Title: Competition between continuously evolving lineages in asexual populations

Authors: Noah Ribeck^{1,2}, Joseph S. Mulka¹, Luis Zaman^{2,3}, Brian D. Connelly^{2,3}, Richard E. Lenski^{1,2}

¹Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan 48824

²BEACON Center for the Study of Evolution in Action, Michigan State University, East Lansing, Michigan 48824

³Department of Biology, University of Washington, Seattle, Washington 98195

Running title: Lineage-interference theory

Keywords: adaptation rate, asexual evolution, beneficial mutations, clonal interference, multiple-mutations regime, periodic selection

ABSTRACT

In an asexual population, the fate of a beneficial mutation depends on how its lineage competes against other mutant lineages in the population. With high beneficial mutation rates or large population sizes, competition between contending mutations is strong, and successful lineages can accumulate multiple mutations before any single one achieves fixation. Most current theory about asexual population dynamics either neglects this multiple-mutations regime or introduces simplifying assumptions that may not apply. Here, we develop a theoretical framework that describes the dynamics of adaptation and substitution over all mutation-rate regimes by conceptualizing the population as a collection of continuously adapting lineages. This model of “lineage interference” shows that each new mutant’s advantage over the rest of the population must be above a critical threshold in order to likely achieve fixation, and we derive a simple expression for that threshold. We apply this framework to examine the role of beneficial mutations with different effect sizes across the transition to the multiple-mutations regime.

INTRODUCTION

Mutation and natural selection are two core processes in evolutionary theory, yet we lack a comprehensive theoretical framework for describing the dynamics of adaptation by natural selection as beneficial mutations sweep through asexual populations. Much of the difficulty arises from the *interference* between lineages with alternative contending beneficial mutations that compete with one another as they move toward (but often fail to achieve) fixation. Gerrish and Lenski developed a model of *clonal interference* based on the probability that a particular beneficial mutation will become fixed before a superior mutation outcompetes it (GERRISH AND LENSKI 1998). Their work has since gained wide acceptance in the field as the standard quantitative description of adaptation. However, clonal-interference theory rests upon a key assumption that each contending mutation

arises on an identical genetic background. This assumption fails for populations that are sufficiently large or have high beneficial mutation rates, such that lineages accumulate multiple beneficial mutations before any single mutation achieves fixation (Fig. 1). Several papers have attempted to relax this assumption, but in so doing they introduced additional assumptions, such as constant fitness effects (ROUZINE *et al.* 2003; DESAI AND FISHER 2007; BRUNET *et al.* 2008; ROUZINE *et al.* 2008; HALLATSCHEK 2011). Thus, we still need an intuitive and quantitatively accurate framework that describes asexual adaptation in both the strong and weak mutation regimes.

Here, we develop a general population-genetic model for describing the dynamics of adaptation by natural selection in asexual populations, one that accounts for mutations arising on genetic backgrounds that already carry beneficial mutations that have not yet fixed. We will follow the mathematical logic of the clonal-interference model, but instead of describing the fates of *clones* in a population, we will describe the fates of *lineages*. In this model of *lineage interference*, we make the key observation that the expected rate of fitness increase for any lineage is proportional to its frequency in the population. Any segregating lineage must then have a large enough initial advantage over the rest of the population such that it reaches 50% frequency before its relative advantage is negated by adaptation of competing lineages. By comparison to simulations, we will show that this model provides a general description of asexual adaptation that accurately predicts the rates of adaptation and substitution across parameter regimes.

METHODS

Lineage-interference model

In clonal-interference theory, the likelihood of a particular beneficial mutation surviving interference is assessed by its probability of sweeping to fixation before any single superior mutation arises on the same background (GERRISH AND LENSKI 1998). In this formulation, each individual mutation competes for fixation against other individual mutations. This simplification breaks down, however, for populations with high $N\mu$, where N is the population size and μ is the beneficial mutation rate (per genome per generation).

To describe more accurately populations with a high mutation supply rate (high $N\mu$) — the *multiple-mutations regime* (DESAI AND FISHER 2007)—we introduce lineage-interference theory. This theory explicitly considers the genetic background in both the treatment of newly occurring mutations and the effect of interference on the fates of those mutations. First, the success of a new mutation depends on the sum of its beneficial fitness effect and its background fitness; for simplicity, we assume there is no epistasis. This “mutant fitness” determines an “effective” selective advantage that includes the effects of any pre-existing mutations carried by the genetic background. Second, the probability of fixation for a new mutation can be increased by its lineage’s acquisition of additional beneficial mutations, and decreased by beneficial mutations in competing lineages. In other words, the “mutant lineage” defined by the new mutation, having just

acquired a bump in fitness, has a head start in a race against the rest of the population, which we will call the “pack lineage.”

As this race progresses, both lineages can acquire additional mutations, and the mutant lineage eventually either wins (achieves fixation) or loses (goes extinct). The original mutant will need a head start in order to win, because the pack lineage initially has many more opportunities to acquire additional beneficial mutations owing to its larger population size. To quantify how much of a head start the mutant lineage needs, we must express the adaptation rates for both the mutant lineage v_{mut} and the pack lineage v_{pack} that result from the accumulation of additional beneficial mutations and their spread by selection. In the absence of epistasis, both lineages have access to the same pool of additional beneficial mutations, and any additional mutation in either lineage is subject to the same potential effects of interference from other members of the population. Early mutations in the mutant lineage have an advantage by virtue of occurring on a higher fitness background: if the effect size of the second mutation is comparable to the original mutation, the double-mutant now has roughly twice the probability of surviving drift as the original. However, because the mutant lineage is initially outnumbered by a factor of N , the pack lineage has an N -times greater supply rate of new mutations, so v_{mut} is still negligible compared to v_{pack} when the mutant lineage is relatively new. Therefore, each lineage’s adaptation rate is approximately proportional to its frequency:

$$\begin{aligned} v_{\text{mut}} &= Kp \\ v_{\text{pack}} &= K(1 - p) \end{aligned} \tag{1}$$

where p is the frequency of the mutant lineage and K is a proportionality constant. The adaptation rate v of the entire population, which is approximately constant in the absence of epistasis, is given by the weighted sum of the two groups:

$$v = pv_{\text{mut}} + (1 - p)v_{\text{pack}} \tag{2}$$

The mutant lineage has a relative fitness advantage s over the pack lineage—i.e., over the rest of the population. This advantage is measured relative to the mean fitness of the pack lineage, so it includes effects from the mutation itself and from its genetic background. The rate at which s changes is equal to the difference between the two groups’ respective adaptation rates:

$$\frac{ds}{dt} = v_{\text{mut}} - v_{\text{pack}} \tag{3}$$

Together with Eqs. (1) and (2), this rate gives an expression for the changing relative advantage of the mutant lineage, in terms of constant v and time-dependent p :

$$\frac{ds}{dt} = -v \frac{1 - 2p}{1 - 2p(1 - p)} \tag{4}$$

The changing relative frequency of the mutant lineage under continuous growth is given by (FISHER 1930) as:

$$\frac{dp}{dt} = sp(1 - p) \quad (5)$$

Eqs. (4) and (5) are the “lineage interference equations,” which provide a framework for determining the fate of the original mutant lineage. Because each lineage’s adaptation rate is proportional to its frequency, the mutant’s frequency must—if it is likely to survive—reach 50% before its mean fitness is surpassed by that of the pack lineage. This implies the existence of a critical threshold s_{crit} for the original mutant’s initial advantage s_0 , above which the original mutation is likely to fix, and below which it is likely to go extinct. Numerical solutions for these equations with different values of s_0 show this critical behavior (see Fig. 2). The following section provides a derivation for an expression of s_{crit} .

This framework is accurate when the mutant lineage is new and therefore rare, such that $v_{\text{pack}} \gg v_{\text{mut}}$ owing to the vastly greater mutation supply rate in the pack lineage. This assumption breaks down for later times when the two lineages are of more comparable frequency. However, we are specifically interested in cases near the critical threshold, where the mutant lineage’s advantage approaches zero when its frequency approaches 50%. Because the two lineages have comparable fitness at this point, the difference in their adaptation rates is still dominated by the mutation supply rate, so proportionality between adaptation rate and frequency still holds. Therefore, this framework is appropriate for analyzing the potential outcomes of a mutant lineage near the critical threshold.

Minimum advantage required for fixation

The differential equation system defined by Eqs. (4) and (5) has no analytical solution, so we must seek an approximation for the critical threshold s_{crit} for a mutation to become fixed. First, consider the dynamics of a beneficial mutation without interference, assuming that it has escaped drift loss while rare. Such a mutation sweeps through a population of size N according to Eq. (5) with constant s , which has the solution:

$$p(t) = \frac{1}{1 + Ne^{-st}} \quad (6)$$

It takes $\log(N)/s$ generations for this mutation to reach 50% frequency. In the lineage-interference model, a mutant lineage near the critical threshold would take roughly twice this long to approach 50%, because its relative advantage spends most of its time decreasing nearly linearly toward $s = 0$ (Fig. 2) and therefore has a time-average advantage of $s_{\text{crit}}/2$.

However, this approximation does not account for any further adaptation achieved by the mutant lineage itself, which occurs mostly at later times when it has reached a substantial frequency. To estimate this effect, consider the integral of p for a constant- s mutation up to the time required to achieve 50% frequency $t_{\text{half}} = \log(N)/s$:

$$\int_0^{\log(N)/s} \frac{1}{1 + Ne^{-st}} dt = \frac{\log(2)}{s} \quad (7)$$

Each lineage's adaptation rate is proportional to p , and so this integral is the cumulative increase in s that results from its ongoing adaptation. And because t_{half} is inversely proportional to s , this integral also represents the decrease in t_{half} caused by its adaptation. Together with the factor of 2 explained above, this process of ongoing adaptation gives an estimate of t_{half} for a mutant lineage near the critical threshold:

$$t_{\text{half}} \approx \frac{2\log(N/2)}{s_{\text{crit}}} \quad (8)$$

At this point, when the mutant lineage approaches 50% frequency, its initial advantage is reduced to zero, which takes roughly s_{crit}/ν generations. Solving for s_{crit} , this gives:

$$s_{\text{crit}} \approx \sqrt{2\nu \log(N/2)} \quad (9)$$

This critical value represents the minimum required fitness difference between a new mutant and the mean fitness of the rest of the population in order for the new mutation to eventually become fixed in a population of size N with an overall rate of adaptation ν . Eq. (9) provides an excellent approximation for numerical solutions of Eqs. (4) and (5), as shown in Fig. 3; it does not depend on assumptions about the magnitude or distribution of mutational effects, although it breaks down for very small population sizes.

Numerical methods and simulations

Numerical solutions throughout were found using Mathematica 10. The `NDSolve` function was used for the lineage interference equations (Eqs. (3) and (4), plotted in Fig. 3), and the `FindMinimum` function was used for the adaptation rate (Eq. (14), plotted in Fig. 4A).

Simulations were written in Python 2.7, using a Wright-Fisher model with discrete generations and infinite alleles. All distinct genotypes were tracked along with their corresponding frequencies, fitness values, and number of mutations. Reproduction was simulated by updating each genotype's population size by drawing from a binomial distribution with N trials and $n_i w_i / \sum_i n_i w_i$ success probability in each trial, where n_i and w_i are the genotype's population size and fitness, respectively. Mutation was simulated by randomly selecting a number of mutant individuals drawn from a binomial

distribution with $\sum_i n_i$ trials and $1 - e^{-\mu}$ success probability in each trial; $\sum_i n_i$ fluctuates slightly about N due to the stochastic implementation of reproduction, and $1 - e^{-\mu}$ is the Poisson probability that an individual receives at least one mutation. To assign a number of mutations to each mutant, numbers were iteratively drawn from the cumulative distribution function of a Poisson distribution with mean μ . Individuals receiving multiple mutations in a single generation were rare, even for the highest mutation rates considered here (up to 10^{-2}). Each mutation was assigned with multiplicative effect on fitness, where the new fitness is $w_i(1 + s_{\text{mut}})$ and the effect size s_{mut} was drawn from an exponential distribution with mean $1/\alpha$.

We explored the evolutionary dynamics of lineages in the multiple-mutations regime using a graph-based implementation of our model (Fig. 1). Here, populations consisted of collections of unique genotypes, each of which was represented as a vertex. Rooted by the ancestral type, each mutation event created a new vertex, which was connected to its parent genotype by an edge. Along with the fitness effects of each mutation, vertices also stored the genotype's abundance at each generation, allowing us to examine each mutation as it arose, spread, and either fixed or was lost. The abundance of a mutation at a given time was equal to the sum of all abundances in the subtree rooted at that vertex. After each generation, vertices were pruned if the corresponding genotype was extinct, had no extant descendants, and had not reached a threshold frequency. Reproduction and mutation were implemented as described above.

Adaptation rates from simulations were estimated by a least-squares linear fit to the population's log-transformed mean fitness as a function of time in generations. One hundred runs of 100,000 generations each were performed for $10^{-9} \leq \mu \leq 10^{-3}$, and 100 runs of 10,000 generations each were performed for $\mu = 10^{-2}$. To remove the transient effect of the initial accumulation of variation in fitness, the fit was started at the earliest point where all individuals had at least one mutation in at least half of the replicate runs; in the long simulations done here, this effect is very small. Substitution rates were measured by fitting a line to the minimum number of mutations (relative to the ancestor) in a population as a function of time in generations, starting at the same point as the fit to the adaptation rate.

RESULTS

Adaptation rate

The lineage-interference framework can be used to compute a population's rate of fitness increase v . In order for a mutation to become fixed, it must occur, survive genetic drift, and survive interference. Following the logic of clonal-interference theory (GERRISH AND LENSKI 1998), we assume that these processes are independent: loss by drift occurs in the first few generations, whereas loss by interference typically occurs in later generations. Taking each of these processes in turn, we compute the expected initial advantage and expected fixation rate for new mutant lineages.

We assume that the effects of beneficial mutations s_{mut} are exponentially distributed, with probability density $\alpha \exp(-\alpha s_{\text{mut}})$, where $1/\alpha$ is the mean effect. This distribution was chosen largely for mathematical convenience, because the exponential is a single-parameter distribution for which mutations of large effect are rare. Theoretical considerations (GILLESPIE 1984; ORR 2002; ORR 2003) and experimental evidence (IMHOF AND SCHLÖTTERER 2001; KASSEN AND BATAILLON 2006; MACLEAN AND BUCKLING 2009; FRENKEL *et al.* 2014) also support using an exponential distribution of beneficial fitness effects.

In the lineage-interference model, the mutant lineage's advantage s_0 depends on both the new beneficial effect and the background fitness. As was done previously (ROUZINE AND COFFIN 2005; DESAI AND FISHER 2007; NEHER *et al.* 2010; GOOD *et al.* 2012), we approximate the distribution of background fitness values as a Gaussian with variance σ^2 . According to Fisher's Fundamental Theorem of Natural Selection (FISHER 1930), a population's rate of fitness increase v is equal to the genetic variance in fitness σ^2 . This relation creates a self-consistency condition that constrains the value of the adaptation rate, which we will use to calculate solutions for v .

The mutant lineage's initial advantage over the population mean fitness is therefore given by the sum of two random variables: a Gaussian with $\sigma^2 = v$ representing background fitness values, and an exponential with mean $1/\alpha$ representing the fitness effects of new mutations. The distribution of their sum is given by their convolution, which is an exponentially modified Gaussian:

$$f(s_0) = \frac{\alpha}{2} \exp\left[\frac{1}{2} \alpha^2 v - \alpha s_0\right] \text{erfc}\left[\frac{\alpha v - s_0}{\sqrt{2v}}\right] \quad (10)$$

where erfc is the complementary error function. For small v , there is negligible effect from the background, and this distribution reduces to an exponential.

After a new beneficial mutation occurs, it must survive genetic drift while it remains at low frequencies. In a large constant-size population where each individual has a Poisson distributed number of offspring, the probability that a mutation with small effect size s_0 survives drift is given by Haldane's formula (HALDANE 1927):

$$\pi(s_0) \approx 2s_0 \quad (11)$$

Finally, in order for a mutation to become fixed, it must survive interference, which we have shown requires $s_0 > s_{\text{crit}}$, where s_{crit} is given by Eq. (9). For multiplicative fitness, the expected contribution to adaptation of a fixed beneficial mutant is:

$$\langle \log(1 + s_0) \rangle = \frac{\int_{s_{\text{crit}}}^{\infty} \log(1 + s_0) f(s_0) \pi(s_0) ds}{\int_{s_{\text{crit}}}^{\infty} f(s_0) \pi(s_0) ds} \quad (12)$$

and the rate at which mutant lineages arise that eventually fix is:

$$\frac{1}{\langle t \rangle} = \mu N \int_{s_{\text{crit}}}^{\infty} f(s_0) \pi(s_0) ds \quad (13)$$

where μ is the beneficial mutation rate per individual per generation. The adaptation rate is then:

$$v = \frac{\langle \log(1 + s_0) \rangle}{\langle t \rangle} \quad (14)$$

where v is the rate of exponential increase in fitness. The right side of Eq. (14) depends on input parameters μ , N , and α , as well as on v itself. Therefore, Eq. (14) cannot be solved explicitly for v , and solutions must be computed numerically. These solutions are compared to individual-based simulations (see Methods). Fig. 4A shows that the lineage-interference model accurately describes adaptation rates in all selection regimes: the low $N\mu$ periodic-selection regime ($N\mu \ll 1/\log(\alpha/N)$, i.e., $\mu < 10^{-8}$ in the simulations) (DESAI AND FISHER 2007), the moderate $N\mu$ clonal-interference regime ($\log(N\mu) \ll 2\log(1/\mu\alpha)$, i.e. $10^{-8} < \mu < 10^{-4}$ in the simulations) (GOOD *et al.* 2012), and the high $N\mu$ multiple-mutations regime where clonal-interference theory fails ($\mu > 10^{-4}$ in the simulations).

We can simplify this expression for v in certain parameter regimes. If the expected fixed advantage is not too large ($\langle s_0 \rangle \ll 1$), and if the population size is not too small ($\sqrt{\log(N/2)} \gg 1$ and $\sqrt{\log(N/2)} \gg \alpha\sqrt{v/2}$), then v can be expressed as a simpler transcendental equation:

$$v \approx 2\mu N \left(s_{\text{crit}}^2 + \frac{2s_{\text{crit}}}{\alpha} + \frac{2}{\alpha^2} \right) \exp \left[-\alpha s_{\text{crit}} + \frac{1}{2} v \alpha^2 \right] \quad (15)$$

This approximation is equivalent to a solution obtained by Good *et al.* (GOOD *et al.* 2012), who used branching-process theory to compute the adaptation rate. A comparison of that work to the lineage-interference model can be found in the Discussion section below.

Substitution rate

By considering a new mutant's fitness that accounts for its genetic background as well as its individual effect, the lineage-interference model provides a convenient framework for

assessing a mutant lineage's fate. However, this treatment complicates the identification of individual fixed mutations because any particular mutant lineage might include additional mutations on the background that have yet to fix, which cannot be directly quantified. Therefore, we must probabilistically estimate the expected number of currently unfixed mutations on the background of a mutation that will likely become fixed.

Consider a new mutant lineage with total advantage s_0 , which is composed of a mutational effect xs_0 and background effect $(1-x)s_0$. Because the two distributions are independent, the non-normalized joint probability density is equal to their product:

$$\frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{((1-x)s_0)^2}{2v}\right] \alpha \exp[-\alpha x s_0] \quad (16)$$

Normalizing this probability density over the interval $0 < x < \infty$ (to infinity because the background advantage can be negative), we have:

$$p(x, s_0) = \sqrt{\frac{2}{\pi v}} s_0 \frac{\exp[-(v\alpha - s_0(1-x))^2/2v]}{1 + \operatorname{erf}[(s_0 - v\alpha)/\sqrt{2v}]} \quad (17)$$

where erf is the error function. Eq. (17) is the probability density of the fraction x of the total mutant advantage s_0 that is attributed to the effect of the mutation itself. The expected fraction $\langle x \rangle$ attributed to a fixed mutational effect is:

$$\langle x \rangle = \int_0^\infty x p(x, \langle s_0 \rangle) dx = 1 - \frac{1}{\langle s_0 \rangle} \left(v\alpha - \sqrt{\frac{2v}{\pi}} \frac{\exp(-z^2)}{(1 + \operatorname{erf}(z))} \right) \quad (18)$$

where $z = (\langle s_0 \rangle - v\alpha)/\sqrt{2v}$. Because any beneficial mutation may in principle be called the “focal” mutation that defines a lineage, the expected number of mutations contained in a fixed lineage is simply $1/\langle x \rangle$. Then, the substitution rate for new mutations is:

$$m = \frac{1}{\langle x \rangle \langle t \rangle} \quad (19)$$

As Fig. 4B shows, this predicted substitution rate agrees closely with results from individual-based simulations (see Methods).

DISCUSSION

We have described a comprehensive conceptual model of competing lineages in asexual populations. The framework for this model is based on four simple, key observations.

- 1) The likely fate of any given beneficial mutation is determined by a combination of its effect size and its background fitness.
- 2) Background fitness effects are directly tied to the overall adaptation rate by Fisher's Fundamental Theorem.
- 3) The adaptation rate of any particular lineage is proportional to its frequency in the population at any given time, which implies a minimum required initial advantage for eventual fixation of mutants.
- 4) Any mutation may be considered one that defines a lineage, which allows for estimating the number of beneficial mutations that a typical successful lineage accumulates on its way to fixation.

This "lineage-interference" model accurately describes the rates of both adaptation and substitution over all selection regimes, including the multiple-mutations regime where clonal-interference theory breaks down (GERRISH AND LENSKE 1998). We have confirmed using simulations that the model works for mutation supply rates up to $N\mu = 10^5$, which is the highest that was computationally feasible; this confirmation even includes high values of μ on the order of $1/\alpha$, where the Gaussian approximation of background fitness values breaks down. Lineage-interference theory also provides a mathematically simple interpretation of interference, which is realized by simply adding up all mutants above a particular threshold (see Eqs. (12) and (13)).

Until recent years, theoretical progress on the multiple-mutations regime had been limited to constant effect-size models, using the "traveling wave" framework (DESAI AND FISHER 2007; ROUZINE *et al.* 2008; NEHER *et al.* 2010). This work provided valuable insights into the transition from stochastic to deterministic fates of new mutants, and the qualitative behavior of adaptation rates with increasing $N\mu$. However, in general real mutation frequency dynamics involves interference among mutations of different effect sizes, which requires incorporating a continuous distribution of effects into the theory.

To our knowledge, there is currently only one other model of adaptation in the multiple-mutations regime that uses a continuous distribution of mutational effects. Good *et al.* (GOOD *et al.* 2012) approximate a sharp transition in the fixation probability for mutations above a particular effect size x_c , which is given by a transcendental integro-differential equation from branching process theory. The sharp transition they describe is directly analogous to the critical threshold s_{crit} in our model—in fact, they obtain an identical large- N approximation of the adaptation rate, where their x_c is equivalent to our s_{crit} . However, by framing the fate of each beneficial mutation directly in the context of its background, we have been able to compute a simple, closed-form solution for the critical advantage $s_{\text{crit}} = \sqrt{2\nu \log(N/2)}$ (see Eqs. (6-9)), which is independent of any assumptions about rates, effect sizes, and distributions of mutations.

Conceptualizing an adapting population as a collection of competing lineages can also help us understand the transition to the multiple-mutations regime at high $N\mu$. As the mutation supply rate increases, multiple mutations occur on the same background before any single one can fix, typically resulting in a better competitor than any single mutant

(Fig. 1). This breaks the adaptation “speed limit” of clonal-interference theory, which only considers the competition between individual mutations. Lineage interference accounts for multiple mutations, and it accurately predicts outcomes on both sides of the regime transition.

The transition to the multiple-mutations regime can be illustrated by the typical number of mutations that comprise a successful lineage (Fig. 5A). At the point where the clonal-interference model breaks down (at $\mu > 10^{-4}$ in the parameterization shown in Figs. 4 and 5), a successful lineage already contains ~ 2 beneficial mutations. This point can be considered the beginning of “nesting” selective sweeps, where a second strong mutation occurs just before the previous one fixes, barely helping to make the first one successful. At even higher $N\mu$, lineages must accumulate additional beneficial mutations earlier and earlier in their respective histories in order to remain competitive. Eventually, sweeps of fixing mutations become more and more highly “nested” until they resemble “cohorts” of mutations sweeping roughly in lockstep. The lineage-interference model may thus help to quantitatively understand the cohorts of multiple beneficial mutations that have been observed in some evolution experiments (LANG *et al.* 2013; MADDAMSETTI *et al.* 2015).

The effects of the multiple-mutations regime can also be seen in the average effect size of a fixed beneficial mutation (Fig. 5B). In the clonal-interference regime, higher beneficial mutation rates result in fixed mutations of larger effect size, as the competition between individual mutations becomes fiercer. But at some point, two individual mutations with smaller beneficial effects are more likely to co-occur in a lineage than any single large effect mutation. As a result, increasing the beneficial mutation rate above some level leads counter-intuitively to the fixation of smaller effect mutations.

The largest anomaly generated by approximation in the lineage-interference model is also well illustrated in Fig. 5B: the model slightly overestimates (by $\sim 15\%$) the average effect size of fixed mutations. Furthermore, predictions of the substitution rate are underestimated by roughly the same amount (see Fig. 4B). These deviations result from the assumption that s_{crit} represents a hard threshold required for fixation of a mutant lineage. More accurately, s_{crit} should be treated as a characteristic value that describes the likelihood of fixation, where some lineages below the threshold become fixed and some above the threshold do not. As a result, the model described here predicts the fixation of slightly fewer mutations of slightly larger effect than those that actually become fixed. Although this approximation does not affect estimates of the overall adaptation rate, it may be possible in future work to relax this assumption to obtain even more accurate descriptions of substitution dynamics.

The lineage-interference model also reveals features of the intermediate $N\mu$ regime that the clonal-interference theory does not capture. To isolate these effects, Fig. 6 shows adaptation rates from the lineage-interference model and from simulations, both plotted relative to the clonal-interference model. Shown this way, simulated adaptation is clearly slower (by $\sim 15\%$ at $N\mu = 1$) than predicted by clonal-interference theory, precisely in the regime that it is meant to describe. In comparison, the lineage-interference model accurately captures this deviation. This deviation indicates that there are systematic

effects of multiple mutations even outside the multiple-mutations regime: specifically, some beneficial mutations can be “wasted” by virtue of their falling on backgrounds of below-average fitness. When interference is not strong, the same mutation would gain negligible advantage by occurring on an above-average fitness background, because it is likely to become fixed anyway by virtue of being “good enough.” The net result is that multiple-mutation effects actually impede adaptation in the intermediate $N\mu$ regime. This effect could be important when interpreting the consequences of mutator alleles in evolving populations (TADDEI *et al.* 1997; SHAVER *et al.* 2002; RAYNES *et al.* 2012; WIELGOSS *et al.* 2012): the potential evolutionary advantage gained by increasing the beneficial mutation rate in this regime could be significantly lessened.

In practice, it is difficult to experimentally test models of adaptation that involve many fixed mutations, because it may require thousands of generations to observe enough selective sweeps to accurately measure adaptation rates. For large population sizes where multiple mutations are most relevant, sweep times are even longer. In Lenski’s long-term evolution experiment (LTEE) with *Escherichia coli*, the adaptation rate was analyzed using a clonal-interference framework (WISER *et al.* 2013). In that study, the upper 95% confidence limit of $N\mu$ was placed at $\sim 2 \times 10^3$ (for populations that had not evolved hypermutability), which would put it barely in the multiple-mutations regime ($N\mu > 10^3$ according to Fig. 4A). Each population in this experiment is given 10 mL of a minimal medium with a low resource concentration and experiences daily 100-fold dilutions; larger laboratory populations (e.g., in rich media, in larger volumes, or in continuous culture without bottlenecks) would have larger effective population sizes and therefore might be well within the multiple-mutations regime.

In order to compare fitness data from laboratory experiments to the lineage-interference model, future research will need to incorporate some quantitative description of epistasis into the theory. Making that challenge more tractable, previous work in several systems has shown that typical populations exhibit a general trend of diminishing-returns epistasis (MACLEAN *et al.* 2010; CHOU *et al.* 2011; KHAN *et al.* 2011; KRYAZHIMSKIY *et al.* 2014), which can be parameterized in a simple way that accurately describes the dynamics of mean fitness (WISER *et al.* 2013). Diminishing-returns epistasis means that the combined effect of beneficial mutations is less than their sum, and lower fitness backgrounds have access to beneficial mutations with larger effect sizes. This form of epistasis would alter the lineage-interference model in three distinct ways. First, the adaptation rate of each lineage would depend not only on its frequency, but also on its mean fitness. Second, $f(s_0)$ would have to be adjusted to account for the non-additivity of fitness effects. Third, v would need to be calculated sequentially for each lineage fixation, because the mean of the distribution of available beneficial effects would decline after each fixation; this third effect was previously incorporated into the clonal-interference framework (WISER *et al.* 2013). It remains to be seen whether these issues can be formally incorporated into the lineage-interference framework, or whether individual-based simulations will have to suffice when dealing with the complications introduced by epistasis.

In conclusion, the lineage-interference model provides a valuable conceptual framework for understanding the likely fates of beneficial mutations and the increased adaptation

rate in the transition to the multiple-mutations regime. We have demonstrated that this model accurately predicts a population's adaptation rate based on a simple integral over all mutant lineages with a fitness advantage above a critical threshold that is proportional to $\sqrt{v}$. This prediction indicates that multiple-mutation effects occur even in the clonal-interference regime, and that they impede adaptation. We have described a probabilistic framework for identifying the likely number of beneficial mutations present in each fixing lineage, which illustrates quantitatively how mutations “team up” when the mutation supply rate is high. We expect this framework will help elucidate the cause of cohorts of beneficial mutations sweeping together, and will be a useful tool for assessing how rapidly large populations adapt to their environments.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation BEACON Center for the Study of Evolution in Action (cooperative agreement DBI-0939454) and by the National Science Foundation (DEB-1451740 to R.E.L.). We are grateful to the Helen Riaboff Whiteley Center and Friday Harbor Laboratories for providing accommodations and working space for our collaboration.

FIGURE LEGENDS

Figure 1. Concurrent sweeps and interference in the multiple-mutations regime. (A) Frequency dynamics of beneficial mutations in a simulated asexual population (see Methods) with size $N = 10^7$, beneficial mutation rate $\mu = 10^{-4}$, and mean beneficial effect $1/\alpha = 0.01$. Mutations that eventually fix (solid lines) often occur on backgrounds with other beneficial mutations that have not yet fixed. There is also intense interference among contending lineages, often resulting in the extinction of beneficial mutations (dashed lines). (B) Corresponding phylogenetic tree. Red branches are genotypes where a beneficial mutation eventually fixes, at a time denoted by a red dot and a fitness increment indicated on the vertical axis. Often, a genotype is no longer extant when its original beneficial mutation becomes fixed, because it has since acquired additional beneficial mutations and its progenitor has been outcompeted.

Figure 2. Lineage interference equations reveal critical threshold required for lineage fixation. Numerical solutions of the lineage-interference model (Eqs. 4-5) for time-dependent values of (A) frequency p , and (B) lineage advantage s ; with initial advantage $s_0 = 0.077 < s_{\text{crit}}$ (red curves), $s_0 = 0.08 > s_{\text{crit}}$ (green curves), and $s_0 = 0.0784 = s_{\text{crit}}$ (black curves) where s is exactly depleted when $p = 0.5$. Population size $N = 10^7$.

Figure 3. Closed-form approximation of the critical threshold for lineage fixation matches numerical solutions. Numerical solutions of s_{crit} from the lineage interference equations (points) closely match the closed-form approximation in Eq. (9) (curves) for

both (A) varying adaptation rate v with $N = 10^7$ (relative error $\sim 0.17\%$ on all points), and (B) varying population size N with $v = 2 \times 10^{-4}$ (relative error $\sim 0.3\%$ on all points with $N \geq 10^4$).

Figure 4. Lineage-interference theory accurately describes rates of adaptation and substitution in simulations over all mutation regimes. (A) Mean adaptation rate and (B) mean fixation rate for individual-based simulations (red symbols) compared to predictions from clonal-interference theory (blue curve) and lineage-interference theory (red curve). Population size $N = 10^7$, mean of beneficial mutation effect distribution $1/\alpha = 0.01$.

Figure 5. Fixed lineages comprise multiple smaller-effect mutations in the multiple-mutations regime. (A) Number of mutations contained in an average fixed lineage predicted by the lineage-interference model, as a function of the beneficial mutation rate. (B) Mean effect size of a fixed beneficial mutation in simulations (red symbols), and the same values predicted by the lineage-interference (red curve) and clonal-interference (blue curve) models as a function of the beneficial mutation rate. Population size $N = 10^7$, mean of beneficial mutation effect distribution $1/\alpha = 0.01$.

Figure 6. Multiple-mutation effects impede adaptation in the clonal-interference regime. The red curve shows the mean adaptation rate predicted by lineage-interference theory divided by the corresponding prediction from clonal-interference theory. The red symbols are simulation results divided by the clonal-interference prediction. Population size $N = 10^7$, mean of beneficial mutation effect distribution $1/\alpha = 0.01$.

LITERATURE CITED

- Brunet, E., I. M. Rouzine and C. O. Wilke, 2008 The stochastic edge in adaptive evolution. *Genetics* 179: 603-620.
- Chou, H.-H., H.-C. Chiu, N. F. Delaney, D. Segrè and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332: 1190-1192.
- Desai, M. M., and D. S. Fisher, 2007 Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* 176: 1759-1798.
- Fisher, R. A., 1930 *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford.
- Frenkel, E. M., B. H. Good and M. M. Desai, 2014 The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. *Genetics* 196: 1217-1226.
- 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., 1984 Molecular evolution over the mutational landscape. *Evolution* 38: 1116-1129.

- Good, B. H., I. M. Rouzine, D. J. Balick, O. Hallatschek and M. M. Desai, 2012 Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl. Acad. Sci. USA* 109: 4950-4955.
- Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection, part V: Selection and mutation. *Math. Proc. Cambridge Philos. Soc.* 23: 838-844.
- Hallatschek, O., 2011 The noisy edge of traveling waves. *Proc. Natl. Acad. Sci. USA* 108: 1783-1787.
- Imhof, M., and C. Schlötterer, 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.
- Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski and T. F. Cooper, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332: 1193-1196.
- Kryazhimskiy, S., D. P. Rice, E. R. Jerison and M. M. Desai, 2014 Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344: 1519-1522.
- Lang, G. I., D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock *et al.*, 2013 Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500: 571-574.
- MacLean, R. C., and A. Buckling, 2009 The distribution of fitness effects of beneficial mutations in *Pseudomonas aeruginosa*. *PLoS Genet.* 5: e1000406.
- MacLean, R. C., G. G. Perron and A. Gardner, 2010 Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* 186: 1345-1354.
- Maddamsetti, R., R. E. Lenski and J. E. Barrick, 2015 Adaptation, clonal interference, and frequency-dependent interactions in a long-term evolution experiment with *Escherichia coli*. *Genetics* 200: 619-631.
- Neher, R. A., B. I. Shraiman and D. S. Fisher, 2010 Rate of adaptation in large sexual populations. *Genetics* 184: 467-481.
- 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.
- Raynes, Y., M. R. Gazzara and P. D. Sniegowski, 2012 Contrasting dynamics of a mutator allele in asexual populations of differing size. *Evolution* 66: 2329-2334.
- Rouzine, I. M., E. 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.
- Rouzine, I. M., and J. M. Coffin, 2005 Evolution of human immunodeficiency virus under selection and weak recombination. *Genetics* 170: 7-18.
- Rouzine, I. M., J. Wakeley and J. M. Coffin, 2003 The solitary wave of asexual evolution. *Proc. Natl. Acad. Sci. USA* 100: 587-592.
- Shaver, A. C., P. G. Dombrowski, J. Y. Sweeney, T. Treis, R. M. Zappala *et al.*, 2002 Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* 162: 557-566.

- Taddei, F., M. Radman, J. Maynard-Smith, B. Toupance, P. H. Gouyon *et al.*, 1997 Role of mutator alleles in adaptive evolution. *Nature* 387: 700-702.
- Wielgoss, S., J. E. Barrick, O. Tenaillon, M. J. Wiser, W. J. Dittmar *et al.*, 2012 Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc. Natl. Acad. Sci. USA* 110: 222-227.
- Wiser, M. J., N. Ribeck and R. E. Lenski, 2013 Long-term dynamics of adaptation in asexual populations. *Science* 342: 1364-1367.