The Stochastic Edge in Adaptive Evolution

Éric Brunet,* Igor M. Rouzine[†] and Claus O. Wilke[‡,1]

*Laboratoire de Physique Statistique, École Normale Supérieure, 75230 Paris Cedex 05, France, [†]Department of Molecular Biology
and Microbiology, Tufts University, Boston, Massachusetts 02111 and [‡]Section of Integrative Biology, Institute for Cell and
Molecular Biology, and Center for Computational Biology and Bioinformatics, University of Texas, Austin, Texas 78712

ABSTRACT

In a recent article, Desai and Fisher proposed that the speed of adaptation in an asexual population is determined by the dynamics of the stochastic edge of the population, that is, by the emergence and subsequent establishment of rare mutants that exceed the fitness of all sequences currently present in the population. Desai and Fisher perform an elaborate stochastic calculation of the mean time τ until a new class of mutants has been established and interpret $1/\tau$ as the speed of adaptation. As they note, however, their calculations are valid only for moderate speeds. This limitation arises from their method to determine τ: Desai and Fisher back extrapolate the value of τ from the best-fit class's exponential growth at infinite time. This approach is not valid when the population adapts rapidly, because in this case the best-fit class grows nonexponentially during the relevant time interval. Here, we substantially extend Desai and Fisher's analysis of the stochastic edge. We show that we can apply Desai and Fisher's method to high speeds by either exponentially back extrapolating from finite time or using a nonexponential back extrapolation. Our results are compatible with predictions made using a different analytical approach (Rouzine et al.) and agree well with numerical simulations.

$\mathbf{F}$OR small asexual populations and low mutation rates, the speed of adaptation is primarily limited by the availability of beneficial mutations: a mutation has the time to reach fixation before the next mutation occurs. Therefore, in this case the speed of adaptation increases linearly with population size and mutation rate. By contrast, for large asexual populations or high mutation rates, beneficial mutations are abundant. In this case, the main limit to adaptation is that many beneficial mutations are wasted: when arising on different genetic backgrounds, they cannot recombine and thus are in competition with each other. The theoretical prediction of the speed of adaptation in the latter case is a formidable challenge even for the simplest models. The earliest attempts to predict this speed go back to CROW and KIMURA (1965) and MAYNARD SMITH (1971). In recent years, several groups have improved upon and extended this work (BARTON 1995; TSIMRING et al. 1996; KESSLER et al. 1997; PRÜGEL-BENNETT 1997; GERRISH and LENSKI 1998; ORR 2000; ROUZINE et al. 2003, 2008; WILKE 2004; BEERENWINKEL et al. 2007; DESAI and FISHER 2007). The recent works can be broadly subdivided into two classes: (i) so-called "clonal-interference models" (GERRISH and LENSKI 1998; ORR 2000; WILKE 2004; PARK and KRUG 2007), which emphasize that different beneficial mutations have different-sized effects and that mutations

with large beneficial effects tend to outcompete mutations with small beneficial effects, and (ii) models in which all mutations have the same effect s (TSIMRING et al. 1996; KESSLER et al. 1997; ROUZINE et al. 2003, 2008; BEERENWINKEL et al. 2007; DESAI and FISHER 2007). The latter type of models emphasizes that in large populations, multiple beneficial mutations frequently occur in quick succession on the same genetic background. These models, however, neglect clonal-interference effects.

For the second class of models, where all mutations have the same fitness effect, each individual can be conveniently described by the number k of beneficial mutations it holds. The whole adapting population can then be seen as a traveling wave (TSIMRING et al. 1996; ROUZINE et al. 2003, 2008) moving with time through fitness space toward increasing values of k. In the traveling-wave approach, the bulk of the population, for which each k value is occupied by many individuals, can be accurately described using a deterministic partial differential equation. However, the partial differential equation breaks down for the rare mutants that have the highest fitness in the population, because these rare mutants are subject to substantial genetic drift and stochasticity. Therefore, the description of this stochastic edge must be approached differently and must be coupled with the description of the bulk of the population. Specifically, the deterministic equation admits a traveling-wave solution for any velocity. The high-fitness tail of that solution ends at a finite point, which is identified with the stochastic edge. To select one solu-

[1]Corresponding author: Integrative Biology, #1 University Station—
C0930, University of Texas, Austin, TX 78712.
E-mail: cwilke@mail.utexas.edu

tion (and thus determine the wave speed), ROUZINE *et al.* (2003, 2008) estimated the average size of the stochastic edge using a stochastic argument and matched this size to the solution of the deterministic equation.

Recently, DESAI and FISHER (2007) proposed a new method to calculate the speed of adaptation for the same model. They mainly carry out an elaborate treatment of the stochastic edge, with little attention paid to the bulk of the population. The full-population model is effectively replaced with a two-class model consisting of the best-fit and the second-best-fit classes only; the best-fit class is treated stochastically, whereas the next-best class is assumed to increase exponentially in time due to selection. Beneficial mutations are neglected compared to the effect of selection, except for mutations into the best-fit class. At the very end of the derivation, the sizes of other fitness classes are estimated to provide a normalization condition.

Both ROUZINE *et al.* (2003, 2008) and DESAI and FISHER (2007) calculate the speed of adaptation once the evolutionary process has become stationary, that is, when the combined effects of mutation and selection maintain the shape of the traveling wave. The transient dynamics generally happen on a short timescale but are hard to quantify analytically (TSIMRING *et al.* 1996; DESAI and FISHER 2007).

ROUZINE *et al.* (2003, 2008) define the speed of adaptation as the change of the population's mean number of mutations over time, $V = d\langle k \rangle / dt$. DESAI and FISHER (2007) consider instead the change in the population's mean fitness, $v = sV$. Both approaches consider as an intermediate quantity the lead q, defined as the difference between the number of mutations of the best-fit individuals and the average number of mutations in the population, and write a relation between q and the mean establishment time $\tau = 1/V$ of a new fitness class at the stochastic edge of the population. (Note that ROUZINE *et al.* 2003, 2008 write the lead as $|x_0|$ rather than q and derive a relation between q and V rather than q and τ. Furthermore, k is in these articles the number of *deleterious* rather than beneficial mutations.) We would expect both approaches to make comparable predictions for V, and indeed they do when the speed of adaptation is moderate. For larger speeds, DESAI and FISHER's (2007) derivation is no longer valid (M. M. DESAI, personal communication), and their result gives a prediction that deviates strongly from the one obtained by ROUZINE *et al.* (2008).

Here, our goal is to provide an extensive reanalysis of the approach of DESAI and FISHER (2007) and to extend it to larger speeds. For completeness, we first rederive the relation between q and τ found by DESAI and FISHER (2007) and point out the approximations made in the process. Then, we show in two different ways how we can extend their work to larger speeds of adaptation. With our modifications, the result of DESAI and FISHER (2007) becomes compatible with the result of ROUZINE *et al.* (2003, 2008). Finally, we substantiate our claims with numerical simulations.

MATERIALS AND METHODS

Model assumptions: We consider exactly the same model as that of DESAI and FISHER (2007). Briefly, we model a population of N sequences evolving in continuous time. A sequence with k beneficial mutations has fitness sk (which means we assume there is no epistasis); such a sequence reproduces with rate $1 + sk - \langle sk \rangle$, where $\langle sk \rangle$ is the average fitness in the population. The population size N is held constant at all times by removing one random sequence from the population for every reproduction event. All sequences are equally likely to be chosen for removal and thus have the same average death rate of 1. Mutation events are decoupled from replication events, and we assume that each sequence may independently undergo a mutation with a rate U_b: if a sequence has k beneficial mutations, it is removed with probability $U_b dt$ and replaced by a sequence with $k + 1$ beneficial mutations. Since there are no differences in mutational effects in this model, all sequences with the same number of mutations k can be lumped together into one fitness class, and we refer to the number of sequences with k mutations at time t as $n_k(t)$.

An evolutionary model in which mutation and replication events are decoupled is called a *parallel mutation–selection model* (BAAKE *et al.* 1997). This model has a long-standing tradition in theoretical population genetics (CROW and KIMURA 1970). Even though the alternative model, in which mutation and selection are coupled, may be more appropriate for rapidly evolving viral populations, both models are biologically relevant. Furthermore, in the limit of small s and U_b, which we consider here, the mutation and selection terms decouple, and the two models become equivalent (see, *e.g.*, ROUZINE *et al.* 2003).

When the number $n_k(t)$ of sequences with mutation number k is large enough, the evolution of $n_k(t)$ becomes nearly deterministic:

$$\frac{dn_k(t)}{dt} = (sk - \langle sk \rangle)n_k(t) + U_b[n_{k-1}(t) - n_k(t)]. \quad (1)$$

Sequence classes that satisfy this condition and follow Equation 1 are called *established*. However, the best-fit sequences in the population are not numerous enough for a deterministic description, and stochasticity and genetic drift play an important role in their evolution. We make the approximation that we give the best-fit fitness class [the class corresponding to the largest k with $n_k(t) > 0$] a precise stochastic treatment, while we regard all other classes as established and treat them deterministically. The validity of this approximation has been discussed in detail (ROUZINE *et al.* 2003, 2008; DESAI and FISHER 2007); in particular, it was shown by ROUZINE *et al.* (2008) that this approximation is valid if the speed of adaptation is much larger than U_b. Moreover, we check this approximation numerically in the present work. We refer to the one stochastic class as the *stochastic edge* and denote the value of k for that class by k_0.

Let $\langle k \rangle$ be the mean number of mutations in the population. We define the *lead* q as $q = k_0 - \langle k \rangle$. The lead is the distance from the stochastic edge to the population center. By the definition of fitness in the model, sequences at the stochastic edge have a fitness advantage of sq over the bulk of the population, and sequences in the first-established (*i.e.*, second-best) class have a fitness advantage of $s(q-1)$. Following DESAI and FISHER (2007), we make the approximation that the second-best class behaves deterministically according to Equation 1 and, neglecting incoming mutations from the third-best class and outgoing mutations to the best class, that it grows approximately exponentially with rate $s(q-1)$. (We discuss or check numerically the validity of these approximations later

on.) While the second-best class is growing, any beneficial mutations that occur to sequences in this class feed the best class. Even though any individual mutant that arrives in the best class has a substantial probability of being lost to drift, the ongoing feeding of the best class guarantees that this class itself will become established at some point in time. At this point, the newly established fitness class becomes the second-best class (which, as we assumed, grows deterministically), a new stochastic edge develops at $k_0 + 1$, and the process repeats.

Note that during one cycle, the values of $\langle k \rangle$ and q change smoothly by 1 unit, but we ignore that change and assume that q remains constant from the creation of a new best class to its establishment. Therefore, the whole approach is valid only if q is large enough so that it makes sense to neglect a change of order 1 in q. We assume also that the stochastic edge becomes established when its size gets large enough compared to $1/(sq)$, which is a well-known stochastic threshold (MAYNARD SMITH 1971; BARTON 1995; ROUZINE et al. 2001). This assumption makes sense only if $sq \ll 1$. Finally, we assume that the stochastic edge does not produce any mutant until it is established, which implies $U_b \ll sq$. These conditions imply, of course, $s \ll 1$. Note that q is not a parameter of the model, but a derived quantity. Therefore, all these assumptions must be checked a posteriori once q is computed as a function of the parameters N, s, and U_b.

Simulations: We carried out three types of numeric simulations: fully stochastic whole-population simulations, semideterministic whole-population simulations, and stochastic-edge simulations. The first two ones are simulations of the whole population, whereas the third one is a simulation of the growth of the best-fit class only, assuming it is fed by an exponentially growing second-best-fit class. Details are given below. In all cases, we simulated continuous time by subdividing one generation into small time steps of length δt and updated the simulation after every such time step. In all results reported, δt was at most 0.01.

We used both the GNU Scientific Library (GALASSI et al. 2006) and the library libRmath from the R project (R DEVELOPMENT CORE TEAM 2007) for generation of Poisson, multinomially, and hypergeometrically distributed random numbers. Source code to all simulations is available upon request from C. O. Wilke.

Fully stochastic whole-population simulations: For each fitness class k, we kept track of a random variable $n_k(t)$ representing the class size at time t. In each time step, we first calculated the number of offspring o_k in fitness class k. The o_k are Poisson random variables with mean $n_k(t)(1 + sk - \langle sk \rangle)\delta t$. We then calculated the number of deaths d_k in each class. The total number of deaths $D = \sum_k d_k$ in one time step equals the number of new offspring in that time step, $D = \sum_k o_k$. We generated the d_k's by drawing a single set of multinomially distributed random numbers with means $\langle d_k \rangle = Dn_k(t)/N$ and $\sum_k d_k = D$. If we obtained one or more d_k with $d_k > n_k(t)$, we redrew the entire set of d_k's. We then computed the state of the population after selection but before mutation as $n_k' = n_k(t) + o_k - d_k$. Next, we generated mutations. For each class k, we generated a binomially distributed random variable m_k with mean $\langle m_k \rangle = n_k' U_b \delta t$ and n_k' trials. We then updated the population to $n_k(t + \delta t) = n_k' + m_{k-1} - m_k$.

The usage of the multinomial distribution to generate d_k's is an approximation, as the distribution of the d_k's is actually hypergeometric. [The hypergeometric distribution describes the probability $p_{hg}(d, D; n, N - n)$ to obtain d white balls after D random draws from an urn containing n white and $N - n$ black balls and is given by $p_{hg}(d, D; n, N - n) = \binom{n}{d}\binom{N-n}{D-d}/\binom{N}{D}$.] We also implemented hypergeometric sampling of deaths, by generating the random variables d_k one by one, going from the

best-fit class to the worst-fit class with the probabilities $\text{Prob}(d_k) = p_{hg}[d_k, D - \sum_{i > k} d_i, n_k(t), \sum_{i < k} n_i(t)]$. We found that the generation of hypergeometrically distributed random variables was much slower than multinomial sampling (up to a factor of 1000) and caused numeric instabilities at large $N \gtrsim 10^8$, even when using an efficient numerical algorithm (KACHITVICHYANUKUL and SCHMEISER 1985). For $N < 10^8$, simulation results with multinomial sampling of deaths and hypergeometric sampling of deaths were virtually identical.

We measure the speed of adaptation V for the stationary evolutionary process, when the population can be considered a traveling wave (ROUZINE et al. 2003, 2008). We know of no good theory for predicting how long it takes for the population to reach stationarity, but simulations indicate that equilibration proceeds rapidly (see also TSIMRING et al. 1996). In our simulations, we considered the population as equilibrated when at least 10 new fitness classes had been established. We then measured the time Δt it took the population to establish 40 additional fitness classes and calculated V as $40/\Delta t$. We averaged V over 10 independent replicates.

Semideterministic simulations: For each fitness class k, we kept track of a variable $n_k(t)$ representing the class size at time t. We updated the size of the stochastic edge class $n_{k_0}(t)$ stochastically and all other variables $n_k(t)$ deterministically. As in the case of the fully stochastic simulations, in each time step we first calculated the number of offspring o_k in fitness class k. For $k < k_0$, $o_k = n_k(t)(1 + sk - \langle sk \rangle)\delta t$. At the stochastic edge, o_{k_0} is a Poisson-distributed random variable with mean $n_{k_0}(t)$ $(1 + sk - \langle sk \rangle)\delta t$. We then calculated the number of deaths d_k. The total number of deaths required is $D = \sum_k o_k$. At the stochastic edge, d_{k_0} is a Poisson-distributed random variable with mean $Dn_{k_0}(t)/N$. We set $d_{k_0} = n_{k_0}$ if $d_{k_0} > n_{k_0}$. For $k < k_0$, we calculated $d_k = (D - d_{k_0})n_k(t)/(N - n_{k_0})$. We then computed the state of the population after selection but before mutation as $n_k' = n_k(t) + o_k - d_k$. Next, we generated mutations. At the stochastic edge, $m_{k_0} = 0$ (the stochastic edge does not produce beneficial mutations). For the second-best class, $m_{k_0 - 1}$ is a Poisson-distributed random variable with mean $n_{k_0 - 1}' U_b \delta t$. For all other $k < k_0 - 1$, $m_k = n_k' U_b \delta t$. We then updated the population to $n_k(t + \delta t) = n_k' + m_{k-1} - m_k$. [In theory, this procedure can lead to a negative $n_{k_0 - 1}(t + \delta t)$. However, this extremely unlikely event never actually occurred in our simulations.] Finally, if $n_{k_0}(t) > 1/(sq)$, we designated the current stochastic edge class as established and set k_0 to $k_0 + 1$.

We measured the speed of adaptation as in the fully stochastic full-population simulations.

Stochastic-edge simulations: We kept track of a single random variable $n(t)$ representing the best-fit class in the population (the stochastic edge), which was set to zero at $t = 0$. Assuming that the population of the second-best-fit class was $e^{s(q-1)t}/(sq)$, we generated at each time step three Poisson random variables o, d, and m, representing the number of offspring, deaths, and incoming mutations in the best-fit class, with means $(1 + sq)n(t)\delta t$, $n(t)\delta t$, and $U_b e^{s(q-1)t}\delta t/(sq)$, respectively, and updated $n(t)$ as $n(t + \delta t) = n(t) + o - d + m$. All measures reported in RESULTS AND DISCUSSION were obtained by averaging over 500 independent realizations of the simulation.

RESULTS AND DISCUSSION

Rederivation of the Desai–Fisher results: In this section, we rederive the main results of DESAI and FISHER (2007), using methods very similar to theirs, but with some simplifications. This section does not contain any new results; it is included here because we need to point

out the various approximations made by DESAI and FISHER (2007) before discussing them and also because we believe that an alternative presentation of their nontrivial results may be helpful to many readers.

DESAI and FISHER (2007) define the *establishment time* τ as the time from the establishment of one new fitness class to the establishment of the next better fitness class. Their approach is based on an elaborate probabilistic calculation of the establishment time τ of a fitness class with advantage sq, given that this class is fed beneficial mutations from the exponentially growing second-best class. Since for large N every beneficial mutation that arrives in the best-fit class forms a clone that is independent of all other clones in the best-fit class, the growth (and potential establishment or extinction) of a single clone can be described using continuous-time branching theory. The treatment of a single clone is standard (ATHREYA and NEY 1972); the single clone follows a birth-and-death process with birth rate $1 + sq$ and death rate 1. The probability-generating function for the size $m(t)$ of a clone at time t that had size 1 at time $t = 0$ is given by ATHREYA and NEY (1972):

$$G(z,\ t) = \langle z^{m(t)} \rangle = \frac{(z-1)(1 - e^{sqt}) + zsq}{(z-1)[1 - (1 + sq)e^{sqt}] + zsq}. \quad (2)$$

Given this result for a stochastically growing individual clone, we now wish to study the size $n(t)$ of the best-fit class at time t. This class grows by itself with a rate sq and is fed by the next-best class, which grows at a rate $s(q - 1)$. We call $f(t)$ the size of the next-best class, and we assume later on that

$$f(t) = \frac{1}{sq}e^{s(q-1)t}. \quad (3)$$

Note that we assume a deterministic growth for this next-best class, and we neglect changes in $f(t)$ due to both outflow of beneficial mutations from the second-best to the best-fit class and inflow of beneficial mutations from the third-best to the second-best class. We set the origin of time such that $t = 0$ when $f(t) = 1/(sq)$, which is the well-known stochastic threshold for a clone with fitness advantage sq (MAYNARD SMITH 1971; BARTON 1995; ROUZINE *et al.* 2001): a clone whose size far exceeds this threshold grows essentially deterministically, whereas a clone whose size falls far below this threshold is subject to genetic drift. The idea is that this second-best-fit class just got established at time $t = 0$ and was the previous stochastic best-fit class at times $t < 0$.

As the size $n(t)$ of the best-fit class grows with rate sq, DESAI and FISHER (2007) suggest to write for large t

$$n(t) = \frac{1}{sq}e^{sq(t-\tau)}, \quad (4)$$

where τ is some random variable. Intuitively, we might interpret τ as the time at which the new best-fit class *appears* to have reached the stochastic threshold, when

sampled at later times when it is already deterministic. Of course, $n(t)$ is really a random variable that in the $t \to \infty$ limit converges to an exponential growth, and $n(\tau)$ has no reason to be equal to $1/(sq)$. But if $n(t)$ had a deterministic exponential growth at all times, then τ would be the time for which $n(t)$ had reached $1/(sq)$. In this case, it would take a (random) time τ to build a new established class from a class that had just crossed the stochastic threshold, the population would move by one fitness class during a time interval τ, and the speed of adaptation would be simply $1/\langle\tau\rangle$, where $\langle\tau\rangle$ is the average of τ. However, things are more complicated than this intuitive picture suggests. In Equation 4, the random variable τ has a distribution that depends on the time t at which $n(t)$ is measured and we from now on write $\tau(t)$ rather than simply τ. When evaluating the speed of adaptation $1/\langle\tau(t)\rangle$, we have to choose a time t at which the average is taken. DESAI and FISHER (2007) chose to take $t = \infty$ and, therefore, to define the establishment time as $\langle\tau(\infty)\rangle$. This choice of t is arbitrary, and we argue later on that it makes more sense to choose t on the order of $\langle\tau\rangle$. As we shall see, when q is large, $\langle\tau(t)\rangle$ converges quite slowly to $\langle\tau(\infty)\rangle$, and the two expressions give different results. For this reason, DESAI and FISHER (2007) considered only moderately large q, for which $\langle\tau(\infty)\rangle$ is a good approximation of $\langle\tau(t)\rangle$.

We put aside for the moment the problem of choosing the best value of t in $\langle\tau(t)\rangle$ and focus on calculating the cumulants of $\tau(t)$. We first calculate the probability-generating function $\langle z^{n(t)} \rangle$ of $n(t)$, under the assumption that the best-fit class is fed by mutations from the second-best class, which itself has size $f(t)$ at time t. In this calculation, we neglect beneficial mutations produced by the best-fit class, as mutations are rare and the number of sequences in this class is small. Assuming that the process starts at some time T_0 with $n(T_0) = 0$, we write $n(t) = \sum_{T_0 < t' < t} m_{t'}(t)$, where $m_{t'}(t)$ is the contribution at time t of a new clone if such a clone appeared at time t'. With a probability $f(t')U_b dt'$, a clone actually appeared at time t' and, according to Equation 2, we have $\langle z^{m_{t'}(t)} \rangle = G(z,\ t - t')$. With a probability $1 - f(t')U_b dt'$, no clone appeared at time t', we have $m_{t'}(t) = 0$, and, of course, $\langle z^{m_{t'}(t)} \rangle = 1$. As all the $m_{t'}(t)$ for a given t are independent random numbers, we can write $z^{n(t)} = \prod_{T_0 < t' < t} z^{m_{t'}(t)}$ and average independently all the terms in the product. We obtain

$$\langle z^{n(t)} \rangle = \prod_{T_0 < t' < t} (f(t')U_b dt'\, G(z,\ t - t') + [1 - f(t')U_b dt'])$$

$$= \exp\left[\sum_{T_0 < t' < t} \ln(1 + dt'\, U_b f(t')[G(z,\ t - t') - 1]) \right]. \quad (5)$$

As dt' is infinitely small, we have $\ln(1 + dt'C) = dt'C$, and we recognize that the summation is actualy an integral. Therefore,

$$\langle z^{n(t)}\rangle = \exp\left(U_{\rm b}\int_{T_0}^{t} dt' f(t')[G(z,\, t-t')-1]\right) \qquad (6)$$

$$= \exp\left(U_{\rm b}\int_{0}^{t-T_0} dt' f(t-t')[G(z,\, t')-1]\right). \quad (7)$$

This equation with $T_0=-\infty$ corresponds to Equation 24 of DESAI and FISHER (2007).

In Equation 7, we use the $G(z,\, t)$ given in Equation 2 and the $f(t)$ given in Equation 3, change the variable of integration to $v=(1+sq)(1-z)e^{sqt'}/[zsq-(1-z)]$, and obtain

$$\langle z^{n(t)}\rangle = \exp\left[-\frac{U_{\rm b}\big[(zsq/(1-z)-1)e^{-sqt}\big]^{(1/q)-1}}{sq(1+sq)^{1/q}}\right.$$
$$\left. \times \int_{(1-z)(1+sq)/(zsq-(1-z))}^{((1-z)(1+sq)/(zsq-(1-z)))e^{sq(t-T_0)}} \frac{v^{(1/q)-1}\,dv}{v+1}\right]. \quad (8)$$

(This equation corresponds to Equation 27 of DESAI and FISHER 2007. Note that because of the way the change of variable was done, it is correct only for $zsq > 1-z$.) To compute the cumulants of $\tau(t)$, it is easier to rewrite Equation 4 as

$$n(t) = \frac{1}{sq}x(t)e^{sqt} \qquad (9)$$

with the random variable $x(t)=e^{-sq\tau(t)}$; the cumulants of $\ln x(t)$ differ from the cumulants of $\tau(t)$ only by a constant multiplicative factor. We obtain the generating function of $x(t)$ from Equation 8 using $\langle e^{-\lambda x(t)}\rangle = \langle z^{n(t)}\rangle$ for $z=\exp(-\lambda sqe^{-sqt})$. For now, we are interested only in the limit of infinite time. Making the substitution for z and taking the limit $t\to\infty$ while holding λ constant, we find

$$\langle e^{-\lambda x(\infty)}\rangle = \exp\left[-\frac{U_{\rm b}}{sq}\frac{\lambda^{1-1/q}}{(1+sq)^{1/q}}\times\int_{0}^{\lambda(1+sq)e^{-sqT_0}} \frac{v^{(1/q)-1}\,dv}{v+1}\right]. \qquad (10)$$

In this expression, T_0 is the starting time at which the second-best class begins feeding the best-fit class. The second-best class can start producing mutants when its size is of order 1, which happens at large negative times. Unfortunately, Equation 10 is, strictly speaking, not valid if $T_0<0$, as we obtained it by using, in Equation 6, the expression Equation 3 for the size of the second-best-fit class, which is correct only for $t>0$. However, as we assumed $U_{\rm b}/(sq)\ll 1$, the mutation events from the second-best class at any negative time are very rare, so that we may expect that the final result will be dominated only by the events with $t>0$ and that it will not depend much on the value of T_0, as long as T_0 is a

negative number. One way to check the validity of this assumption is to verify that we reach the same results for $T_0=-\infty$ (equivalent to the assumption that Equation 3 is a good approximation for the size of the second-best class at negative times) and for $T_0=0$ (equivalent to the assumption that the second-best class is empty at negative times). Therefore, we first follow DESAI and FISHER (2007) by taking the limit $T_0\to-\infty$, and, at the end of this section, we consider briefly the case $T_0=0$ to validate this approximation. For $T_0=-\infty$, using $\int_{0}^{\infty} dv\, v^{(1/q)-1}/(v+1)=\pi/\sin(\pi/q)$, we obtain

$$\langle e^{-\lambda x(\infty)}\rangle = \exp(-b\lambda^{1-1/q}) \qquad (11)$$

with

$$b = \frac{\pi U_{\rm b}}{sq(1+sq)^{1/q}\sin(\pi/q)} \approx \frac{\pi U_{\rm b}}{sq\sin(\pi/q)}. \qquad (12)$$

(The first expression for b in Equation 12 is exact, but we use only the second, approximate expression in the rest of the article as we need $s\ll 1$ anyway in the biological applications of that model. In fact, we often use $b\approx U_{\rm b}/s$ when we suppose $q\gg 1$.)

Equation 11 is the generating function of $x(\infty)$, but we need the generating function of $\ln x(\infty)$. For any random variable x, we can turn the former into the latter using the following identity, which is valid for $\mu<0$ and follows from the definition of the Gamma function:

$$\langle x^{\mu}\rangle = \frac{1}{\Gamma(-\mu)}\int_{0}^{\infty} d\lambda\, \lambda^{-\mu-1}\langle e^{-\lambda x}\rangle. \qquad (13)$$

(Actually, the equality holds without the averages.) Then, expanding $\ln\langle x^{\mu}\rangle$ in powers of μ allows us to recover all the cumulants of $\ln x$:

$$\ln\langle x^{\mu}\rangle = \mu\langle\ln x\rangle + \frac{\mu^2}{2}\mathrm{Var}[\ln x] + \mathcal{O}(\mu^3). \qquad (14)$$

Alternatively, if all we need is $\langle\ln x\rangle$, we can integrate by part $\lambda^{-\mu-1}$ in Equation 13 (assuming that $\langle e^{-\lambda x}\rangle$ goes to 0 for large λ) and expand directly to the first order in μ. We obtain

$$\langle\ln x\rangle = -\gamma + \int_{0}^{\infty} d\lambda\ln(\lambda)\frac{d}{d\lambda}\langle e^{-\lambda x}\rangle, \qquad (15)$$

where $\gamma=-\Gamma'(1)\approx 0.5772$ is the Euler gamma constant. Applying this procedure to the random variable $x(\infty)$, we get from Equations 13 and 11

$$\langle x(\infty)^{\mu}\rangle = \frac{1}{\Gamma(-\mu)}\int_{0}^{\infty} d\lambda\, \lambda^{-\mu-1}\exp(-b\lambda^{1-1/q})$$
$$= \frac{\Gamma(1-\mu q/(q-1))}{\Gamma(1-\mu)}b^{\mu q/(q-1)}. \qquad (16)$$

Making use of the expansion $\ln\Gamma(1-\varepsilon)=\gamma\varepsilon+(\pi\varepsilon)^2/12+\mathcal{O}(\varepsilon^3)$, we obtain from Equation 14

$$\langle \ln x(\infty) \rangle = \frac{q}{q-1} \ln(be^{\gamma/q}), \qquad (17)$$

$$\mathrm{Var}[\ln x(\infty)] = \frac{\pi^2}{6}\left[\left(\frac{q}{q-1}\right)^2 - 1\right]. \qquad (18)$$

Converting $\ln x(\infty)$ back into $\tau(\infty)$, we arrive at our final expressions

$$\langle \tau(\infty) \rangle = \frac{1}{s(q-1)} \ln\left(\frac{1}{be^{\gamma/q}}\right)$$

$$\approx \frac{1}{s(q-1)} \ln\left(\frac{sq\sin(\pi/q)}{U_{\mathrm{b}}\pi e^{\gamma/q}}\right) \qquad (19)$$

and

$$\mathrm{Var}[\tau(\infty)] = \frac{\pi^2}{6}\left[\frac{1}{[s(q-1)]^2} - \frac{1}{(sq)^2}\right]. \qquad (20)$$

We emphasize that these quantities were obtained in the limit $t \to \infty$.

When we compare our results for mean and variance of $\tau(\infty)$ to the results of DESAI and FISHER (2007), we find that our expression for the variance agrees with their Equation 37. Our expression for $\langle \tau(\infty) \rangle$ is similar to their Equation 36, except that the factor q in the logarithm was accidently replaced by a factor $q-1$ in DESAI and FISHER (2007) (M. DESAI, personal communication). As in DESAI and FISHER (2007), we neglected the factor $(1 + sq)^{1/q}$ in the expression (12) of b as we need $s \ll 1$ in the context of the full biological model.

We now consider what happens if we use $T_0 = 0$ (and $sq \ll 1$) in Equation 10 instead of $T_0 = -\infty$. Clearly, for large q, this integral is dominated by small v, so the value of the upper bound should not matter much to the final result. Indeed, if λ is not too small, we have

$$\int_0^{\lambda} \frac{v^{(1/q)-1}dv}{v+1} = \int_0^{\infty} \frac{v^{(1/q)-1}dv}{v+1} - \int_{\lambda}^{\infty} \frac{v^{(1/q)-1}dv}{v+1}$$

$$\approx \frac{\pi}{\sin(\pi/q)} - \ln\left(1 + \frac{1}{\lambda}\right) \approx \frac{\pi}{\sin(\pi/q)}. \quad (21)$$

We neglected $v^{1/q}$ in the last integral, which is valid if λ is large enough, namely if either $\lambda > 1$ or $-\ln \lambda \ll q$. The same condition on λ allows the last simplification in Equation 21. Therefore, the generating function Equation 10 is identical for $T_0 = 0$ or $T_0 = -\infty$, except for very small λ, and the probability distribution function of $x(\infty)$ does not depend on T_0 except for very large values of $x(\infty)$ such that $\ln x(\infty) \gg q$. As it is easy to check from Equation 11 that Equation 15 is dominated by values of λ of order $1/b \approx s/U_{\mathrm{b}}$, we finally obtain that the result for $\langle \ln x(\infty) \rangle$ and hence $\langle \tau(\infty) \rangle$ is approximately the same for $T_0 = 0$ or $T_0 = -\infty$ if either $s/U_{\mathrm{b}} > 1$ or $q \gg \ln(U_{\mathrm{b}}/s)$. As DESAI and FISHER (2007) assumed $s/U_{\mathrm{b}} \gg 1$ in their work, their approximation of taking $T_0 = -\infty$ is justified.

The stochastic threshold $1/(sq)$, which enters the calculation through Equations 3 and 4, is central to the entire approach of DESAI and FISHER (2007). Yet it is not exactly true that all classes that exceed this threshold will behave purely deterministically. All we know for certain is that classes whose size is $\gg 1/(sq)$ will generally not show much stochasticity. Therefore, it is reasonable to ask what would happen if the true threshold were actually $\alpha/(sq)$, with some constant $\alpha \neq 1$. We find that replacing the threshold $1/(sq)$ in Equations 3 and 4 by $\alpha/(sq)$ results only in an additional factor $\alpha^{-1/q}$ inside the large logarithm in the final result for $\langle \tau(\infty) \rangle$, Equation 19. Because q is assumed to be large, the effect of that change is minor.

As a side matter, note that the generating function Equation 11 describes a distribution with a long tail; in particular, the average of $x(\infty)$ is infinite, which is not biologically possible and is an artifact of taking $T_0 = -\infty$. If we were interested in the average of $x(\infty)$, we would need to keep T_0 finite and we would obtain, after some algebra, $\langle x(\infty) \rangle = (U_{\mathrm{b}}/s)e^{-sT_0}$.

The case of large q: In general, a weak selective pressure ($s \ll 1$) results in a broad fitness distribution, $q \gg 1$. To gain better insight into the predictions of Equation 19 for this case, we consider the limit q large and s small. We find

$$\langle \tau(\infty) \rangle \approx \frac{1}{sq} \ln(s/U_{\mathrm{b}}). \qquad (22)$$

By studying the deterministic evolution of the bulk of the population in the same $q \gg 1$ limit, ROUZINE *et al.* (2008) obtained in their Equation 39 a relation very similar to Equation 22. Using τ and q instead of the notations $V = 1/\tau$ and $x_0 = -q$ of the cited work, we can write their result as

$$\tau = \frac{1}{sq}\left[\ln(1/(U_{\mathrm{b}}\tau)) - 1\right] = \frac{1}{sq}\left[\ln\left(\frac{sq}{U_{\mathrm{b}}|\ln(eU_{\mathrm{b}}\tau)|}\right) - 1\right]$$

$$= \frac{1}{sq}\left[\ln(sq/U_{\mathrm{b}}) - 1 - \ln|\ln(eU_{\mathrm{b}}\tau)|\right]. \qquad (23)$$

Ignoring subleading corrections, we find that the main difference between Equation 22 and Equation 23 is a term q within the logarithm, which can become large in some situations.

We claim that when q is large, Equation 22 is not an accurate prediction for the mean establishment time. In particular, we obtain $\langle \tau(\infty) \rangle < 0$ for $s < U_{\mathrm{b}}$. (Note that we assume throughout this work that $sq \gg U_{\mathrm{b}}$, but unlike DESAI and FISHER 2007, we do not require $s > U_{\mathrm{b}}$.) This result is problematic, because the whole point of this calculation was to interpret $\langle \tau(\infty) \rangle$ as the mean time between the establishment of a best-fit class and the establishment of the next best-fit class in the full model describing a population of N sequences. Clearly, the establishment time in the full model cannot be negative, and this result would seem to suggest that the whole approach of approximating the full model by the sole

behavior of its stochastic edge does not work for large values of q. However, we believe that the method can be fixed by replacing some of the assumptions that led to Equation 19 by improved and more accurate assumptions.

Approximations made in Desai and Fisher's approach: DESAI and FISHER (2007) made several approximations to obtain the relation Equation 19 between the establishment time and the lead q:

1. All the classes are evolving deterministically, except the stochastic edge.
2. The lead q does not vary in time between the creation and the establishment of a new mutant class.
3. The stochastic edge does not produce any mutant until it is established.
4. The second-best-fit class has an exactly exponential growth with a rate $s(q-1)$, as in Equation 3.
5. One can take the limit $T_0 \to -\infty$ when evaluating the mean establishment time.
6. At large times, the size $n(t)$ of the stochastic edge is well fit by an exponential growth of rate sq, as in Equation 4 or Equation 9.
7. One can interpret the establishment time as $\langle \tau(t) \rangle$ for $t \to \infty$.

DESAI and FISHER (2007) discussed the validity of these approximations in the context of their parameter range of interest, *i.e.*, for moderate q (see their Appendixes E–G). We reevaluate the approximations here in the context of large q.

ROUZINE *et al.* (2003, 2008) gave detailed analytical arguments why approximation 1 is valid for $q \gg U_b/s$. In our work, we verify this approximation numerically, using the semideterministic full-population simulation. Getting rid of this approximation and treating all classes stochastically is a formidable mathematical challenge that would be of limited interest because the approximation is quite good.

Approximations 2 and 3 are valid in, respectively, the limits $q \gg 1$ and $sq \gg U_b$, which we have assumed throughout. For more moderate values of q (between 2 and 5), DESAI and FISHER (2007) discussed the validity of approximation 2 in their Appendix H.

Approximation 4 is more problematic. Saying that the second-best-fit class grows exponentially implies that we are ignoring the contribution from mutations originating in the third-best-fit class. On one hand the mutation rate U_b is supposed to be small compared to the effect of selection $s(q-1)$, but on the other hand the third-best-fit class is much larger than the second-best class. In APPENDIX A, we present an argument indicating that approximation 4 is justified only at smaller times and is incorrect by a large factor for values of t close to the establishment time, which is unfortunately precisely the time at which most of the mutations occur. To what extent this deviation from approximation 4 affects our final result is difficult to assess at this point. Improving

upon this approximation would require having a theory of at least the third-best-fit class.

We have already discussed the validity of approximation 5.

Approximations 6 and 7 are closely related: one can always decide to write Equation 4 for a well-chosen time-dependent random variable $\tau(t)$. But saying that $n(t)$ is well fit by an exponential (approximation 6) is then equivalent to saying that $\tau(t)$ actually does not depend too much on time and that, consequently, one can choose any value of t to evaluate the establishment time, including $t = \infty$ (approximation 7). But, as we now argue, $n(t)$ is not well fit by an exponential growth for large q. This implies that $\tau(t)$ has a strong t dependence and that choosing the best value of t when evaluating the establishment time is important; we argue that the proper value of t is of the order of the establishment time. Alternatively, one can get rid of approximation 6 and replace Equation 4 by a better fit of $n(t)$. When carrying out this procedure, we find that the new random variable τ has indeed a weak time dependence and taking the limit $t \to \infty$ makes sense. We presently explore both possible improvements.

Finite extrapolation time: We are still fitting the best-fit class by an exponential, as in Equation 4 or Equation 9, but this time we try to evaluate $\langle \tau(t) \rangle$ for some finite time t. We go back to Equation 8, set $T_0 = -\infty$, and substitute $z = \exp(-\lambda sq e^{-sqt})$ as before. However, this time we keep all terms to the first order in e^{-sqt}. We find

$$\langle e^{-\lambda x(t)} \rangle = \exp\left[-\frac{U_b \lambda^{1-1/q}}{sq(1+sq)^{1/q}}\left[1 + \lambda\frac{q-1}{q}\left(\frac{sq}{2}+1\right)e^{-sqt}\right]\right.$$
$$\left. \times \int_{(1+sq)\lambda e^{-sqt}}^{\infty} \frac{v^{1/q-1}dv}{v+1}\right]. \tag{24}$$

For $\lambda e^{-sqt} \ll 1$, the integral assumes the value

$$\frac{\pi}{\sin(\pi/q)} - q(1+sq)^{1/q}\lambda^{1/q}e^{-st} + \mathcal{O}(e^{-s(q+1)t}). \tag{25}$$

Note that the small term appearing in the integral of Equation 24 is e^{-sqt}, but the first-order correction for finite time is actually proportional to e^{-st}, which is much larger. Compared to this correction, we neglect the term proportional to e^{-sqt} before the integral in Equation 24 and find

$$\langle e^{-\lambda x(t)} \rangle \approx \exp\left[-b\lambda^{1-1/q} + \frac{U_b}{s}\lambda e^{-st}\right] \quad \text{for } \lambda e^{-sqt} \ll 1. \tag{26}$$

(DESAI and FISHER 2007 write a similar expression in their Equation G2, but do not exploit it.) When λe^{-sqt} is not small but q is large, we can obtain another expression by neglecting the $v^{(1/q)}$ in Equation 8. Assuming sq small, we obtain after some algebra

$$\langle e^{-\lambda x(t)}\rangle \approx \exp\left[-\frac{U_{\mathrm{b}}}{sq}e^{-s(q-1)t}\times\frac{\lambda e^{-sqt}}{\lambda e^{-sqt}-1}\ln(\lambda e^{-sqt})\right]$$
$$\text{for } e^{-q}\ll\lambda e^{-sqt}\ll 1/(sq). \tag{27}$$

Note that Equations 26 and 27 are both valid in the range $e^{-q}\ll\lambda e^{-sqt}\ll 1$.

We want, as before, to compute $\langle x(t)^{\mu}\rangle$ by using Equation 26 into Equation 13. Expanding inside the integral in powers of the small parameter $\exp(-st)$, we would get

$$\langle x(t)^{\mu}\rangle\approx\frac{1}{\Gamma(-\mu)}\sum_{n\geq 0}\frac{1}{n!}\left(\frac{U_{\mathrm{b}}}{s}\right)^{n}e^{-nst}\int_{0}^{\infty}d\lambda\,\lambda^{-\mu-1}$$
$$\times\exp\left[-b\lambda^{1-1/q}\right]\lambda^{n}, \tag{28}$$

but writing this equation is not justified *a priori*, because we may not use Equation 26 for arbitrarily large λ, and Equation 28 is actually a divergent series. We show, however, that the first terms of that series are nevertheless correct. Indeed, the integral in the nth order term of the series Equation 28 is mainly contributed from values of λ of order $n/b\approx ns/U_{\mathrm{b}}$. (This result is obtained by looking at the maximum of the integrand, in the limit of large q and large n.) Given the validity range of Equation 26, this means that the series Equation 28 is correct up to $n=n_{\max}$ with $n_{\max}\ll(U_{\mathrm{b}}/s)e^{sqt}$. With this condition in mind, we compute the integrals and find

$$\langle x(t)^{\mu}\rangle\approx\frac{\Gamma(1-\mu q/(q-1))}{\Gamma(1-\mu)}b^{\mu q/(q-1)}$$
$$\times\left[1-\sum_{n=1}^{n_{\max}}\frac{1}{n!}\left(\frac{U_{\mathrm{b}}}{s}\right)^{n}e^{-nst}\frac{\mu}{n-\mu}\right.$$
$$\left.\times\frac{\Gamma(1-(n-\mu)q/(q-1))}{\Gamma(1-\mu q/(q-1))}b^{-(nq/(q-1))}\right]+o(e^{-n_{\max}st}). \tag{29}$$

Using $\langle\ln x\rangle=\lim_{\mu\to 0}(1/\mu)\ln\langle x^{\mu}\rangle$ as before (see Equation 14), we arrive at

$$\langle\ln x(t)\rangle\approx\frac{q}{q-1}\ln(be^{\gamma/q})-\sum_{n=1}^{n_{\max}}\frac{\Gamma(1+nq/(q-1))}{nn!}\left(\frac{U_{\mathrm{b}}}{s}\right)^{n}$$
$$\times b^{-(nq/(q-1))}e^{-nst}+o(e^{-n_{\max}st}). \tag{30}$$

The first term corresponds to the result for $t\to\infty$, Equation 17, while the second term gives a correction for finite time. This expression can be simplified further for large q by using $b\approx U_{\mathrm{b}}/s$ and $\Gamma(1+nq/(q-1))\approx n!$, where the latter simplification is valid only if $n_{\max}\ln n_{\max}\ll q$. We recognize then the expansion of $\ln$ and obtain

$$\langle\ln x(t)\rangle\approx\ln(U_{\mathrm{b}}/s)+\ln\left[1-e^{-st+(1/q)\ln(s/U_{\mathrm{b}})}\right]$$
$$+o(e^{-n_{\max}st}), \tag{31}$$

where we recall that $n_{\max}$ is such that $n_{\max}\ln n_{\max}\ll q$ and $n_{\max}\ll(U_{\mathrm{b}}/s)e^{sqt}$. Furthermore, the $o(e^{-n_{\max}st})$ is indeed a small correction only if $n_{\max}st\gg 1$. Therefore, Equation 31 is valid only if $q\gg 1$ (from the first condition above) and $tU_{\mathrm{b}}e^{sqt}\gg 1$ (from the second and third conditions). [We made some simplifications using $q\gg 1$ to reach Equation 31, but as we will see, we need to keep the term $(1/q)\ln(s/U_{\mathrm{b}})$ given the relevant values of t.] In terms of $\tau(t)$, we finally get

$$\langle\tau(t)\rangle\approx\frac{1}{sq}\ln\left[\frac{s/U_{\mathrm{b}}}{1-e^{-st+(1/q)\ln(s/U_{\mathrm{b}})}}\right] \tag{32}$$

for sufficiently large q and t.

Another way to reach Equation 32 is to use the integral expression Equation 15 to compute $\langle\ln x(t)\rangle$. Making the change of variable $y=\lambda e^{-sqt}$, we find

$$\langle\ln[x(t)]\rangle=-sq\langle\tau(t)\rangle=-\gamma-sqt+\int_{0}^{\infty}dy\ln(y)\frac{d}{dy}\langle e^{-\lambda x(t)}\rangle. \tag{33}$$

We rewrite $\langle e^{-\lambda x(t)}\rangle$ from Equation 26 as a function of y,

$$\langle e^{-\lambda x(t)}\rangle\approx\exp\left[-R(t)\times qy^{1-1/q}(1-y^{1/q})\right]\quad\text{for } y\ll 1,$$

with

$$R(t)=\frac{U_{\mathrm{b}}}{sq}e^{s(q-1)t}\quad\text{and}\quad y=\lambda e^{-sqt}. \tag{34}$$

(We assumed q large and used $b\approx U_{\mathrm{b}}/s$.) In fact, without any approximation, one can check that $\langle e^{-\lambda x(t)}\rangle$ can be written as $\exp[-R(t)\times F(y)]$, where the function $F(y)$ has no explicit dependency on λ or t. Clearly, as y is proportional to λ and $\langle e^{-\lambda x(t)}\rangle$ decreases with λ, the function $F(y)$ is an increasing function of y. [See, for instance, Equation 27, which shows how $F(y)$ increases for $y\ll 1/(sq)$.] Moreover, despite the presence of the large parameter q, this function varies neither slowly nor rapidly with y, so that the speed with which $\langle e^{-\lambda x(t)}\rangle$ changes with y depends only on the magnitude of $R(t)$. When $R(t)$ is large, $\langle e^{-\lambda x(t)}\rangle$ interpolates very quickly between 1 and 0, and its derivative can be approximated by a delta function. This interpolation occurs at some value y_{c} of y that is very small; hence it is justified to use Equation 34 to compute y_{c}. Moreover, for $R(t)$ large enough, $\langle e^{-\lambda x(t)}\rangle$ becomes negligibly small within the range of validity of Equation 34 and will go on decreasing for larger values of y [because $F(y)$ is an increasing function] so that values of y outside the validity range of Equation 34 do not contribute to the integral. All these remarks allow us to compute the integral in Equation 33; we find, for $R(t)\gg 1$,

$$-sq\langle\tau(t)\rangle\approx-\gamma-sqt-\ln y_{\mathrm{c}},$$
$$R(t)\times qy_{\mathrm{c}}^{1-1/q}(1-y_{\mathrm{c}}^{1/q})\approx 1\quad\text{with } y_{\mathrm{c}}\ll 1. \tag{35}$$

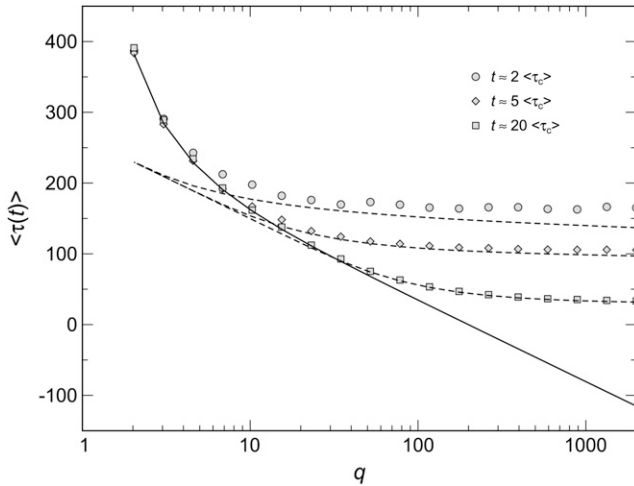Eliminating y_{c} in the previous equation gives

FIGURE 1.—Numerical evaluation of the average $\langle \tau(t) \rangle$ in simulations of the stochastic edge, as a function of q, for $U_b = 10^{-4}$ and $sq = 0.02$ held constant throughout. Points are simulation results; the standard error from the simulations is smaller than the symbol size. As measurement times t, we used three multiples of $\langle \tau_c \rangle$ and determined $\langle \tau_c \rangle$ from the approximation formula Equation 57. The dashed lines were calculated from Equation 32 (valid only for large q). The solid line represents $\langle \tau(\infty) \rangle$ (Equation 19, valid for all q).

$$\langle \tau(t) \rangle \approx \frac{1}{s(q-1)} \ln \left[\frac{e^{((q-1)/q)\gamma} s/U_b}{1 - e^{-st + s\langle \tau(t) \rangle - \gamma/q}} \right]. \qquad (36)$$

Equation 36 is an equation for $\langle \tau(t) \rangle$; iterating it once and using q large, we recover Equation 32, up to some negligible terms. Note that the validity condition $R(t) \gg 1$ is approximately the same as in the first method, as either one can be rewritten as $t - 1/(sq)\ln(sq/U_b) \gg 1/(sq)$. Numerical simulations (see Figure 1 confirm that our analytical argument is sound and that Equation 32 gives indeed a good numerical approximation of the measured $\langle \tau(t) \rangle$ in stochastic edge simulations for values of t larger (but not very much larger) than $\langle \tau(t) \rangle$.

We now exploit Equation 32 to test the validity of DESAI and FISHER's (2007) result. The purpose of computing $\langle \tau(t) \rangle$ is to compute the mean establishment time of a new class, which we call T in the remainder of this section. DESAI and FISHER (2007) take $T = \langle \tau(\infty) \rangle$. But, as we argue now, it makes more sense to take $T \approx \langle \tau(T) \rangle$. Indeed, the reason why $\tau(t)$ depends on t stems from the fact that fitting the growth of the new class by an exponential (see Equation 4) is not a perfect description of what is really happening, and the best value of the parameter τ in this fit depends on the range of values of t where we want this exponential fit to be the most precise. This range of values is precisely t of order T, because it is at this moment that the new class becomes the second-best-fit class, starts feeding an even newer class, and becomes approximated by a deterministic exponential growth (see Equation 3). The whole theory can be made self-consistent only if the value of the best-fit class just before it is established matches its value just after its establishment, which happens only if the size of the best-fit class is well described for t of the order of T. Consequently,

$$T \approx \langle \tau(T) \rangle. \qquad (37)$$

We can try to solve Equation 37 directly from Equation 32; as the argument of the exponential in Equation 32 is small for $t \approx T$, we may expand it and we obtain, ignoring subleading logarithmic terms inside the logarithm,

$$T \approx \frac{1}{sq} \ln \frac{qs}{U_b}, \qquad (38)$$

which is quite different from the result of Equation 22 of DESAI and FISHER (2007) and is rather closer to that of Equation 23. Note that in this procedure we are operating slightly outside the range of validity of Equation 32: we need to use this equation at $t = T$ with T given in Equation 38, but we have shown it is valid only for $t - T \gg 1/(sq)$, that is, for t slightly larger than T. As we use $t \approx T$ only in the argument of a large logarithm, we do not believe that this approximation should affect at all the final result of Equation 38 to the leading order. Our more precise method presented in the next section of this article confirms this claim.

DESAI and FISHER's (2007) result in Equation 22 is obtained by making the approximation $T = \langle \tau(\infty) \rangle$, which is equivalent to neglecting the exponential in Equation 32. Clearly, this procedure is justified only when it gives a result compatible to that of Equation 38. This is the case only if

$$\ln q \ll \ln(s/U_b). \qquad (39)$$

[Note that DESAI and FISHER 2007 stated the condition for the validity of their work incorrectly in their Appendix G (M. DESAI, personal communication).]

Note that in this whole section, the derivation begins by assuming that the time T_0 at which the second-best class starts producing mutants is $-\infty$. We now briefly check that this is a sound hypothesis by showing that we would have reached, to the leading order, the same final result in Equation 38 by taking $T_0 = 0$. As can be checked from Equations 34 and 35, the values of λ contributing most to the integral are $\sim \lambda_c = y_c e^{sqt} \approx e^{sq\langle \tau(t) \rangle - \gamma}$. To reach Equation 38, we are interested in the time $t \approx T \approx \langle \tau(T) \rangle$, for which we obtain $\lambda_c \approx qs/U_b$, which we assumed is large. Now, if $T_0 = 0$, the upper bound of the integral in Equation 24 should be λ (since we use $sq \ll 1$, see Equation 10), which is large for the relevant values of λ. As in Equation 21, this means we need to subtract $\ln(1 + 1/\lambda) \approx 1/\lambda$, which is small, from the evaluation of this integral, Equation 25. But, within our working hypothesis $q \gg 1$ and $\lambda e^{-sqt} \ll 1$, the value of that integral is large: it diverges logarithmically for large q and small $\varepsilon = \lambda e^{-sqt}$; for $q > 3$ and $\varepsilon < 0.1$, it is >1, which

is much larger than the small correction $1/\lambda$. Therefore, considering $T_0 = 0$ instead of $T_0 = -\infty$ does not change the final result in Equation 38 at the leading order.

When using a finite back-extrapolation time, we run into another difficulty that we have not mentioned yet. The mathematically exact value of $\langle \tau(t) \rangle$ for any finite t is $-\infty$, because there is a nonzero probability p_0 that the size of the best-fit class is 0. For relevant values of t, the value of p_0 is incredibly small [one can show that $p_0 < \exp[-R(t)\ln(1 + 1/(sq))]$, with $R(t)$ given in Equation 35]. Of course, this event never occurred during all our simulations, and the only biologically observable quantity that makes sense is the average of $\tau(t)$ given that the new best-fit class is not empty. This quantity can be calculated in a precise way by replacing $\langle e^{-\lambda x(t)} \rangle$ everywhere in the previous derivation with $((\langle e^{-\lambda x(t)} \rangle - p_0)/(1 - p_0)$. As a close inspection of our derivations would show, we use only the function $\langle e^{-\lambda x(t)} \rangle$ in regions where it is much larger than p_0, so that nothing in our final result in Equation 32 should be changed because of that p_0. We now, however, present a better, more satisfying approach where none of these issues occurs.

A better back extrapolation: In the previous subsection, we have seen that $\langle \tau(t) \rangle$ depends strongly on t. At first glance, this result is somewhat unexpected. We intended the quantity τ to be the time at which the best-fit class crosses the stochastic threshold (*i.e.*, the establishment time of a new fitness class), and this time should have a specific, well-defined value. Instead, we have found that the expected value $\langle \tau(t) \rangle$ decays as t increases; *i.e.*, the longer we wait before we evaluate the system, the smaller the mean establishment time appears to be. This result indicates that $\tau(t)$, as defined above, is a poor method for getting an approximation of the mean establishment time.

We can understand the origin of the strong time dependence of $\langle \tau(t) \rangle$ from Equation 26. We rewrite this equation as

$$\langle e^{-\lambda[x(t) + (U_b/s)e^{-st}]} \rangle \approx \exp(-b\lambda^{1-1/q}). \tag{40}$$

In this form, we see that the variable $x(t)$ (defined in Equation 9) has a deterministic part $-(U_b/s)e^{-st}$ and that the fluctuations around that deterministic part have a nearly time-independent distribution described by the generating function on the right-hand side. The deterministic part has its origin in beneficial mutations fed into the best-fit class from the second-best class and can easily be understood by considering the deterministic approximation for the size $n(t)$ of the best-fit class:

$$\frac{dn(t)}{dt} = sqn(t) + \frac{U_b}{sq}e^{s(q-1)t}. \tag{41}$$

[This equation follows from Equation 1 with $n_{k-1}(t)$ given by Equation 3 and outgoing mutations neglected.] The origin of time is such that $n_{k-1}(0) = 1/(sq)$, and we fix the integration constant by imposing that $n(\tau_c) = 1/$
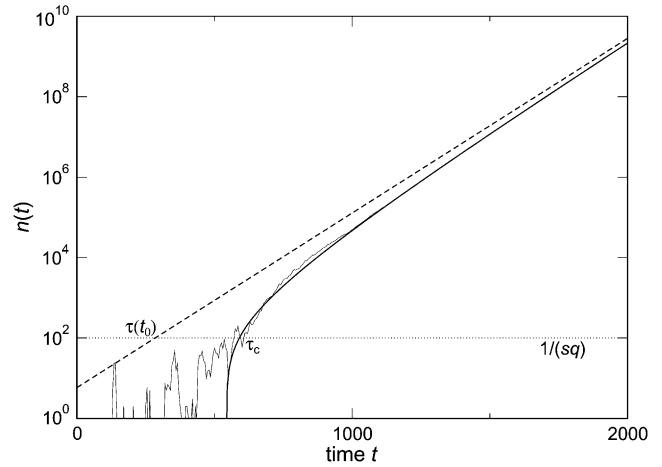

FIGURE 2.—Stochastic-edge simulation and back extrapolation to obtain $\tau(t_0)$ and τ_c. The thin solid line represents the size $n(t)$ of the best-fit class in a typical stochastic-edge simulation run for $s = 0.001$, $U_b = 0.0001$, and $q = 10$. The thick solid line is Equation 42 with $\tau_c = 590$ and the dashed line is Equation 4 with $\tau(t_0) = 284$. The values of τ_c and $\tau(t_0)$ have been determined at time $t_0 = 10{,}000$, which means that the stochastic value $n(t_0)$ is indeed given by, respectively, Equations 42 and 4. For large times ($t \gtrsim 2000$), both fits are good but for intermediate times, the stochastic $n(t)$ is best captured by the thick solid line. The time at which $n(t)$ reaches the stochastic threshold $1/(sq)$ (represented as a horizontal dotted line) is much closer to τ_c than to $\tau(t_0)$. Moreover, the value of $\tau(t_0)$ would have depended much more on the choice of t_0. For instance, taking $t_0 = 1000$ would have given $\tau(t_0) = 380$ and $\tau_c = 578$.

(sq). This choice will allow us to interpret τ_c later on as the establishment time, that is the time it takes to move one notch in the periodic motion of the wave. We find

$$n(t) = \frac{1}{sq}e^{sq(t-\tau_c)} + \frac{U_b}{s^2 q}e^{sqt}(e^{-s\tau_c} - e^{-st}). \tag{42}$$

Thus, because of incoming mutations, $n(t)$ does not grow purely exponentially, even in the deterministic limit. If we try to approximate this deterministic $n(t)$ or the stochastic $n(t)$ by a pure exponential as in Equation 4, the optimal fit of the parameter (τ in the case of Equation 4) depends on the time at which we want a good fit. This deviation from pure exponential growth is the source of the strong time dependence in $\langle \tau(t) \rangle$. It makes more sense to fit the stochastic $n(t)$ by Equation 42 but with τ_c now a random variable (see Figure 2. We may expect that in this way the distribution of τ_c will be largely independent of time. This is indeed the case. If we define $x(t) = sqe^{-sqt}n(t)$ as before (see Equation 9), we obtain from Equation 42 the deterministic evolution of $x(t)$:

$$x(t) + \frac{U_b}{s}e^{-st} = \frac{U_b}{s}e^{-s\tau_c} + e^{-sq\tau_c}. \tag{43}$$

Then, comparing this equation with Equation 40, we find that the deterministic component of $x(t)$ in the

probabilistic calculation corresponds exactly to the time-dependent part of $x(t)$ in a fully deterministic model of the stochastic edge. Interpreting τ_c as a random variable, we see that the generating function of the right-hand side of Equation 43 is given by Equation 40 and is nearly time independent. (Only "nearly" because we neglected terms of order e^{-sqt} in the right-hand side of Equation 24 to reach Equations 26 and 40.)

To sum up, we write the stochastic size $n(t)$ of the best-fit class as in Equation 42, where τ_c is a random variable. We equate the mean establishment time in the full population model with $\langle \tau_c \rangle$. In our new approach, $\langle \tau_c \rangle$ does not depend much on time (the subscript "c" stands for constant) and we avoid the difficulty of Desai and Fisher's approach. From Equations 40 and 43 the distribution of τ_c is determined by

$$\langle e^{-\lambda K} \rangle \approx \exp(-b\lambda^{1-1/q}) \tag{44}$$

with

$$K = \frac{U_b}{s} e^{-s\tau_c} + e^{-sq\tau_c}. \tag{45}$$

The new difficulty, of course, is to obtain $\langle \tau_c \rangle$ from these two equations.

Scaling function for $\langle \tau_c \rangle$: The equations determining $\langle \tau_c \rangle$ are transcendental, and we have not been able to obtain a simple, closed-form expression for $\langle \tau_c \rangle$. Nevertheless, we can gain substantial insight into how $\langle \tau_c \rangle$ depends on the parameter values s, U_b, and q. We make the change of variables

$$K' = K \left(\frac{U_b}{s} \right)^{-(q/(q-1))} \quad \text{and} \quad X = \left(\frac{U_b}{s} \right)^{-(q/(q-1))} e^{-sq\tau_c} \tag{46}$$

and obtain, from Equations 45 and 46,

$$\langle \tau_c \rangle = \frac{1}{s(q-1)} \ln \frac{s}{U_b} - \frac{1}{sq} \langle \ln X \rangle, \tag{47}$$

$$K' = X + X^{1/q}, \tag{48}$$

and, from Equation 44,

$$\langle e^{-\lambda K'} \rangle \approx \exp(-c\lambda^{1-1/q}) \quad \text{with} \quad c = \frac{bs}{U_b}. \tag{49}$$

The constant b is defined in Equation 12. Inserting this definition into $c = bs/U_b$, we obtain

$$c \approx \frac{\pi}{q \sin(\pi/q)}. \tag{50}$$

We note that the constant c does not depend on U_b or on s. Therefore, $\langle \ln X \rangle$ does not depend on U_b or s, and Equation 47 fully captures the dependency of $\langle \tau_c \rangle$ both on U_b and on s. (Actually, using the most precise version
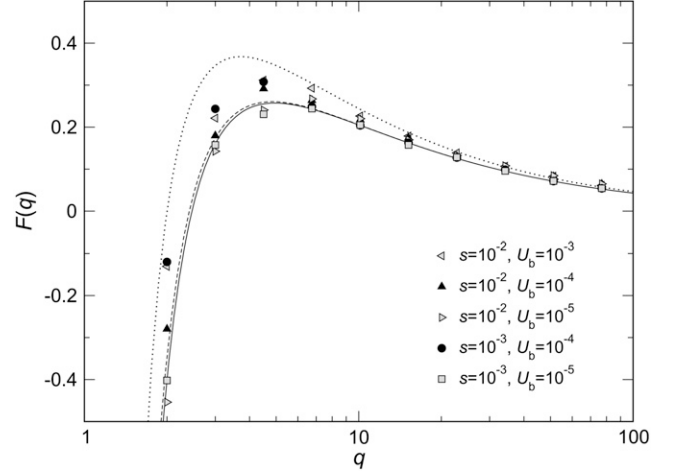

FIGURE 3.—Measured values $\langle \tau_c \rangle$ collapse onto a single scaling function $F(q)$. Data points are simulation results obtained from stochastic-edge simulations. For each parameter setting, we measured $\langle \tau_c \rangle$ and then plotted $s\langle \tau_c \rangle + (1/(q-1))\ln(U_b/s)$ as a function of q. The solid line represents a numerical evaluation of Equation 52 and the dashed line represents the approximate analytic expression Equation 53. The dotted line is the scaling function $F(q) = (1/(q-1))\ln(q-1)$ derived from Equation 57.

of Equation 12, there is a very weak dependency on s in c, but for any biologically relevant case, s is small and this dependency can be neglected.) We can then write

$$\langle \tau_c \rangle = \frac{1}{s} \left[F(q) + \frac{1}{(q-1)} \ln \frac{s}{U_b} \right], \tag{51}$$

where $F(q)$ is a function depending only on q and given by

$$F(q) = -\frac{1}{q} \langle \ln X \rangle. \tag{52}$$

In APPENDIX B, we show that we can write $\langle \ln X \rangle$ as a single integral, see Equation B20, which can be easily numerically evaluated for any value of q. We obtain

$$F(q) \approx \frac{1}{q-1} [\ln(q-1) - 0.345]. \tag{53}$$

The leading term comes from an analytical argument and the corrective term -0.345 is numerical. Figure 3 shows that the measured values of $\langle \tau_c \rangle$ in stochastic edge simulations can be reasonably well collapsed on the scaling function in Equation 53 for small values of s and U_b in a broad interval of q.

Inserting Equation 53 into Equation 51, we obtain

$$\langle \tau_c \rangle \approx \frac{1}{s(q-1)} \left[\ln \frac{s(q-1)}{U_b} - 0.345 \right]. \tag{54}$$

Here again, the establishment time given by Equation 54 is very similar for large q to the result in Equation 23 obtained by ROUZINE et al. (2003, 2008).

A simple approximation formula: With Equation 54, we have a good approximation for $\langle\tau_c\rangle$, but the derivation of this approximation was quite tedious. We can alternatively derive a simple approximation formula for $\langle\tau_c\rangle$ on the basis of biological considerations. A similar derivation was first presented by DESAI *et al.* (2007; DESAI and FISHER 2007), and was also used by ROUZINE *et al.* (2008) in the context of traveling-wave theory.

The average total number of mutations $m(t)$ produced by the second-best class up to time t is

$$m(t) = U_b \int_{-\infty}^{t} \frac{1}{sq} e^{s(q-1)t'} dt'$$
$$= \frac{U_b}{s^2 q(q-1)} e^{s(q-1)t}. \tag{55}$$

Each of these mutations has a probability of going to fixation of $sq/(1+sq) \approx sq$ (LENSKI and LEVIN 1985). Since a single mutation that fixes is sufficient to establish a new fitness class, we have

$$sqm(\langle\tau_c\rangle) \approx 1. \tag{56}$$

We rearrange this equation and find

$$\langle\tau_c\rangle \approx \frac{1}{s(q-1)} \ln\left(\frac{s(q-1)}{U_b}\right). \tag{57}$$

Despite the simplicity of this argument, we find that this expression has good accuracy, in particular for large q. Equation 57 differs from Equation 54 only in the constant 0.345 subtracted from the logarithm.

In the remainder of this article, we do not use Equation 57. We included its derivation primarily to show that the edge treatment of ROUZINE *et al.* (2008) is consistent with our derivation of $\langle\tau_c\rangle$.

Predicting the speed of adaptation: The goal of calculating $\langle\tau\rangle$ in the previous subsections was to obtain the speed of adaptation V, which is approximately given by $1/\langle\tau\rangle$. [Throughout this subsection, we mean $\langle\tau\rangle$ to stand for either $\langle\tau(t)\rangle$ or $\langle\tau_c\rangle$.] Since $\langle\tau\rangle$ depends on q, which is a derived property of the adapting population and not known in advance, we need a second, independent expression linking $\langle\tau\rangle$ and q. DESAI and FISHER (2007) obtained this second expression from the normalization condition that the sum over all fitness classes has to yield the population size N. They argued that at the time of establishment of the best class, the size n_{k_0-r} of a fitness class r mutations away from the best class is given approximately by

$$n_{k_0-r} \approx \frac{1}{sq} \exp([rq - r(r+1)/2]s\langle\tau\rangle), \tag{58}$$

because the second-best class has on average been growing exponentially at rate $s(q-1)$ for a time interval $\langle\tau\rangle$, the third-best class has in addition been growing at rate $s(q-2)$ for an additional time interval $\langle\tau\rangle$, and so on. As the largest term in the sum arises for $r \approx q$, DESAI

and FISHER (2007) simplified the normalization condition $N = \sum_k n_k$ to $N \approx n_{k_0-q}$, which yields

$$s\langle\tau\rangle q(q-1) \approx 2\ln(sqN). \tag{59}$$

Inserting the expression for $\langle\tau\rangle$ from DESAI and FISHER (2007) (their Equation 36) into this expression recovers their Equation 39, an expression that implicitly determines q as a function of s, U_b, and N. Note, however, that Equation 59 works only if $s\langle\tau\rangle$ is large; *i.e.*, $s \gg V$. When $s\langle\tau\rangle$ is small, *i.e.*, $s \ll V$, a better approximation to $N = \sum_k n_k$ is to replace the summation with an integral, $N \approx \int dk\, n_k$ which gives

$$s\langle\tau\rangle\left(q - \frac{1}{2}\right)^2 \approx 2\ln(sqN) + \ln[s\langle\tau\rangle/(2\pi)]. \tag{60}$$

Equations 59 and 60 correspond, respectively, to the two limits of a narrow wave and a broad wave discussed in ROUZINE *et al.* (2008). Here, to better compare our results to Desai and Fisher's approach, we use only Equation 59 even for $s < V$ as in Figure 4B. Using the more correct Equation 60 would have resulted in only a small correction of $\sim$5% at $N = 10^9$ to $<$14% at $N = 10^4$ for the parameter settings of Figure 4B (data not shown).

To sum up, the final prediction in this model for the speed of adaptation V is

$$V = \frac{1}{\langle\tau_c\rangle}, \tag{61}$$

where $\langle\tau_c\rangle$ as a function of N, s, and U_b is obtained by eliminating q in Equations 54 and 59. As we cannot analytically eliminate q, we have only two options: either to derive an approximate expression for q from these equations or to solve them numerically.

DESAI and FISHER (2007) derived an approximate expression for q, neglecting some large logarithm inside of another logarithm (see their Equations 39 and 40). The final result shown in their Figure 5 agrees well with simulation results. However, when we compared this approximate expression to the corresponding exact numerical solution of their Equation 39, we found that the term DESAI and FISHER (2007) neglected is not small in their parameter range and that, compared to the results of numerical simulations, the solution obtained by eliminating q numerically from Equations 54 and 59 performed worse than their approximate expression. Thus, the performance of the approximate expression is partly due to cancellation of errors, and we do not further consider this approximate expression here.

Figure 4 compares how the work of DESAI and FISHER (2007) and the present work perform in predicting the speed of adaptation V. The dashed lines represent the exact numerical solution to the expression derived by DESAI and FISHER (2007). This expression works reasonably well for low wave speeds [$\ln q < \ln(s/U_b)$, Figure 4A], but performs poorly at high wave speeds [$\ln q > \ln(s/U_b)$, Figure 4B], as expected. The poor perfor-
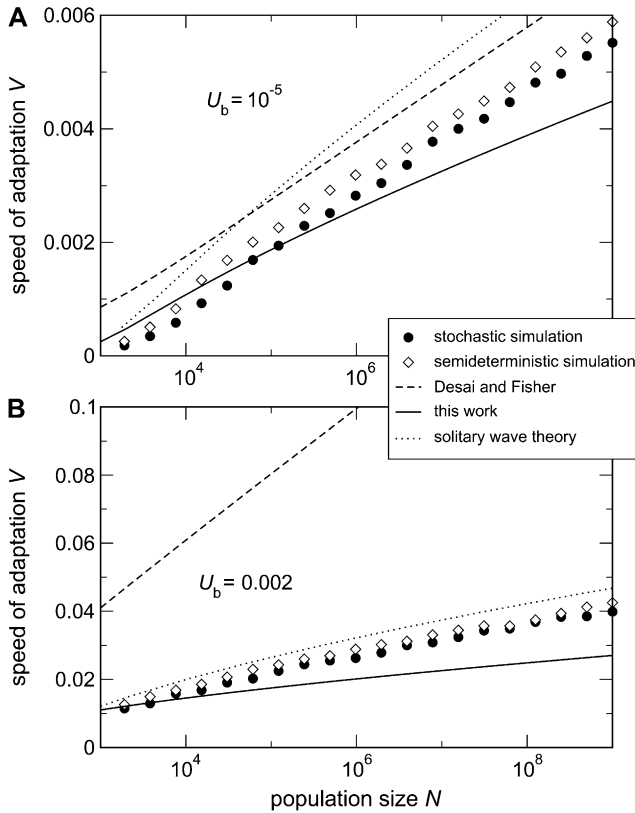
FIGURE 4.—Speed of adaptation as a function of population size N. Points are simulation results: the solid circles come from stochastic simulations of the full model, while the open diamonds come from semideterministic simulations where only the best-fit class is stochastic. Dashed lines were obtained by numerically solving Equations 36 and 39 of DESAI and FISHER (2007). Solid lines were obtained by numerically solving Equations 54 and 60 in the present work. Dotted lines are Equation 52 (for A) and Equation 51 (for B) from ROUZINE *et al.* (2008). Parameters are $s = 0.01$ and $U_b = 10^{-5}$ (A) and $s = 0.01$ and $U_b = 0.002$ (B). Note that our simulation results are in excellent agreement with simulation results reported by DESAI and FISHER (2007).

mance at high wave speeds is caused by the breakdown of the $\langle \tau(\infty) \rangle$ approximation. If we instead use $\langle \tau_c \rangle$, we get a significant improvement in the prediction accuracy at high wave speeds (solid lines in Figure 4). At low wave speeds, the two methods have comparable accuracies.

For comparison, we also plotted the predictions from traveling-wave theory (Figure 4, dotted lines), as derived by ROUZINE *et al.* (2003, 2008). At low wave speeds (Figure 4A), traveling-wave theory performs approximately as well as both the original approach by DESAI and FISHER (2007) and our revision of it. While all three methods show reasonable performance in this parameter region, none has excellent accuracy. At high wave speeds (Figure 4B), traveling-wave theory performs better than our revised version of the Desai-and-Fisher approach and comes close to the speed found in semi-

deterministic simulations (see also next paragraph). Traveling-wave theory takes into account the effect of mutation pressure on intermediate fitness classes and thus incorporates their nonexponential growth. By contrast, we have neglected this effect in the present work, and have assumed that the second-best class grows purely exponentially (approximation 4). Certainly, the present work tends to underestimate the speed of adaptation because of approximation 4. It is less clear why traveling-wave theory always overestimates the wave speed. Possibly, the assumption made in traveling-wave theory that the wave speed is determined by the mean size of the stochastic edge might underestimate the drag exerted by the stochastic edge when it is very small.

We also carried out semideterministic simulations in which the best-fit class was treated stochastically and all other classes were treated deterministically. The semideterministic simulation tests the fundamental assumption, made both by DESAI and FISHER (2007) and in traveling-wave theory (ROUZINE *et al.* 2003, 2008), that only a single stochastic fitness class is necessary to describe an adapting population. Any analytical treatment of adaptive evolution based on this assumption can ever perform only as well as the semideterministic simulations. We found that the wave speed in the semideterministic simulations was close, but not exactly the same, as the true wave speed (Figure 4). In general, the semideterministic simulations tended to overestimate the wave speed, in particular for small wave speeds.

Conclusions: The work by DESAI and FISHER (2007) constitutes an interesting new approach to calculating the speed of adaptation. However, their work does not apply to high adaptation speeds, *i.e.*, populations with large q. This limitation arises because the growth of the best-fit class cannot be described as a purely exponential growth times a random constant when the population evolves rapidly. Because the best-fit class is continuously being fed beneficial mutations from the second-best class, the random variable that modifies the exponential growth of the best-fit class is actually time dependent, and its mean changes with time.

Here, we have modified Desai and Fisher's method to handle correctly the nonexponential growth of the best-fit class. Our modification leads to a substantial improvement in the prediction of the speed of adaptation for rapidly adapting populations and agrees with predictions from traveling-wave theory. However, we have relied on an exponentially growing second-best class throughout this work, even though beneficial mutations from classes with lower fitness contribute significantly to the growth of the second-best class. A more accurate treatment of adaptive evolution than we have presented here will have to take this fact into account.

LITERATURE CITED

ABATE, J., and P. P. VALKÓ, 2004 Multi-precision laplace transform inversion. Int. J. Numer. Methods Eng. **60:** 979–993.

ATHREYA, K. B., and P. E. NEY, 1972 *Branching Processes.* Springer, New York.

BAAKE, E., M. BAAKE and H. WAGNER, 1997 Ising quantum chain is equivalent to a model of biological evolution. Phys. Rev. Lett. **78:** 559–562.

BARTON, N. H., 1995 Linkage and the limits to natural selection. Genetics **140:** 821–841.

BEERENWINKEL, N., T. ANTAL, D. DINGLI, A. TRAULSEN, K. W. KINZLER *et al.*, 2007 Genetic progression and the waiting time to cancer. PLoS Comput. Biol. **3:** e225.

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

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

DESAI, M. M., and D. S. FISHER, 2007 Beneficial mutation-selection balance and the effect of linkage on positive selection. Genetics **176:** 1759–1798.

DESAI, M. M., D. S. FISHER and A. W. MURRAY, 2007 The speed of evolution and maintenance of variation in asexual populations. Curr. Biol. **17:** 385–394.

GALASSI, M., J. DAVIES, J. THEILER, B. GOUGH, G. JUNGMAN *et al.*, 2006 *GNU Scientific Library Reference Manual*, Ed. 2. Network Theory Limited, Bristol, UK.

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

KACHITVICHYANUKUL, V., and B. SCHMEISER, 1985 Computer generation of hypergeometric random variates. J. Stat. Comput. Simul. **22:** 127–145.

KESSLER, D. A., H. LEVINE, D. RIDGWAY and L. TSIMRING, 1997 Evolution on a smooth landscape. J. Stat. Phys. **87:** 519–544.

LENSKI, R. E., and B. R. LEVIN, 1985 Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. Am. Nat. **125:** 585–602.

MAYNARD SMITH, J., 1971 What use is sex? J. Theor. Biol. **30:** 319–335.

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

PARK, S.-C., and J. KRUG, 2007 Clonal interference in large populations. Proc. Natl. Acad. Sci. USA **104:** 18135–18140.

PRÜGEL-BENNETT, A., 1997 Modelling evolving populations. J. Theor. Biol. **185:** 81–95.

R DEVELOPMENT CORE TEAM, 2007 *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna.

ROUZINE, I. M., A. RODRIGO and J. M. COFFIN, 2001 Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. Microbiol. Mol. Biol. Rev. **65:** 151–181.

ROUZINE, I. M., J. WAKELEY and J. M. COFFIN, 2003 The solitary wave of asexual evolution. Proc. Natl. Acad. Sci. USA **100:** 587–592.

ROUZINE, I. M., É. BRUNET and C. O. WILKE, 2008 The traveling wave approach to asexual evolution: Muller's ratchet and speed of adaptation. Theor. Popul. Biol. **73:** 24–46.

TSIMRING, L. S., H. LEVINE and D. A. KESSLER, 1996 RNA virus evolution via a fitness-space model. Phys. Rev. Lett. **76:** 4440–4443.

VALKÓ, P. P., and J. ABATE, 2004 Comparison of sequence accelerators for the Gaver method of numerical Laplace transform inversion. Comput. Math. Appl. **48:** 629–636.

WILKE, C. O., 2004 The speed of adaptation in large asexual populations. Genetics **167:** 2045–2053.

APPENDIX A: VALIDITY OF APPROXIMATION 4

The goal of this APPENDIX is to test approximation 4, namely that the second-best-fit class grows exponentially with a rate $s(q-1)$:

$$n_{k_0-1}(t) \approx \frac{1}{sq} e^{s(q-1)t}. \tag{A1}$$

[Remember that k_0 was defined as the location of the stochastic edge, so that the second-best-fit class is at position $k_0 - 1$. The origin of time is when that class just got established: $n_{k_0-1}(0) = 1/(sq)$.]

The size of any established class can be obtained from Equation 1, which reads for the second-best-fit class:

$$\frac{dn_{k_0-1}(t)}{dt} = s(q-1)n_{k_0-1}(t) + U_b n_{k_0-2}(t) - U_b n_{k_0-1}(t). \tag{A2}$$

Equation A1 is the solution of Equation A2 only if the second and third terms on the right-hand side of Equation A2 are negligible. The third term is easily dealt with as we assumed throughout this work that $s(q-1) \approx sq \gg U_b$. For the second term, we need to evaluate the size $n_{k_0-2}(t)$ of the third-best class.

We proceed by assuming that the third-best class is described by a deterministic exponential formula, analogous to the equation used in the two-class model for the second-best class. Then, from Equation 58, we obtain for $0 \le t \le \tau$ the expression

$$n_{k_0-2}(t) = \frac{1}{sq} e^{s(q-1)\tau + s(q-2)t}. \tag{A3}$$

This expression is based on the assumptions that the class $k_0 - 2$ got established at time $-\tau$ when it reached the size $1/(sq)$, grew from time $-\tau$ to time 0 with rate $s(q-1)$, and grows from time 0 to time τ with rate $s(q-2)$.

Using the value of τ given in Equation 54, we find

$$n_{k_0-2}(t) = \frac{\alpha}{U_b} e^{s(q-2)t}, \tag{A4}$$

where α is of order 1. Using $n_{k_0-1}(0) = 1/(sq)$ and neglecting the third term on the right-hand side of Equation A2, we find as the solution to Equation A2

$$n_{k_0-1}(t) = \frac{1}{sq}e^{s(q-1)t} + \frac{\alpha}{s}\left(e^{s(q-1)t} - e^{s(q-2)t}\right). \tag{A5}$$

For moderate times such that $t \ll 1/(sq)$, the second term in Equation A5 is negligible and we recover Equation A1. However, the most relevant time interval is when t is very close to τ, when most of the mutations from the second-best class to the not-yet-established best class occur (ROUZINE *et al.* 2008). For $t \sim \tau$, further estimates depend on whether product $s\tau$ is small or large (*i.e.*, whether $V \gg s$ or $V \ll s$). For $s\tau \ll 1$, using $\exp[s(q-2)t] \approx (1-st)\exp[s(q-1)t]$, we obtain

$$n_{k_0-1}(t) \approx \left(\frac{1}{sq} + \alpha t\right)e^{s(q-1)t}. \tag{A6}$$

This expression deviates from Equation A1 due to the second term in parentheses. The deviation is by a factor of order 2 when t becomes of the order of $1/(sq)$, which happens early in a cycle, as $\tau \gg 1/(sq)$ from Equation 54. At the end of a cycle, $t \sim \tau$, the second term is larger than the first term by a factor of $\ln(sq/U_b) \gg 1$. Therefore, approximation 4 is not valid, and the second-best-fit class cannot be described by Equation A1.

At $s\tau \gg 1$ and $t \sim \tau$, we can neglect the third exponential in Equation A5. Then, instead of Equation A6, we obtain

$$n_{k_0-1}(t) \approx \left(\frac{1}{sq} + \frac{\alpha}{s}\right)e^{s(q-1)t}. \tag{A7}$$

The first term in parentheses is negligible and the result differs from Equation A1 by the large factor $\alpha q \gg 1$. Therefore, approximation 4 is not valid in this case either.

Thus, taking into account the third-best class creates an additional large factor in the size of the second-best class at the most relevant times $t \sim \tau$. This factor is on the order of either q or $\ln(sq/U_b)$, whichever is smaller, and approximation 4 is not valid by itself. One could try to fix this issue by using Equation A5 instead of Equation A1 for the size of the second-best class, but it would make the derivation much more complicated. Note, however, that, as the effects of mutations enter the final result only through the logarithm of the mutation rate, it is plausible (but remains to be checked) that the large corrective factors of Equation A6 or Equation A7 will enter the final result as a logarithmic correction. On the other hand, there is no guarantee that taking into account the third-best class is sufficient, and it might be that one needs also to consider the effects of the fourth- or fifth-best class. In all cases, the replacement of the full-population model by a two-class model with an exponentially growing second-best class is problematic and deserves a more careful investigation.

APPENDIX B: CALCULATING $\langle \ln X \rangle$

To calculate $F(q)$, we have to calculate $\langle \ln X \rangle$, where X is the only positive root of

$$X + X^{1/q} = K \tag{B1}$$

and we have written K instead of K' for simplicity. The moment-generating function for K is (Equation 49)

$$\langle e^{-\lambda K} \rangle = \exp[-c\lambda^{1-1/q}]. \tag{B2}$$

A first approach is to evaluate $\langle \ln X \rangle$ numerically using an inverse Laplace transform. First, we calculate the density function $p_K(y)$ of the probability distribution of K from the inverse Laplace transform $\mathcal{L}^{-1}$ of the moment-generating function of K,

$$p_K(y) = \mathcal{L}^{-1}\{\exp(-c\lambda^{1-1/q})\}, \tag{B3}$$

where $\mathcal{L}^{-1}$ can be written as an integral. In practice, this integral can be evaluated with efficient numerical algorithms (VALKÓ and ABATE 2004; ABATE and VALKÓ 2004). A transformation of variables gives us the density function $p_X(y)$ of the probability distribution of X:

$$p_X(y) = \left(1 + \frac{1}{q}y^{(1/q)-1}\right)p_K(y + y^{1/q}). \tag{B4}$$

Finally, we integrate to obtain $\langle \ln X \rangle$:

$$\langle \ln X \rangle = \int_0^\infty p_X(y) \ln y \, dy. \tag{B5}$$

This method can be worked out, but it is delicate and computationally expensive to evaluate numerically with a good accuracy these not so well-behaved double integrals, especially for large values of q. We now present an alternative method that allows us to write $\langle \ln X \rangle$ as a simple integral, which is much easier to evaluate.

Writing ln X as a series: Our first step is to invert Equation B1. By using Cauchy's integral formula from complex analysis, we can write for any analytical function f the quantity $f(X)$ as

$$f(X) = \frac{1}{2\pi i} \oint f(z) \frac{1 + (1/q)z^{(1/q)-1}}{z + z^{1/q} - K} dz, \tag{B6}$$

where the integration is on a contour surrounding the only positive root of Equation B1. We set $f(z) = z^\mu$ and make use of the Taylor series

$$\frac{1}{a+b-K} = \sum_{n\geq 0}(-1)^n \frac{b^n}{(a-K)^{n+1}}. \tag{B7}$$

Setting $a = z$, $b = z^{1/q}$ in the above expansion, we obtain

$$X^\mu = \frac{1}{2\pi i} \oint z^\mu \sum_{n\geq 0}(-1)^n \frac{(1+(1/q)z^{(1/q)-1})z^{n/q}}{(z-K)^{n+1}} dz \tag{B8}$$

$$= \sum_{n\geq 0}(-1)^n \frac{1}{2\pi i} \oint \frac{(z+K)^{\mu+n/q} + (1/q)(z+K)^{\mu+n/q+(1/q)-1}}{z^{n+1}} dz \tag{B9}$$

$$= \sum_{n\geq 0}(-1)^n \left[\binom{\mu+n/q}{n} K^{\mu+n/q-n} + \frac{1}{q} \binom{\mu+n/q+(1/q)-1}{n} K^{\mu+n/q+(1/q)-1-n} \right] \tag{B10}$$

$$= K^\mu + \sum_{n\geq 1}(-1)^n K^{\mu+n/q-n} \frac{q\mu}{q\mu+n} \binom{\mu+n/q}{n}. \tag{B11}$$

[We use the Binomial symbol $\binom{x}{n}$ for the x noninteger, with the convention that $\binom{x}{n} = x(x-1)\cdots(x-n+1)/n!$.]
Expanding both sides of the equation to first order in μ and comparing the coefficients of the linear term, we find

$$\ln X = \ln K + \sum_{n\geq 1}(-1)^n \frac{q}{n} \binom{n/q}{n} K^{n/q-n}. \tag{B12}$$

Taking the average: We can now calculate $\langle \ln X \rangle$ by averaging Equation B12 term by term. Following the same steps as in the derivation of Equations 16 and 17 in the main text, we obtain from Equation B2

$$\langle K^\mu \rangle = \frac{\Gamma(1 - \mu q/(q-1))}{\Gamma(1-\mu)} c^{\mu q/(q-1)} \quad \text{for } \mu < 0, \tag{B13}$$

$$\langle \ln K \rangle = \frac{q}{q-1} \ln(ce^{\gamma/q}). \tag{B14}$$

Using these two equations, we find

$$\langle \ln X \rangle = \frac{q}{q-1} \ln(ce^{\gamma/q}) + \sum_{n \geq 1} (-1)^n \frac{q}{n} \binom{n/q}{n} \frac{\Gamma(1+n)}{\Gamma(1+n-n/q)} c^{-n} \tag{B15}$$

$$= \frac{q}{q-1} \ln(ce^{\gamma/q}) - \sum_{n \geq 1} \frac{q^2}{\pi n^2 (q-1)} \Gamma(1+n/q) \sin(\pi n/q) c^{-n}, \tag{B16}$$

where we have made use of Euler's reflection formula $\Gamma(x)\Gamma(1-x) = \pi/\sin(\pi x)$. The resulting series diverges. However, we treat it as a formal expansion of $\langle \ln X \rangle$ and continue. We replace $\Gamma(1+n/q)$ by its integral representation, integrate by parts once, and obtain

$$\langle \ln X \rangle = \frac{q}{q-1} \left[\ln(ce^{\gamma/q}) - \int_0^\infty d\lambda \frac{e^{-\lambda}}{\pi\lambda} \sum_{n \geq 1} \frac{1}{n} \lambda^{n/q} \sin(\pi n/q) c^{-n} \right]. \tag{B17}$$

We now write $\sin(\pi n/q)$ as the imaginary part of $e^{i\pi n/q}$ and note that the remaining sum is the Taylor expansion of the complex logarithm. Thus, we arrive at

$$\langle \ln X \rangle = \frac{q}{q-1} \left[\ln(ce^{\gamma/q}) - \int_0^\infty d\lambda \frac{e^{-\lambda}}{\lambda\pi} \mathfrak{J}\left(-\ln\left[1 - \frac{\lambda^{1/q} e^{i\pi/q}}{c} \right] \right) \right], \tag{B18}$$

where $\mathfrak{J}(z)$ indicates the imaginary part of z. Let $\rho > 0$ and ϕ be such that $\rho e^{-i\phi} = 1 - \lambda^{1/q} e^{i\pi/q}/c$. Then, we have

$$\tan\phi = \frac{\sin(\pi/q)\lambda^{1/q}/c}{1 - \cos(\pi/q)\lambda^{1/q}/c} \quad \text{with } \sin\phi \geq 0 \tag{B19}$$

and

$$\mathfrak{J}\left(-\ln\left[1 - \frac{\lambda^{1/q} e^{i\pi/q}}{c} \right] \right) = \phi. \tag{B20}$$

The angle ϕ is defined only up to a multiple of 2π, but the result must be $\phi = 0$ when $\lambda = 0$ and the result must be a continuous function of λ. (While the sum converges only when $\lambda^{1/q}/c < 1$, we are considering here the analytical continuation of this function.) This reasoning implies that $0 \leq \phi \leq \pi$.

The final result is then

$$\langle \ln X \rangle = \frac{q}{q-1} \left[\ln(ce^{\gamma/q}) - \int_0^\infty d\lambda \frac{e^{-\lambda}}{\lambda\pi} \phi(\lambda) \right], \tag{B20}$$

where

$$\tan\phi(\lambda) = \frac{\sin(\pi/q)}{c\lambda^{-1/q} - \cos(\pi/q)} \quad \text{and} \quad 0 \leq \phi(\lambda) \leq \pi. \tag{B21}$$

Note that

$$\phi(\lambda) = \begin{cases} \arctan\left(\frac{\sin(\pi/q)}{c\lambda^{-1/q} - \cos(\pi/q)} \right) & \text{when } c\lambda^{-1/q} > \cos(\pi/q), \\ \pi + \arctan\left(\frac{\sin(\pi/q)}{c\lambda^{-1/q} - \cos(\pi/q)} \right) & \text{when } c\lambda^{-1/q} < \cos(\pi/q). \end{cases} \tag{B22}$$

We evaluated both Equations B20 and B5 numerically and found excellent agreement between the two formulas.

Leading asymptotic of $\langle \ln X \rangle$: We now evaluate the integral in Equation B20 in the large q limit. For a fixed small λ and $q \to \infty$, it is easy to see that

$$\phi(\lambda) \approx -\frac{\pi}{\ln \lambda} \quad \text{for fixed (small) } \lambda \text{ and } q \to \infty. \tag{B23}$$

(Remember that $c \approx 1$ for large q.) However, replacing $\phi(\lambda)$ by that expression leads to a diverging integral. What happens is that for a given large q, the approximation Equation B23 breaks for extremely small values of λ, and we obtain

$$\phi(\lambda) \approx \frac{\pi\lambda^{1/q}}{q} \quad \text{for fixed (large) } q \text{ and } \lambda \to 0. \tag{B24}$$

With the latter approximation, the integral converges. Looking more closely at the approximations made, we can check that Equation B23 is valid for $e^{-q} \ll \lambda \ll 1$ and that Equation B24 is valid for $\lambda \ll e^{-q}$.

Therefore, it makes sense to cut the integral into three parts: one for $0 < \lambda < e^{-q}$, where we use Equation B24, one for $e^{-q} < \lambda < \varepsilon$, where we use Equation B23 and where ε is some fixed small number, and one for $\lambda > \varepsilon$. It is easy to check that the first and third parts give a number of order 1 (in other words, they do not diverge when $q \to \infty$) and that the second part dominates the integral:

$$\int_0^\infty d\lambda \frac{e^{-\lambda}}{\lambda \pi} \phi(\lambda) \approx \int_{e^{-q}}^\varepsilon d\lambda \frac{-1}{\lambda \ln \lambda} = \ln[-\ln(e^{-q})] - \ln[-\ln \varepsilon] \approx \ln q. \tag{B25}$$

Once this leading term has been identified, we can evaluate numerically the integral for many values of q and extract an asymptotic expansion of the correction to the leading term. We found

$$\int_0^\infty d\lambda \frac{e^{-\lambda}}{\lambda \pi} \phi(\lambda) \approx \ln q - 0.345 - \frac{0.45}{q} + \frac{1.0}{q^2} - \frac{0.3}{q^3} + \cdots. \tag{B26}$$

Another possibility is to write an expansion of the integral in powers of the variable $q - 1$:

$$\int_0^\infty d\lambda \frac{e^{-\lambda}}{\lambda \pi} \phi(\lambda) \approx \ln(q-1) - 0.345 + \frac{0.58}{q-1} + \frac{0.19}{(q-1)^2} + \cdots. \tag{B27}$$

Both asymptotic expansions are, of course, equally good for large q, but it happens that truncated to its first terms, the second expansion is better than the first at approximating the integral for smaller q. Inserting the latter expansion into Equation B20 and using Equation 52, we recover Equation 53.

We have not been able to find a theory for the numerical coefficients of this asymptotic expansion, and this remains an interesting challenge. The expansion of Equation B27 is a very good approximation of the integral in the range $q \in [2, \infty)$.