A. Stec, J. Hey, L. Lukens and J. Doebley, 1999 The limits of selection during maize domestication. Nature **398:** 236–239.
Whitlock, M., 2003 Fixation probability and time in subdivided populations. Genetics **164:** 767–779.
Wilke, C. O., 2004 The speed of adaptation in large asexual populations. Genetics **167:** 2045–2053.
Wolfram Research, 2005 *Mathematica*. Wolfram Research, Champaign, IL.

Communicating editor: N. Takahata

APPENDIX

Assuming weak selection and a fluctuating population size, the average probability of fixation of a beneficial mutation is $\sim 2sN_e/\overline{N}$ (Ewens 1967; Kimura and Ohta 1974; Otto and Whitlock 1997). Here, the arithmetic average population size is $\overline{N}$, and the "effective" population size is N_e, whose calculation depends on the nature of the population fluctuations (Otto and Whitlock 1997; Wahl *et al.* 2002). Unfortunately, we lack an analytical expression for the fixation probability when selection is strong and population size varies. We conjecture that an adequate approximation for the average fixation probability under strong selection is given by $\overline{P} = 1 - e^{-2sN_e/\overline{N}}$, which is nearly $2sN_e/\overline{N}$ when selection is weak but has the advantage of remaining <1 when selection is strong. This approximation is equivalent to the one used when the population size is constant; *i.e.*, $N_e = \overline{N}$ (see Figure 1). This functional form is also suggested by diffusion analysis in populations of large effective size (Kimura 1957, 1964), which assumes weak selection. Simulations confirm that $\overline{P}$ provides a satisfactory approximation for the fixation probability over a range of parameter values in populations undergoing repeated bottlenecks (within a factor of two; supplemental Figure 3 at http://www.genetics.org/supplemental/).

We next consider the time to fixation of a beneficial mutation. If the mutation arises when the population size is N_a and fixes when the population size is N_f, a deterministic model of selection can again be used to predict that

$$T = \ln((N_a - 1)(N_f - 1))/\ln(1 + s) \approx \ln(N_a N_f)/\ln(1 + s).$$

Mutations are more likely to arise when the population size is large, but they are more likely to fix when the population size is small. Averaging the time to fixation over all possible events requires precise knowledge of the fluctuations in population size and the strength of selection. Assuming that mutations arise and fix uniformly over time, however, provides a generic approximation for the time to fixation,

$$\overline{T} \approx \sum_{f=1}^{\tau}\sum_{a=1}^{\tau} \frac{\ln(N_a N_f)}{\tau^2 \ln(1+s)} = \frac{2\ln(N_{gm})}{\ln(1+s)}, \qquad \text{(A1)}$$

where N_{gm} is the geometric mean population size over time. In Equation A1, τ represents the period of the population size cycle if population size changes cyclically. If not, Equation A1 is evaluated by taking the limit as τ goes to infinity. Simulations indicate that $\overline{T}$ provides a satisfactory approximation for the average time to fixation over a range of parameter values in populations undergoing repeated bottlenecks (within a factor of two; supplemental Figure 4 at http://www.genetics.org/supplemental/).

To account for clonal interference, we should determine the expected number of mutations that compete for fixation when the focal mutation appears at time t (see Equation 5) and then average over all possible times at which the focal mutation could arise. To do so exactly requires a precise description of the manner in which the population size fluctuates. As a first-order approximation, we estimate the number of competing mutations using

$$\overline{\lambda}(s) = \overline{PN}\,\mu\frac{\overline{T}}{2}. \qquad \text{(A2)}$$

This approximation ignores the covariance between the number of contending mutations and the time to fixation of a focal mutation, which should be generated by the fluctuations in population size.

Using Equation A2 to rederive Equation 7, the cdf among fixed beneficial mutations becomes

$$H(s) = N_{gm}^{-\overline{N}\mu\kappa/\ln(1+s)}\left(1 - \frac{\kappa}{\Pi}\right), \qquad \text{(A3)}$$

where κ is again given by Equation 8 and the average probability of fixation across the distribution of new mutations is now

$$\Pi = 1 - \left(1 + 2\sigma\,\mathrm{cv}^2\frac{N_e}{\overline{N}}\right)^{-1/\mathrm{cv}^2}. \qquad \text{(A4)}$$

The corresponding probability density function for fixed mutations is $h(s) = dH(s)/ds$.

We assessed the accuracy of Equation A3 against simulations of a population whose size cycles from N_0 to 2^7N_0 via seven doubling events followed by a $1/2^7$ serial dilution. In these simulations, the growth of the population was assumed to be deterministic (no sampling except during the dilution or "bottleneck" generation), and births occurred at a rate proportional to the fitness of an individual. Under this scenario, the size of the bottleneck, N_0, and the period of the cycle, τ, determine $N_e \approx N_0 \ln(2)\tau$ (Wahl *et al.* 2002), $\overline{N} \approx N_0(2^\tau - 1)/\tau$, and $N_{gm} \approx N_0\sqrt{2^{\tau-1}}$ for use in Equation A3. Every combination of the following parameters was explored: selection coefficients ($\sigma = 0.01$, 0.1, 1, 2, and 10), beneficial mutation rate ($\mu = 10^{-7}$ and 10^{-9}), and initial population size ($N_0 = 10^5$, 10^6, and 10^7), assuming that the fitness effects of new mutants were exponential (cv = 1).

Figure A1 indicates that Equation A3 accurately predicts the distribution of fixed selective effects across this range of parameters. Interestingly, Equation 2 of Rozen *et al.* (2002) provides a more accurate prediction of the distribution of fixed beneficial mutations with a fluctuating population size (with N_e in place of N) than with a constant population size (Figure 2). The improved performance of their method is due to the fact that the fixation probability used, $2sN_e/\overline{N}$, remains reasonably accurate even when selection is strong ($s \gg 0.1$) because of the reduction in effective population size caused by the fluctuations ($N_e \ll \overline{N}$).

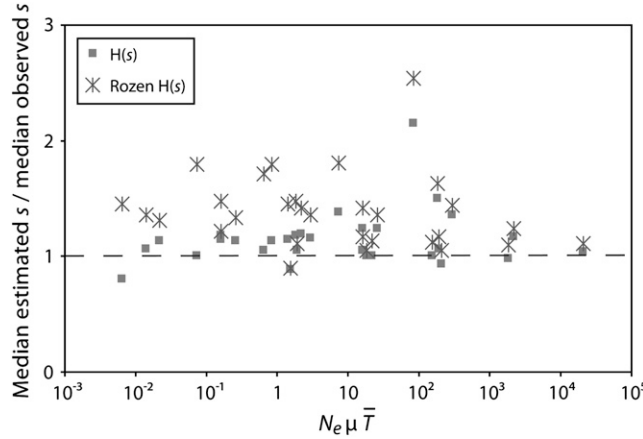

Figure A1.—Median selection coefficient of fixed beneficial mutations estimated from numerical simulations *vs.* analytical results. Median estimated s is given from the results of our fluctuating population size model, given by $H(s)$ in Equation A3, or from those of a model assuming weak selection given by Equation 2 in Rozen *et al.* (2002), but replacing N with N_e from Wahl *et al.* (2002). The horizontal axis measures the number of mutations that appear within the population over the average time to fixation, $\overline{T}$, for a new mutation with selection coefficient given by the median observed s.