

# Genetics submission

**Title:**

Dynamics and fate of beneficial mutations under lineage contamination by linked deleterious mutations.

**Authors:**

Sophie Péniçon\*, Tanya Singh<sup>†</sup>, Paul Sniegowski<sup>†</sup>, Philip Gerrish<sup>‡,§</sup>

**Author affiliations:**

\*Université Paris-Est, Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS F-94010 Créteil, France

<sup>†</sup>Department of Biology, Leidy Laboratories, 433 S University Ave, University of Pennsylvania, Philadelphia PA 19104 USA

<sup>‡</sup>Theoretical Biology & Biophysics (T-6), Theoretical Division, Los Alamos National Lab, Los Alamos NM 87545 USA

<sup>§</sup>School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive, Atlanta, GA 30332

**Running title:**

Lineage contamination of beneficial mutations

**Corresponding author:**

Philip J Gerrish

School of Biological Sciences  
Georgia Institute of Technology  
950 Atlantic Drive, Atlanta, GA 30332

Theoretical Biology & Biophysics (T-6)  
Theoretical Division  
MailStop K710  
Los Alamos National Lab  
Los Alamos NM 87545 USA

TEL: +1 505 695 0176

email: [pgerrish@gatech.edu](mailto:pgerrish@gatech.edu)

# Dynamics and fate of beneficial mutations under lineage contamination by linked deleterious mutations

Sophie Pénisson\*, Tanya Singh<sup>†</sup>, Paul Sniegowski<sup>‡</sup> and Philip Gerrish<sup>‡ § 1</sup>

\*Université Paris-Est, Laboratoire d'Analyse et de Mathématiques Appliquées (UMR 8050), UPEM, UPEC, CNRS, F-94010 Créteil, France, <sup>†</sup>Department of Biology, Leidy Laboratories, 433 S University Ave, University of Pennsylvania, Philadelphia PA 19104 USA, <sup>‡</sup>Theoretical Biology & Biophysics (T-6), Theoretical Division, Los Alamos National Lab, Los Alamos NM 87545 USA, <sup>§</sup>School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive, Atlanta, GA 30332 USA

**ABSTRACT** Beneficial mutations drive adaptive evolution, yet their selective advantage does not ensure their fixation. Haldane's application of single-type branching process theory showed that genetic drift alone could cause the extinction of newly-arising beneficial mutations with high probability. With linkage, deleterious mutations will affect the dynamics of beneficial mutations and might further increase their extinction probability. Here, we model the lineage dynamics of a newly-arising beneficial mutation as a multitype branching process; this approach allows us to account for the combined effects of drift and the stochastic accumulation of linked deleterious mutations, which we call *lineage contamination*. We first study the lineage contamination phenomenon in isolation, deriving extinction times and probabilities of beneficial lineages. We then put the lineage contamination phenomenon into the context of an evolving population by incorporating the effects of background selection. We find that the survival probability of beneficial mutations is simply Haldane's classical formula multiplied by the correction factor  $e^{-\kappa U/\bar{s}_b}$ , where  $U$  is deleterious mutation rate,  $\bar{s}_b$  is mean selective advantage of beneficial mutations,  $\kappa \in (1, \varepsilon]$ , and  $\varepsilon = 2 - e^{-1}$ . We also find there exists a genomic deleterious mutation rate,  $\hat{U}$ , that maximizes the rate of production of surviving beneficial mutations, and that  $\hat{U} \in [\frac{1}{\varepsilon}\bar{s}_b, \bar{s}_b)$ . Both of these results, and others, are curiously independent of the fitness effects of deleterious mutations. We derive critical mutation rates above which: 1) lineage contamination alleviates competition among beneficial mutations, and 2) the adaptive substitution process all but shuts down.

**KEYWORDS** adaptation; mutation rates; fixation; multitype branching processes

Beneficial mutations are the ultimate source of the genetic variation that fuels evolutionary adaptation, but deleterious mutations are likely to be far more abundant (Muller 1950; Sturtevant 1937). Perhaps for the sake of simplicity, the evolutionary effects of these two types of fitness-affecting mutations were generally considered separately in early studies. For example, Muller (1964) assumed that beneficial mutations were negligible and reasoned verbally that deleterious mutations should have disastrous consequences for populations in the absence of recombination because of the recurrent, stochastic loss of geno-

typic classes with the fewest deleterious mutations—Muller's ratchet (Felsenstein 1974). Haldane (1927), on the other hand, focused on the fate of single beneficial mutations in the absence of other fitness-affecting mutations and used single-type branching process theory to show that most such beneficial mutations are lost to what is now called genetic drift: the fixation probability of such a beneficial mutation is only about twice its selective effect,  $s_b$ , for small  $s_b$ .

In reality, of course, multiple fitness-affecting mutations (both beneficial and deleterious) can be present simultaneously in populations, and these mutations can influence each others' fates and evolutionary effects as a consequence of linkage (reviewed in Gordo and Charlesworth (2001); Charlesworth (2013, 2009); Barton (2009)). Interactions between beneficial and deleterious mutations are of particular interest in this regard, because such

interactions—in contrast to interactions between beneficial mutations alone—can determine whether a population will increase or decrease in fitness. Indeed, recent studies (Bachtrog and Gordo 2004a; Poon and Otto 2000; Goyal *et al.* 2012; Kaiser and Charlesworth 2009; Silander *et al.* 2007) have indicated that beneficial mutations (including reversions of deleterious mutations) can impede or halt the fitness loss predicted in asexual populations under Muller’s ratchet, as originally suggested by Haigh (1978a). Moreover, a number of theoretical studies (Johnson and Barton 2002a; Jiang *et al.* 2011; McFarland *et al.* 2014; Bachtrog and Gordo 2004a; Charlesworth 2013; Peck 1994) have shown that Haldane’s classical fixation probability of  $2s_b$  for a beneficial mutation can be reduced by the effects of selection against linked deleterious mutations (Good and Desai 2014; Hartfield *et al.* 2010; C W Birky and Walsh 1988): in principle, such effects include both background selection against deleterious mutations already present in the genome on which the beneficial mutation appears, and selection against deleterious mutations that arise and accumulate in genomes carrying the beneficial mutation. The latter form of selective effect has not previously been analyzed in isolation; it is the primary focus of the current paper and will be referred to as *lineage contamination*.

In preliminary computer simulations, we observed that the fixation probability of a beneficial mutation appearing in an otherwise initially homogeneous asexual population with a high genomic mutation rate is considerably reduced below Haldane’s classical  $2s_b$  expectation. We hypothesized that the lower probability of fixation of a beneficial mutation in this situation can be attributed to lineage contamination: specifically, Muller’s ratchet operates at a much faster rate in the small lineage founded by the beneficial mutation than in the rest of the population. Here, we present the results of analytical modeling and further computer simulations that support this hypothesis and show how lineage contamination affects fixation probabilities, dynamics of differential load (relative fitness dynamics), sojourn dynamics, and fitness effects of surviving beneficial mutations.

We model the influences of background selection and lineage contamination, both singly and jointly, on the fate of beneficial mutations. Under background selection alone, a beneficial mutation that lands on the best genetic background (the one least loaded with deleterious mutations) always has a non-zero probability of achieving fixation in an asexual population. In contrast, under lineage contamination alone, a beneficial mutation can have a probability of fixation that is zero if the mutation rate is high enough. Our simulations and analytical results suggest that when both background selection and lineage contamination are operating – as they do in real populations – asexual populations traverse a continuum of evolutionary regimes as the genomic mutation rate increases: at low mutation rates, beneficial mutations appear infrequently enough that they do not interfere with each others’ progress to fixation (the “periodic selection” regime; Sniegowski and Gerrish (2010)); as the mutation rate increases, alternative beneficial mutations begin to compete with each other (the “clonal interference” regime); as mutation rate increases further, we find that lineage contamination can suppress a fraction of beneficial mutations that is sufficient to cause a population to revert to the periodic selection regime; ultimately, at very high mutation rates, a regime can be reached in which beneficial mutations are no longer substituting. Significantly, these last two regimes would not obtain without the operation of lineage contamination: at high mutation rates, background selection alone cannot shut down clonal interference or the adap-

tive substitution process, but lineage contamination can. Our results, therefore, indicate that the lineage contamination effect is central to determining the adaptive fate of a population when both beneficial and deleterious mutations are arising (Bull and Wilke 2008; Bull *et al.* 2007; Springman *et al.* 2009).

## Beneficial lineages in a homogeneous population

We model the random accumulation of deleterious mutations within a growing lineage founded by the occurrence of a beneficial mutation (henceforth, *beneficial lineage*). Our main objective in this section is to study the effects of lineage contamination in isolation, and to this end we model beneficial lineages arising within initially homogeneous populations.

We are interested in how the accumulation of *linked* deleterious mutations affects the dynamics and fate of beneficial mutations; as a first approximation, we assume complete asexuality. We assume that relative fitness effects of mutations do not change over the relevant time span, i.e., the environment remains constant over this time span, and there are no frequency-dependent effects other than the one examined here (due to differential rates of Muller’s ratchet). All models assume that fitness effects of mutations are multiplicative, i.e., no epistasis. Finally, our multitype branching process model assumes that reproduction is by binary fission (e.g., bacteria, cell lines): thus, individuals can have a maximum of two offspring. Simulations that relax this assumption give qualitatively similar results; quantitatively, however, fixation probabilities derived from the binary fission model may be approximately halved for the more general model in which numbers of offspring are Poisson-distributed.

Finally, it will facilitate further reading to precisely define three terms: *extinction probability*,  $p_{ext}^b$ , is the probability that a beneficial lineage, arising in an otherwise infinite population, becomes extinct in finite time (the superscript  $b$  indicates that this probability pertains to the beneficial lineage in question and not the whole population); *survival probability* is the complement of the extinction probability:  $p_{svl} = 1 - p_{ext}^b$ ; finally, *fixation probability*,  $p_{fix}$ , is the probability that a lineage will displace the rest of a finite population (becomes fixed) in finite time. We note that we have dropped the superscript in  $p_{svl}$  and  $p_{fix}$ , as reference to the beneficial lineage is implied. We further note that it is possible for an ultimately doomed lineage to become fixed in a finite population, implying  $p_{fix} > p_{svl}$ .

## Multitype branching process model

Our stochastic model is a discrete-time multitype branching process, where a “type”  $i \in \mathbb{N}$  corresponds to the number of acquired deleterious mutations. The model describes the evolution of the composition of the population  $\mathbf{X}_t = (X_{t,0}, X_{t,1}, \dots)$ ,  $X_{t,i}$  being the number of individuals carrying  $i$  deleterious mutations at time  $t \in \mathbb{N}$ . We denote by  $U \geq 0$  the deleterious mutation rate and by  $0 \leq s_b \leq 1$  and  $0 < s_d < 1$  the selective advantage of beneficial mutations and disadvantage of deleterious mutations, respectively. The model can be described as follows: at each time-step, each individual produces two descendants carrying as many deleterious mutations as itself. Each descendant might accumulate during this reproduction  $k$  additional deleterious mutations, with probability  $e^{-U} U^k / k!$ . If the parent was of type  $i$ , the descendant is then of type  $i + k$  and is selected according to its fitness, i.e. with probability proportional to  $(1 + s_b)(1 - s_d)^{i+k}$ . Therefore, an individual of type  $i$  produces a total number of 0, 1 or 2 descendants, each of them being of

a type greater than or equal to  $i$ . We refer to the Supporting Information (SI) for a more detailed description of the model.

We consider a sub-population carrying a beneficial mutation (i.e., a single beneficial lineage) arising in a large wild-type population. In order to study the lineage contamination effect in isolation, we assume that both populations initially do not carry any deleterious mutations. For this purpose we consider two independent branching processes:  $(X_t)_{t \in \mathbb{N}}$  describing the evolution of a wild-type population of initial size  $N$ , hence with  $s_b = 0$  and initial state  $X_0 = (N, 0, 0, \dots)$ , and  $(X_t^b)_{t \in \mathbb{N}}$  describing the evolution of a single beneficial lineage, with  $s_b > 0$  and initial state  $X_0^b = (1, 0, 0, \dots)$ .

### Mean demographic dynamics of each sub-population

The mean wildtype population size (all types combined) at time  $t$  is given by  $N e^{-U t} e^{U(1-(1-s_d)^t)(1-s_d)/s_d}$  (S4). A beneficial mutation occurring within the wildtype population founds a beneficial lineage whose mean size is given by  $(1+s_b)^t e^{-U t} e^{U(1-(1-s_d)^t)(1-s_d)/s_d}$  (S5). Note that as time tends to infinity this quantity tends to  $+\infty$  if  $U < \ln(1+s_b)$ , to  $e^{U(1-s_d)/s_d}$  if  $U = \ln(1+s_b)$ , and to 0 if  $U > \ln(1+s_b)$ . The latter convergence will typically not be monotonic (Figure S1).

### Extinction and survival probabilities

The previous result concerning the evolution of the mean beneficial lineage population size can be refined by looking at the extinction probability of the beneficial lineage. By this we mean the probability  $p_{ext}^b$  that the process  $(X_t^b)_{t \in \mathbb{N}}$  does become extinct, i.e.  $p_{ext}^b = \mathbb{P}(\forall i \in \mathbb{N}, \lim_{t \rightarrow \infty} X_{t,i}^b = 0)$ . We show (Proposition 1; see SI) that although the number of types is infinite, this probability also equals  $\mathbb{P}(\lim_{t \rightarrow \infty} \sum_{i \in \mathbb{N}} X_{t,i}^b = 0)$ . The beneficial lineage almost surely becomes extinct if and only if  $U \geq U_c$ , where the critical deleterious mutation rate is given by

$$U_c = \ln(1+s_b). \quad (1)$$

Of course, this implies that if  $U < U_c$  then  $p_{ext}^b < 1$ , i.e. the beneficial lineage can survive with positive probability. We find that survival probability  $p_{svl} = 1 - p_{ext}^b$  is bounded by:

$$\begin{cases} f_l \leq p_{svl} < f_u, & U < \ln(1+s_b) \\ p_{svl} = 0, & U \geq \ln(1+s_b) \end{cases} \quad (2)$$

where

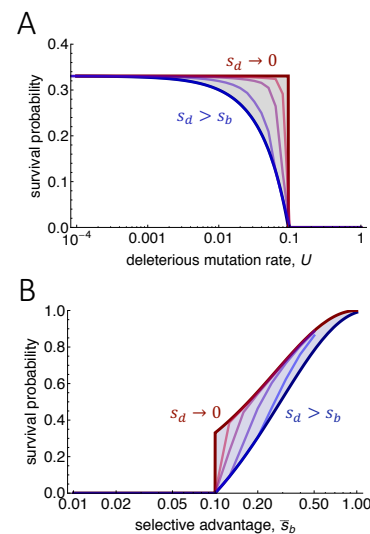
$$f_l = 4 \frac{e^U}{1+s_b} \left(1 - \frac{e^U}{1+s_b}\right) \quad \text{and} \quad f_u = 4 \frac{1}{1+s_b} \left(1 - \frac{1}{1+s_b}\right)$$

Lower bound  $f_l$  is achieved when  $s_d > s_b$ , and the upper bound  $f_u$  is achieved when  $s_d \rightarrow 0$ . Exact computation of  $p_{svl}$  is achieved numerically using algorithm (S11) derived in the SI. Figure 1 plots examples of such computations (thin intermediate curves) as well as limiting cases  $f_l$  and  $f_u$  (thick curves) as a function of the deleterious mutation rate  $U$  and selective advantage  $s_b$ , respectively.

### Fixation probabilities

In this branching process model the fixation probability of the beneficial lineage is the probability  $p_{fix}$  that at some point the whole population carries the beneficial mutation. Note that because we take into account stochastic variation in population size, the beneficial mutation might not be *permanently* established even after fixation, because the population might eventually become

extinct afterward. The fixation probability corresponds here exactly to the probability that the wild-type population dies out before the single beneficial lineage does:  $p_{fix} = \mathbb{P}(T_{ext} < T_{ext}^b)$ , where  $T_{ext} = \inf\{t \in \mathbb{N} : X_t = \mathbf{0}\}$  (alternatively,  $T_{ext}^b$  is the extinction time of the wild-type population (alternatively, beneficial lineage). From what precedes we know that  $T_{ext}$  is almost surely finite, whereas  $T_{ext}^b$  is almost surely finite if and only if  $U \geq U_c$ . Note also that because of the strict inclusion of the probability events  $\{T_{ext}^b = +\infty\} \subsetneq \{T_{ext} < T_{ext}^b\}$ , we know that  $1 - p_{ext}^b < p_{fix}$ . This implies in particular that in this model the fixation probability is never zero. Although we cannot provide a closed-form expression for  $p_{fix}$ , this probability can be well approximated numerically by Equations (S12) - (S13). We illustrate this result in Figure S5 where we plot  $p_{fix}$  as a function of the deleterious mutation rate.



**Figure 1** Survival probabilities as functions of: (A) deleterious mutation rate, and (B) selective advantage of the beneficial mutation. Solid curves plot survival probabilities of a beneficial mutation in an otherwise homogeneous population (lineage contamination in isolation). Blue curves plot the case  $s_d > s_b$ , red curves plot the limit  $s_d \rightarrow 0$ , and curves of intermediate color plot survival probabilities for intermediate values of  $s_d$  (numerical solutions to the survival probability equations given in the SI). Parameters are:  $s_b = 0.1$  (for A), and  $U = 0.1$  (for B).

### Fitness dynamics of a beneficial lineage within a population

The fitness of the beneficial lineage at time  $t$  is given by

$$W(X_t^b) = \sum_{i \in \mathbb{N}} (1+s_b) (1-s_d)^i \frac{X_{t,i}^b}{\sum_{i \in \mathbb{N}} X_{t,i}^b}.$$

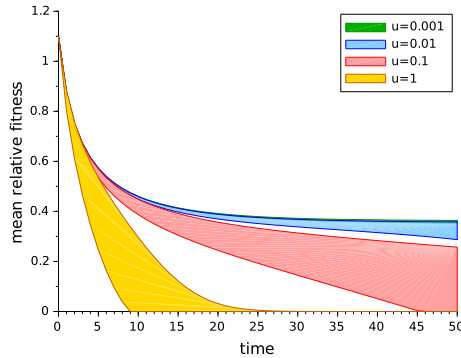
Because of the potential extinction of the population, the random variable  $W(X_t^b)$  is only defined for  $t < T_{ext}^b$ . We similarly define the fitness of the whole population  $W(X_t + X_t^b)$ , and focus our study on the dynamics of the relative fitness  $W(X_t^b)/W(X_t + X_t^b)$ . Because we assume in our model that the wild-type population is initially large, we approximate the relative fitness by its almost certain limit as  $N$  tends to infinity



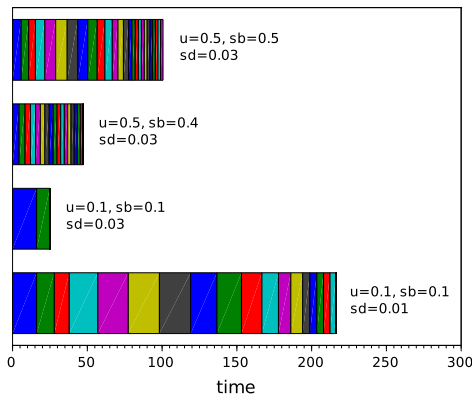
(S14), namely

$$e^{U(1-s_d)(1-(1-s_d)^t)} W(\mathbf{X}_t^b).$$

We also prove that the mean value of this relative fitness tends as time tends to infinity to  $(1+s_b)(1-p_{ext})$  (S15)-(S16). A plot of this long-term limit is given in Figure S6. In order to have a more accurate description of the evolution over time of the relative fitness, we provide in addition an upper and lower bound (S17)-(S18) of its mean value for each  $t \in \mathbb{N}$ , as illustrated in Figure 2.



**Figure 2** Evolution over time of the mean relative fitness of a single beneficial lineage ( $s_b = 0.1$ ) in a large wild-type population, initially homogeneous, with  $s_d = 0.03$ . Upper and lower bounds were obtained as described in the SI. The decrease in relative fitness is caused by the comparatively rapid accumulation of deleterious mutations owing to the small size of emerging beneficial lineages, i.e., lineage contamination.



**Figure 3** Mutational meltdown of beneficial lineages. Sequence  $\mathbb{E}_{\mathbf{n}_0}(T_0), \mathbb{E}_{\mathbf{n}_1}(T_1), \dots$  of the mean extinction times of the least-loaded classes, for one single beneficial lineage.

### Mutational meltdown of a beneficial lineage

Our goal here is to study the synergy between the loss of the least-loaded classes and the potentially decreasing size of the beneficial lineage. For technical reasons detailed in the SI, we consider in this section the continuous-time analog  $(\mathbf{X}_t^b)_{t \geq 0}$  of the branching processes  $(\mathbf{X}_t^b)_{t \in \mathbb{N}}$  studied previously. Assuming

that at time  $t$  the least-loaded class in the beneficial lineage's population is of type  $i$ , the process at this time is of the form  $\mathbf{X}_t^b = \mathbf{n}_i = (0, \dots, 0, n_{ii}, n_{i,i+1}, \dots)$ . Conditionally on  $\mathbf{X}_0^b = \mathbf{n}_i$ , we define the extinction time of the least-loaded class as  $T_i = \inf\{t \geq 0 : X_{t,i}^b = 0\}$ . The mutational meltdown effect is then fully described by the sequence of random variables  $T_0, T_1, \dots$

Note that  $T_1$  strongly depends on the random value  $\mathbf{X}_{T_0}^b$  taken by the process at the beginning of the time interval  $[T_0, T_0 + T_1]$ . Note also that assuming  $\mathbf{X}_{T_0}^b = \mathbf{n}_1$ , the strong Markov property enables study of the process on the latter interval to be reduced to its study on  $[0, T_1]$ , conditionally on  $\mathbf{X}_0^b = \mathbf{n}_1$ . We thus provide in Proposition 2 (SI) an explicit computation of the cumulative distribution function  $\mathbb{P}_{\mathbf{n}_i}(T_i \leq t)$  of the time to extinction of the least-loaded class of type  $i$ , for any  $i$  and any initial condition  $\mathbf{X}_0^b = \mathbf{n}_i$ . From this we deduce its mean value  $\mathbb{E}_{\mathbf{n}_i}(T_i)$ . Again, three different regimes appear depending on whether  $U < U_c$ ,  $U = U_c$  or  $U > U_c$ . We illustrate this result in Figures S2 and S3, where we plot the cumulative distribution function and mean value of the extinction time  $T_0$  of the first least-loaded class, with  $\mathbf{X}_0^b = \mathbf{n}_0 = (1, 0, 0, \dots)$ .

Finally, in order to study not only the behavior of each extinction time separately but to take into account the stochastic evolution of the process  $(\mathbf{X}_t^b)_{t \geq 0}$ , we compute the sequence of the mean extinction times  $\mathbb{E}_{\mathbf{n}_0}(T_0), \mathbb{E}_{\mathbf{n}_1}(T_1), \dots$ , where the deterministic sequence  $(\mathbf{n}_i)_{i \in \mathbb{N}}$  is chosen to reflect as accurately as possible the mean evolution of  $(\mathbf{X}_t^b)_{t \geq 0}$ . We naturally choose  $\mathbf{n}_0 = (1, 0, 0, \dots)$ , and then define  $\mathbf{n}_1$  as the mean value of the process at the end of the first time interval  $[0, T_0]$ . Because this mean value might not be integer-valued, we round each of its coordinates to the closest integer. Hence we set  $\mathbf{n}_1 = \text{round}(\mathbb{E}_{\mathbf{n}_0}(\mathbf{X}_{T_0}^b))$ , and iteratively define in a similar manner  $\mathbf{n}_2, \mathbf{n}_3, \dots$ . As proved in Proposition 2 (SI), we can explicitly compute each  $\mathbb{E}_{\mathbf{n}_i}(\mathbf{X}_{T_i}^b)$ , which combined with the previously mentioned computation of  $\mathbb{E}_{\mathbf{n}_i}(T_i)$  for any initial condition  $\mathbf{n}_i$ , enables us to obtain the desired sequence  $\mathbb{E}_{\mathbf{n}_0}(T_0), \mathbb{E}_{\mathbf{n}_1}(T_1), \dots$ . Figure 3 illustrates this result and provides a visualization of the mutational meltdown effect in a single beneficial lineage for different values of  $U, s_b$  and  $s_d$ .

### Beneficial lineages in an evolving population

Until now, we have examined the process of lineage contamination in isolation; that is, the accumulation of deleterious mutations occurring *after* the production of a beneficial mutation. In addition, we have assumed that we know the selective advantage of the focal mutation.

In real populations, however: 1) deleterious mutations can occur both after and before the appearance of a beneficial mutation, and 2) we generally will not know the selective advantages of beneficial mutations. Deleterious mutations that appear before create a deleterious background upon which the beneficial mutation arises; selection against this deleterious background is *background selection* (Charlesworth *et al.* 1993; Stephan 2010). Here, we model the growth and fate of beneficial mutations of varying selective advantages arising in a population already contaminated with deleterious mutations.

Angled-bracket notation in this section indicates average over all possible trajectories, or “states”, (*ensemble average*) of a beneficial lineage emerging in an otherwise heterogeneous (evolving) population. (The absence of angled-brackets indicates that the focal beneficial lineage arises in an otherwise homogeneous population, as in the previous section.)

### Incorporating background selection

If a beneficial mutation is produced on a background carrying  $j$  deleterious mutations, the initial growth rate of the resulting lineage is:

$$W_j = (1 - \bar{s}_d)^j (1 + S_b) \quad (3)$$

where  $S_b$  is a random variable denoting the selective advantage of the beneficial mutation, with mean  $\bar{s}_b = \mathbb{E}(S_b)$ . Essentially, to incorporate background selection, we simply replace  $1 + s_b$  in the previous section with  $W_j$ .

Of course, we do not know beforehand how many deleterious mutations will be present in the background upon which a beneficial mutation arises. But we do have accurate predictions for both the average number of deleterious mutations in the population, as well as the probability that a beneficial mutation will arise on a background with a given number of deleterious mutations.

When mutation rates are low and population sizes are large, classical theory (Haigh 1978a; Johnson 1999b) predicts that individuals in a population will acquire a Poisson-distributed number of deleterious mutations with parameter  $\theta = U/\bar{s}_d$ . For our purposes, the assumptions of low mutation rate and large population size may be too restrictive, as we wish to explore effects of high mutation rates in finite populations. For exact computation of results, therefore, we will rely on the more encompassing results derived by Gessler (1995) that relax these assumptions, giving the probability of a background having  $j$  mutations (re-derived in SI):

$$\mathcal{P}_{k,\beta}(j) = \frac{\lambda^{j-k}}{(j-k)!} e^{-\lambda} \left( 1 - \frac{\Gamma(\beta-k, \lambda)}{\Gamma(\beta-k)} \right)^{-1} \quad \forall j \geq \beta \quad (4)$$

where  $\lambda = \theta - k$ , and  $k$  and  $\beta$  are integers defined by  $k = \min\{j | N e^{-\theta} \theta^j / j! \geq 1\}$ , and  $\beta = \min\{j | N \mathcal{P}_{k,j}(j) \geq 1\}$ . To derive approximate analytical expressions, where defensible, we nevertheless resort to the straight Poisson distribution from classical theory, the rationale being that tail probabilities lower than  $1/N$  will have negligible effects on the quantities being derived.

### Survival probability

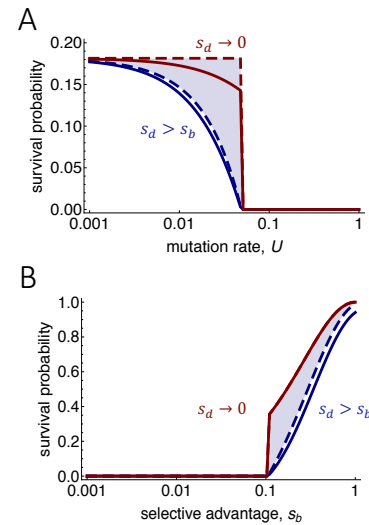
The ensemble-averaged probability of survival is bounded by:

$$\langle p_{svl} \rangle = \begin{cases} K \mathbb{E} \left[ \frac{e^U}{W_l} \left( 1 - \frac{e^U}{W_l} \right) \right], & \bar{s}_d \geq \bar{s}_b \\ K \mathbb{E} \left[ \frac{1}{W_u} \left( 1 - \frac{1}{W_u} \right) \right], & \bar{s}_d \rightarrow 0 \end{cases} \quad (5)$$

where  $K = 4$  for binary fission, and  $K \approx 2$  for Poisson-distributed offspring;  $W_l = (1 - \bar{s}_d)^{J_l} (1 + S_b)$  and  $J_l \sim \mathcal{P}_{k,\beta}(j) \quad \forall j \in [\beta, \hat{j}_l]$ ;  $W_u = (1 - \bar{s}_d)^{J_u} (1 + S_b)$  and  $J_u \sim \mathcal{P}_{k,\beta}(j) \quad \forall j \in [\beta, \hat{j}_u]$ ;  $S_b$  is also a random variable and, for each value of  $S_b$ ,  $\hat{j}_l = \max\{j | W_j > e^U\}$  and  $\hat{j}_u = \max\{j | W_j > 1\}$ , and  $W_j$  is defined by Equation (3). Figure 4 plots exact calculations of  $p_{svl}$  and of  $\langle p_{svl} \rangle$  by fixing  $S_b = s_b$ . Monte Carlo integration of (5), in which  $S_b$  was drawn from an exponential distribution with mean  $\bar{s}_b$ , gives probabilities that are indistinguishable from the approximations we now derive.

### Approximate survival probability

This approximation is suggested by the observation (Figure 4) that fixation probability increases sharply at the critical selective



**Figure 4** Survival probabilities, computed exactly from (5) by fixing  $S_b = s_b$ , plotted as functions of: (A) deleterious mutation rate, and (B) selective advantage of the focal beneficial mutation. Solid curves plot survival probabilities of a beneficial mutation in a heterogeneous (evolving) population. Dashed curves plot survival probabilities in an otherwise homogeneous population (lineage contamination in isolation). Blue curves plot the case  $\bar{s}_d > s_b$ , and red curves plot the limit  $\bar{s}_d \rightarrow 0$ . Parameters are:  $N = 50000$ ,  $s_b = 0.05$  (for A), and  $U = 0.1$  (for B).

advantage,  $s_b^*$ , above which fixation probability becomes positive. We thus explored the possibility that fixation probability might be approximated as simply the probability that  $S_b > s_b^*$  times the probability that the beneficial mutation survives. If the beneficial mutation in question arises on a background with  $j$  deleterious mutations, then  $s_b^* = s_b^*(j) = c_0(1 - \bar{s}_d)^{-j} - 1$ , where  $c_0 = 1$  for  $\bar{s}_d \rightarrow 0$  and  $c_0 = e^U$  for  $\bar{s}_d \geq \bar{s}_b$ , and  $P\{S_b > s_b^*\} = e^{-s_b^*/\bar{s}_b}$ . Taking the logarithm of this probability, multiplying by the corresponding Poisson probabilities and summing over  $j$ , we employ Jensen's inequality to derive expressions providing exact minimums on both upper and lower bounds for the ensemble survival probability:

$$\langle p_{svl} \rangle \geq \begin{cases} K \hat{s}_b e^{-(1/\bar{s}_b)} (e^{U \frac{2-\bar{s}_d}{1-\bar{s}_d}} - 1), & \bar{s}_d \geq \bar{s}_b \\ K \hat{s}_b e^{-(1/\bar{s}_b)} (e^{U \frac{1}{1-\bar{s}_d}} - 1), & \bar{s}_d \rightarrow 0 \end{cases} \quad (6)$$

where  $\hat{s}_b = \bar{s}_b / (1 + \bar{s}_b)^2$ . This expression is a bound-of-bounds and thus of questionable utility. Comparison with simulations reveals the upper bound (for  $\bar{s}_d \rightarrow 0$ ) to be quite accurate but the lower bound (for  $\bar{s}_d \geq \bar{s}_b$ ) to be overly conservative.

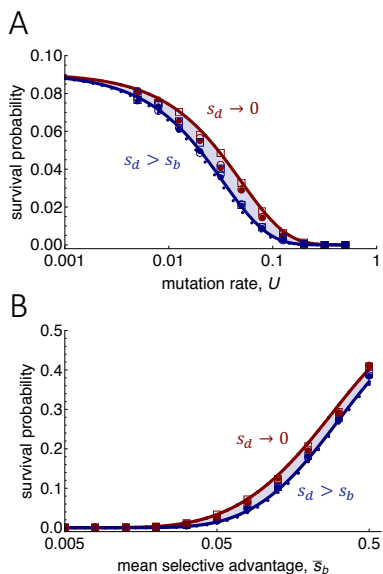
Employing a different approach that does not rely on Jensen's inequality (SI), we find that, to a very good approximation:

$$\langle p_{svl} \rangle \approx \begin{cases} K \hat{s}_b e^{-U/\bar{s}_b} e^{-U/\bar{s}_d (1 - e^{-\bar{s}_d/\bar{s}_b})}, & \bar{s}_d \geq \bar{s}_b \\ K \hat{s}_b e^{-U/\bar{s}_b}, & \bar{s}_d \rightarrow 0 \end{cases} \quad (7)$$

From this, it is apparent that the smallest value of  $\langle p_{svl} \rangle$  is achieved when  $\bar{s}_d = \bar{s}_b$ , so that survival probability is bounded as:

$$K \hat{s}_b e^{-\epsilon U/\bar{s}_b} \lesssim \langle p_{svl} \rangle \lesssim K \hat{s}_b e^{-U/\bar{s}_b}, \quad (8)$$

where  $\varepsilon = 2 - e^{-1} \approx 1.63$ . Remarkably, the foregoing bounds on survival probability are independent of  $\bar{s}_d$ . Comparison with simulations (Figure 5) reveals that the bounds given by Equations (7) and (8) are very accurate. From (7), we can see that the upper bound in (8) is approximated under the wider range of circumstances, because: 1) when  $\bar{s}_d$  is small the upper bound is approximated, and 2) when  $\bar{s}_d \geq \bar{s}_b$  the lower bound obtains, but as  $\bar{s}_d$  becomes increasingly larger than  $\bar{s}_b$ ,  $\langle p_{svl} \rangle$  moves away from the lower bound and back towards the upper bound. There is nevertheless a restricted range of values for  $\bar{s}_d$  – namely when  $\bar{s}_d$  is close to  $\bar{s}_b$  – over which the lower bound obtains.

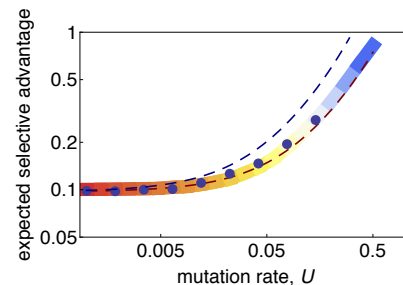


**Figure 5** Ensemble survival probabilities. Solid curves plot Equation (7); dotted curves plot Equation (8). Red dotted curves are not visible because they are indistinguishable from solid red curves; blue dotted curves are very close to solid blue curves. Points each plot the fraction of 15,000 individual-based stochastic simulations in which the focal beneficial mutation survived. Lineage survival was defined as either achieving a size of 5000 or lasting for 5000 generations. Solid points plot simulation results for which no further beneficial mutation occurred ( $c = 0$ ); open circles plot simulations that allowed further beneficial mutations to occur at rate  $c = 0.001$  times the deleterious rate  $U$ ; open squares plot simulations for which  $c = 0.01$ . For panel A, beneficial mutations have mean selective advantage  $\bar{s}_b = 0.05$ ; for panel B deleterious mutation rate is  $U = 0.05$ . When  $\bar{s}_d$  is small, the upper bound obtains (red curve); when  $\bar{s}_d > \bar{s}_b$ , the lower bound obtains (blue curve). In simulations, red points plot the case  $\bar{s}_d = 0.001$  and blue points plot the case  $\bar{s}_d = 0.1$ .

### Selective advantages of surviving beneficial mutations

Because of lineage contamination, beneficial mutations of small effect will have a very small or zero chance of survival; beneficial mutations that do survive, therefore, will tend to be of larger selective advantage. Following logic similar to that of the previous subsection, we derive the ensemble survival probability for a beneficial mutation of given selective advantage  $s_b$ . Given this survival probability, expected selective advantages of surviving beneficial mutation rates, as well as approximate bounds, are

plotted in Figure 6 as a function of deleterious mutation rate.



**Figure 6** Expected selective advantage of surviving beneficial mutations as a function of deleterious mutation rate. Survival probabilities are computed in the same way as in Figure 4, employing Equation (5) and fixing  $S_b = s_b$ . Survival probability at each  $s_b$  was multiplied by  $(1/\bar{s}_b)e^{-s_b/\bar{s}_b}$ , the probability that a mutation of selective advantage  $s_b$  is produced. This product resulted in a unimodal curve, which we then normalized to create a probability density; from this, the expected selective advantage was computed. The thick solid curve has a color gradient indicating log probability of occurrence of beneficial mutations with the plotted selective advantage  $\mathbb{E}_s(S_b)$  (red high, blue low). Expected selective advantages are approximated (dashed lines) as the expected value of a Gamma distribution with shape parameter 2 (Rozen *et al.* 2002; Gerrish and Lenski 1998) and truncated from below at  $e^U - 1$ ; this expected value is  $\bar{s}_b \approx \bar{s}_b + U + \bar{s}_b^2/(\bar{s}_b + U)$ . Parameters are:  $\bar{s}_b = 0.05$ , and  $\bar{s}_d = 0.03$ . Employing the same parameters, each blue dot plots the average selective advantage of surviving beneficial mutations from 15,000 simulations. We note the absence of blue dots above  $U \approx 0.3$ , owing to the fact that none of the 15,000 beneficial mutations survived above this mutation rate (alluding to the “fixation threshold” derived below).

### Mutation rate that maximizes production of surviving beneficial mutations

The recruitment rate of beneficial mutations increases with genomic mutation rate, but because of lineage contamination, the survival probability of beneficial mutations decreases with genomic mutation rate. Therefore, there must exist a genomic mutation rate that maximizes the rate of production of surviving beneficial mutations. Setting  $\partial_U U \langle p_{svl} \rangle = 0$  and solving for  $U$ , we find this maximum production rate occurs at mutation rate  $\hat{U}$ , bounded as:

$$\hat{U} = \begin{cases} [1/\bar{s}_b + (\varepsilon - 1)/\bar{s}_d]^{-1}, & \bar{s}_d \geq \bar{s}_b \\ \bar{s}_b, & \bar{s}_d \rightarrow 0. \end{cases} \quad (9)$$

Figure 7 compares the foregoing predictions to simulation results and shows them to be quite accurate. The smallest value of  $\hat{U}$  is achieved when  $\bar{s}_d = \bar{s}_b$ , resulting in the bounds:

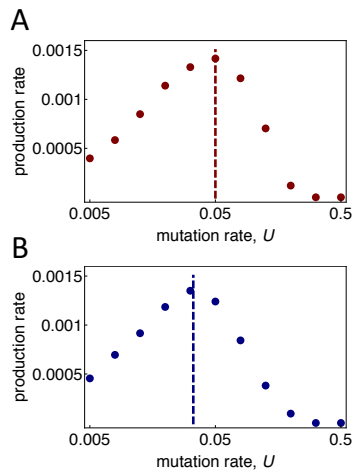
$$\frac{1}{\varepsilon} \bar{s}_b \leq \hat{U} < \bar{s}_b, \quad (10)$$

again displaying a curious independence of  $\bar{s}_d$ . For reasoning similar to that given after Equation (8), the upper bound on  $\hat{U}$  is approximated under the wider range of circumstances.

### Effects of excessive mutation

The lineage contamination effect we describe will increase with increasing deleterious mutation rate. When the mutation rate





**Figure 7** Theory accurately predicts mutation rate  $\hat{U}$  that maximizes production rate of surviving beneficial mutations (9). Beneficial mutations have mean selective advantage  $\bar{s}_b = 0.05$ . Each point represents 15,000 stochastic, individual-based simulations. (A) The weak-deleterious case ( $\bar{s}_d = 0.001$ ) for which  $\hat{U} = \bar{s}_b$  (vertical dashed line). (B) The strong-deleterious case ( $\bar{s}_d = 0.06$ ) for which  $\hat{U} = [1/\bar{s}_b + (\epsilon - 1)/\bar{s}_d]^{-1}$  (vertical dashed line).

is high enough, this effect can cause the within-population mutational meltdown of many newly-arising beneficial lineages causing a reduction in competition and clonal interference. At even higher mutation rates, this effect can suppress most or all newly-arising beneficial lineages, resulting in the partial or complete cessation of adaptive evolution.

### Clonal interference threshold

Evolutionary dynamics may be naturally partitioned into different regimes, depending on the recruitment rate of beneficial mutations. At very low rates of recruitment of beneficial mutations, adaptive evolution proceeds through isolated selective sweeps – a regime that has been dubbed the “periodic selection” regime (Sniegowski and Gerrish 2010). As the recruitment rate of beneficial mutations increases, a point is reached at which two or more alternative beneficial mutations may coexist and compete for fixation (the “clonal interference” regime; Gerrish and Lenski (1998)). As recruitment rate of beneficials continues to increase, it may become likely that competition occurs not among single beneficial mutations but among genotypes carrying multiple beneficial mutations (the “multiple-mutations clonal interference” regime; Desai and Fisher (2007); Desai *et al.* (2007)).

What much of this previous work failed to account for (c.f. Orr (2000); Bachtrog and Gordo (2004b)) was the fact that, as beneficial recruitment rate increases via an increase in overall genomic mutation rate, the rate of deleterious mutation increases in parallel. The findings we have presented so far suggest an intriguing implication of this parallel increase: whereas beneficial recruitment rate increases linearly with genomic mutation rate, survival probability of beneficial mutations decreases exponentially with genomic mutation rate. This fact suggests that, at high genomic mutation rates, the effects of lineage contamination can overwhelm the increased production of beneficial mutations, such that the effective recruitment rate of beneficials

(i.e., the rate of production of *surviving* beneficial mutations) can decrease as mutation rate increases further. As mutation rate increases, therefore, adaptive evolution may eventually revert to a regime in which it proceeds only through isolated selective sweeps; put differently, the population may revert from one of the clonal interference regimes back to the periodic selection regime at high mutation rates.

As delineated in Sniegowski and Gerrish (2010), the clonal interference regime is entered when a second, alternative beneficial mutation is likely to be produced on the ancestral background before the first, or focal, beneficial mutation becomes fixed. Mathematically, this transition occurs when  $N\mu \langle p_{svl} \rangle \ln(N\bar{s}_b/2)/\bar{s}_b = 1$  (SI), where  $\mu = cU$  is the beneficial mutation rate and  $c$  is thus the ratio of beneficial to deleterious rates. In previous work, the transition considered was that which occurs as very low mutation rates increase, and  $\langle p_{svl} \rangle$  was taken to be some function ( $2\bar{s}_b$  or some variant thereof) that was independent of  $U$ . Here, we have shown that, at high mutation rates,  $\langle p_{svl} \rangle$  can depend strongly on  $U$ . We define the “clonal interference threshold” to be the critical mutation rate above which adaptive evolution *reverts* from a clonal interference regime back to the periodic selection regime. This threshold is defined as:

$$U_{ci} = \max\{U \mid N\mu c \langle p_{svl} \rangle \ln(N\bar{s}_b/2)/\bar{s}_b = 1\}, \quad (11)$$

where  $\langle p_{svl} \rangle$  is defined by (8),  $c$  is the ratio of numbers of potential beneficial to deleterious mutations, i.e.,  $c = \mu/U$ , and  $\bar{s}_b = \bar{s}_b + U + \bar{s}_b^2/(\bar{s}_b + U)$ , the expected selective advantage of beneficial mutations that survive (SI).

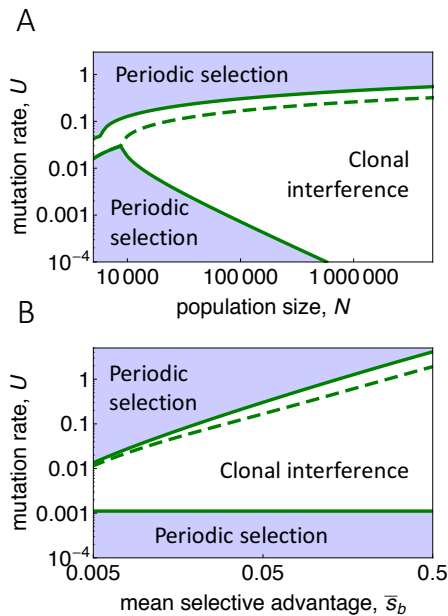
### Fixation threshold

The critical selective advantage below which a beneficial mutation does not survive increases approximately linearly with mutation rate when lineage contamination is considered in isolation, and faster than linearly when background selection is also accounted for. In contrast, the fittest mutation produced by a population has a selective advantage that increases approximately linearly with the *log* of the mutation rate. This necessarily implies that, as mutation rate increases, eventually a point will be reached at which even the selective advantage of the fittest beneficial mutation will not be sufficient to overcome the effects of lineage contamination. This point defines the “fixation threshold”, and its existence follows from the fact that the critical selective advantage required and the maximum selective advantage produced by a population have qualitatively different relationships with mutation rate.

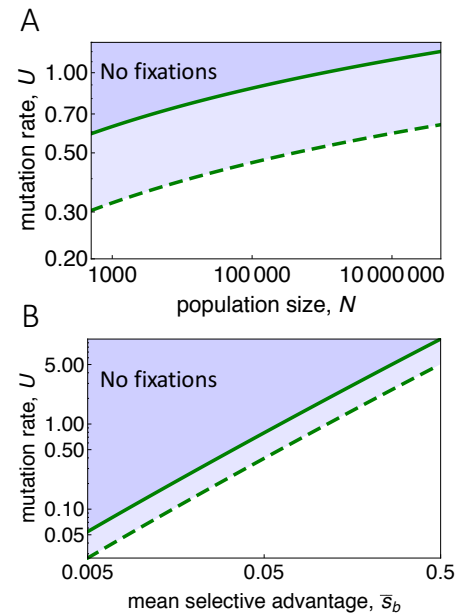
The fixation threshold is exceeded when no beneficial mutation produced by a population has a selective advantage strong enough to survive the effects of lineage contamination. Concretely, in a given interval of time  $\tau$ , we suppose a population produces a total of  $n$  beneficial mutations; then, the fixation threshold is defined as the mutation rate that ensures extinction of even the fittest of these mutations. This critical mutation rate, which defines the fixation threshold and which we will denote by  $U_f$ , thus ensures the extinction of all  $n$  beneficial mutations produced with specified probability  $p_c$ ; it is given by:

$$U_f = \max\{U \mid U = -\hat{U} \ln\left(-\frac{\ln p_c}{n(U)}\right)\}, \quad (12)$$

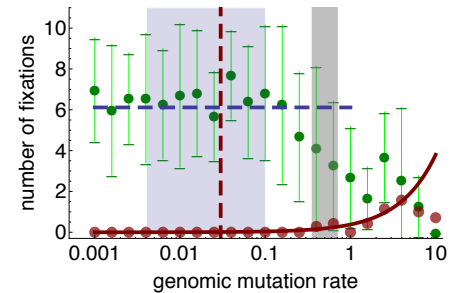
where  $\hat{U}$  is given by (9) and bounded by (10),  $n(U) = N\mu c K \bar{s}_b \tau$ , and  $\tau$  denotes the relevant time period; for example, to compute the mutation rate at which, over a time period of 5000 generations, all fixations will be suppressed with probability 95%, we set  $\tau = 5000$  and  $p_c = 0.95$ .



**Figure 8 Clonal interference threshold.** Mutation rates  $U_{ci}$  that define boundaries between clonal interference and periodic selection regimes:  $U_{ci} = \{U \mid N U c \langle p_{svl} \rangle \ln(N \bar{s}_b / 2) / \bar{s}_b = 1\}$ . The greater of the two solutions (upper green solid curves, given by Equation (11)), represents the transition, as high mutation rate increases, from a clonal interference regime back to a periodic selection regime (the *clonal interference threshold*); the lesser of the two solutions (lower green solid curves) represents the first transition, as low mutation rate increases, from a periodic selection regime into the clonal interference regime. The dashed green curve plots the lower bound (for the case  $\bar{s}_d \geq \bar{s}_b$ ) on the threshold. (A) Parameters are:  $\bar{s}_b = 0.05$  (upper solid curve);  $\bar{s}_b = 0.05$  and  $\bar{s}_d = 0.06$  (lower dashed curve). (B) Parameters are:  $N = 500,000$  (upper solid curve);  $N = 500,000$  and  $\bar{s}_d = 1.1 \times \bar{s}_b$  (lower dashed curve).



**Figure 9 Fixation threshold.** Green solid and dashed curves plot upper and lower bounds of  $U_f$ , as given by (12). (A) Parameters are:  $\bar{s}_b = 0.05$  (upper solid curve);  $\bar{s}_b = 0.05$  and  $\bar{s}_d = 0.06$  (lower dashed curve). (B) Parameters are:  $N = 500,000$  (upper solid curve);  $N = 500,000$  and  $\bar{s}_d = 1.1 \times \bar{s}_b$  (lower dashed curve).



**Figure 10** Total numbers of fixations in simulations of population size 50000 over the course of 2000 generations, as a function of genomic mutation rate (green dots), with error bars of 1.96 standard deviations. Deleterious and beneficial mutation rates were fixed fractions 0.5 and 0.0005 of the genomic mutation rate, respectively. Blue horizontal dashed line indicates the number of fixations above which is the clonal interference regime; light-blue shaded area delineates the mutation rates for which the population is predicted to be in the clonal interference regime. Red vertical dashed line plots the upper-bound mutation rate at which the production rate of surviving beneficial mutations is maximized (closely corresponding the maximum number of fixations), given by Equation (9). Gray vertical bar plots the mutation-rate interval bounded by the predicted “fixation threshold” bounds (12), predicting suppression of all fixations with probability  $p_c = 0.95$ . Red solid curve plots the theoretical neutral expectation: numbers of neutral fixations expected to occur in the same time period; red points plot numbers of neutral fixations in separate simulations. Above the predicted “fixation threshold” (gray bar), numbers of fixations are not statistically distinguishable from the neutral expectation (green bars overlap with red curve).

## Discussion

### Summary

Evolutionary interactions between linked deleterious and beneficial mutations have received increasing attention in recent years. It is now well accepted, for example, that background selection caused by the continual rain of deleterious mutations into regions of low recombination decreases the fixation probability of beneficial mutations (Charlesworth *et al.* 1993; Peck 1994) and decreases nucleotide diversity (Stephan 2010; Kim and Stephan 2000; Birky 1988; Keightley and Otto 2006); moreover, theoretical and empirical studies have shown that selective sweeps of beneficial mutations can cause the fixation of linked deleterious mutations (McDonald *et al.* 2011; Bachtrog and Gordo 2004b; Good and Desai 2014; Hartfield *et al.* 2010; C W Birky and Walsh 1988). To date, work in this area has been focused on populations with relatively low genomic mutation rates. In such populations, the key consideration in analyzing the interaction between beneficial and deleterious mutations is the number of deleterious mutations *already* present in the linked genomic background on which a new beneficial mutation arises. In the current paper, we have focused, in contrast, on populations in which genomic mutation rates may be very high: We have examined the possibility that the genomic background on which a beneficial mutation arises can become progressively contaminated with newly arising deleterious mutations even as the beneficial mutation spreads into the larger population. Our work is motivated in part by numerous studies indicating that adapting asexual populations tend to evolve high mutation rates through genetic hitchhiking (Sniegowski *et al.* 1997, 2000; Johnson 1999a; Elena and Sanjuán 2005; Gentile *et al.* 2011; Söderberg and Berg 2011; M’Gonigle *et al.* 2009; Raynes *et al.* 2011) and by the substantial literature that has been devoted to the question of when the genomic mutation rate will be sufficiently high to cause population extinction (Gerrish *et al.* 2007, 2013; Gerrish and Sniegowski 2012; Bull and Wilke 2008; Bull *et al.* 2007; Springman *et al.* 2009; Biebricher and Eigen 2005; Eigen 2002, 2000, 1971; Eigen and Schuster 1977).

### Multiple beneficial mutations

Our multitype branching process model assumes that beneficial mutations occur infrequently enough that acquiring a second beneficial mutation in linkage with the focal beneficial mutation is improbable in the time required for the focal mutation to either survive or go extinct. In reality, it might be the case that multiple beneficial mutations arise on the same background and sweep to fixation, collectively overcoming the lineage contamination effect.

To assess the strength of our assumption, we studied the effects of allowing additional beneficial mutations to arise at different rates within the lineage founded by the focal beneficial mutation. To this end we varied the parameter  $c$ , introduced above and defined as:  $c = \mu/U$ , or the ratio of numbers of potential beneficial to deleterious mutations. We assessed the effects of doing so in both an extension of our analytical model and in simulations.

To assess the effects of additional within-lineage beneficial mutations on lineage contamination in isolation, we extended our multitype branching process model so as to allow a beneficial lineage to acquire a second beneficial mutation with the same selective advantage  $s_b$  as the first. This additional beneficial mutation could be acquired during reproduction with probability  $1 - e^{-cU}$ . Figure S7 compares, for different values of

$c$ , survival probabilities of a single beneficial lineage in an otherwise homogeneous population (lineage contamination only) as a function of the deleterious mutation rate. It is apparent from this figure that, for reasonable values of  $c$ , there is minimal quantitative difference in survival probabilities and only a slight increase in the apparent threshold. This indicates that our assumption of no additional within-lineage beneficial mutations is a weak assumption. Mathematically, there is a qualitative difference in that, for  $c > 0$ , the critical deleterious mutation rate above which a beneficial lineage becomes extinct almost surely (the “hard” threshold) is twice what it is without the additional beneficial mutation (i.e.,  $2 \ln(1 + s_b)$ ). Practically, this is of little consequence, however, because the survival probabilities are typically minuscule for mutation rates in the region between the hard threshold for which no additional beneficial mutation is allowed ( $\ln(1 + s_b)$ ) and the hard threshold for which one additional beneficial mutation is allowed ( $2 \ln(1 + s_b)$ ). And, while not shown here, survival probabilities become even smaller in regions between higher thresholds that allow more beneficial mutations.

To assess the effects of additional within-lineage beneficial mutations arising in evolving populations, where both background selection and lineage contamination are operating, we performed simulations in which there was technically no limit on the number of additional beneficial mutations. Figure 5 plots survival probabilities computed from simulations, for the cases  $c = 0, 0.001$ , and  $0.01$ , and would seem to indicate, again, that our original assumption of no additional within-lineage beneficial mutation is a very weak assumption. Survival probabilities such as those plotted in Fig. 5 were computed for a range of different parameters and in all cases, for what we considered to be reasonable beneficial-to-deleterious ratios ( $c \leq 0.01$ ), survival probabilities were essentially unaffected by the incorporation of additional beneficial mutations.

### Independence of our results from the selective effects of deleterious mutations

The critical mutation rate above which lineage contamination in isolation ensures extinction of a beneficial mutation, derived in the first section (1), depends only on the selective advantage of the focal beneficial mutation; it does not depend on the selective disadvantages of deleterious mutations. In the next section, when we incorporate background selection, most of the solution bounds we derive are also independent of the selective disadvantages of deleterious mutations. These results stand in contrast to some previously published results that have focused primarily on the effects of background selection. Particularly striking is the contrast between our results, which find a surprising lack of dependence on  $\bar{s}_d$ , and the results of Orr (2000), which instead find a surprising lack of dependence on  $\bar{s}_b$ . For example, we find that  $\hat{U} \approx \bar{s}_b$ , whereas Orr finds that  $\hat{U} = \bar{s}_d$ . A thorough exploration of the relevant parameter space and assessment of Orr’s result is found in Johnson and Barton (2002b). The single factor that accounts for the qualitative discrepancy between our results and Orr’s is lineage contamination: when only background selection is accounted for,  $\hat{U} = \bar{s}_d$ , yet when lineage contamination is also accounted for,  $\hat{U} \approx \bar{s}_b$ . These “opposite” results shine a light on the impact of lineage contamination generally.

### Lineage contamination in nature

The effects of lineage contamination only become significant under linkage and relatively high mutation rates. While we have



focused exclusively on the case of asexuality, lineage contamination should also operate in organisms that undergo some form of genetic exchange as well: the fitness of a newly-arising beneficial mutation will be eroded at a faster rate than the same linkage region in the rest of the population – on principles similar to those studied here. The requirement of high mutation rates would seem to restrict the relevance of our findings to organisms like RNA viruses, although the evolution of high mutation rates has been predicted (Gerrish *et al.* 2007) and increasingly reported in natural (Matic *et al.* 1997) and laboratory (Sniegowski *et al.* 1997; Shaver *et al.* 2002; Denver *et al.* 2009; Wichman 2005; Pal *et al.* 2007; Gentile *et al.* 2011; Chao and Cox 2008; Cox and Gibson 1974) populations of RNA and clonal DNA organisms and in somatic (esp. cancerous) cells. Indeed, the process we have analyzed has been implied in conjunction with background selection as a mechanism that can slow the evolution of tumors (McFarland *et al.* 2013; Solé 2004).

### Lineage contamination, mutational meltdown, and lethal mutagenesis

As alluded to in the presentation of our branching process model, lineage contamination may be thought of as within-population mutational meltdown. If this meltdown is induced by treatment of a population with a mutagenic agent, then it may be thought of as within-population lethal mutagenesis. Put differently, our findings may be thought of as the population-genetic analogues of these processes. Indeed, one of the processes we model – Muller’s ratchet in a growing beneficial lineage – is similar to previous models of Muller’s ratchet in freely-growing populations (Fontanari *et al.* 2003; Bull *et al.* 2007; Bull and Wilke 2008). Our work differs from these previous studies, however, in that we model the fitness erosion of a growing lineage within the context of a larger population.

In a recent experiment, lethal mutagenesis failed to cause extinction in a laboratory population of the bacteriophage T7 (Springman *et al.* 2009) because the accumulation of deleterious mutations opened up new genetic pathways that could increase fitness, i.e., it increased the number of available beneficial mutations. The theory we present here may offer insight into what would be required to thwart the evolutionary rescue afforded by these newly-available beneficial mutations. In particular, our “fixation threshold” might offer an appropriate quantitative guideline for the mutation rate required.

### Lineage contamination and the error threshold

There is an intriguing relationship between our findings and predictions of “error threshold” models (Eigen 2002, 1971; Eigen and Schuster 1977; Biebricher and Eigen 2005; Bagnoli and Bezzi 1998; Nowak and Schuster 1989; Bonhoeffer and Stadler 1993). Generally and somewhat loosely speaking, an error threshold is a critical mutation rate,  $U_{et}$ , above which all genotypes deterministically converge to the same *equilibrium* frequency, independent of their fitness (in the absence of mutational biases).

*Single-peak model.* The simplest model of the error threshold – the so-called “single peak” model – assumes that there is a single fittest genotype of fitness  $1 + s_b$  (the beneficial mutant) and all other (mutationally accessible) genotypes have fitness equal to one (Wiehe 1997; Tejero *et al.* 2011), i.e., the fitness landscape has a set of two possible fitness classes  $\{1, 1 + s_b\}$ . This fitness landscape is obviously unrealistic; its original conjecture may have been based on the fact that many other such “phase transition” phenomena are robust to severe model simplification. When the

number of possible genotypes may be assumed to be infinite, the error threshold is  $U_{et} = \ln(1 + s_b)$  (Wiehe 1997). Curiously, while the assumed fitness landscapes are very different, this critical mutation rate is identical to the one we derive for lineage contamination in isolation (1).

*Multiplicative model.* The set of possible fitness classes on the “multiplicative” fitness landscape is  $\{(1 + s_b)(1 - s_d)^i \mid i \in [0, D]\}$ , where  $D$  is the maximum number of deleterious mutations allowed. Curiously, for the case  $D < \infty$  and  $s_b = s_d$ , the error threshold is  $U_{et} = s_b$  (we note that for small  $s_b$ ,  $U_{et} \approx \ln(1 + s_b)$ , in agreement with the “single-peak” model), whereas when  $D = \infty$ , there is no error threshold:  $U_{et} = \infty$  (Wiehe 1997). Oddly, our lineage contamination model corresponds most closely to the case  $D = \infty$ , for which there is no error threshold, but there is a lineage contamination threshold.

### Concluding remarks

Wittingly or not, the presence of lineage contamination has been implicit in many previous models of mutation-induced fitness erosion. To our knowledge, however, it has not previously been modeled in isolation, as a process separate from background selection. Our theoretical framework partitions these two processes, and allows lineage contamination to be scrutinized separately from other processes. We find, for example, that newly-arising beneficial mutations can be driven extinct almost surely by lineage contamination whereas background selection alone cannot ensure their extinction.

### Acknowledgements

We thank Thomas Bataillon, Guillaume Martin, Alan Perelson, Nick Hengartner, and Thomas Burr for helpful discussions. SP received financial support from the “Soutien à la recherche des jeunes maîtres de conférences” program at the Université Paris-Est Créteil. PG carried out much of this work in, and received financial support from, visiting faculty programs at Aarhus University, Denmark, and at University of Montpellier (Labex/CeMEB/ISEM), France. PS and PG received financial support from NASA grant NNA15BB04A.

### Literature Cited

- Athreya, K. B. and P. E. Ney, 2004 *Branching processes*. Dover Publications, Inc., Mineola, NY, Reprint of the 1972 original [Springer, New York; MR0373040].
- Bachtrog, D. and I. Gordo, 2004a Adaptive evolution of asexual populations under Muller’s ratchet. *Evolution* **58**: 1403–1413.
- Bachtrog, D. and I. Gordo, 2004b Adaptive evolution of asexual populations under Muller’s ratchet. *Evolution* **58**: 1403–1413.
- Bagnoli, F. and M. Bezzi, 1998 Eigen’s error threshold and mutational meltdown in a quasispecies model. *Int J Mod Phys* **9**: 999–1005.
- Barton, N. H., 1995 Linkage and the limits to natural selection. *Genetics* **140**: 821–841.
- Barton, N. H., 2009 Why Sex and Recombination? Cold Spring Harbor Symposia on Quantitative Biology **74**: 187–195.
- Biebricher, C. K. and M. Eigen, 2005 The error threshold. *Virus Research* **107**: 117–127.
- Birky, C. W. J. B., 1988 Effects of linkage on rates of molecular evolution. *Proc. Nat. Acad. Sci. USA* **85**: 6414–6418.
- Bonhoeffer, S. and P. F. Stadler, 1993 Error thresholds on correlated fitness landscapes. *Journal of Theoretical Biology* **164**: 359–372.

- Bull, J. J., R. Sanjuán, and C. O. Wilke, 2007 Theory of lethal mutagenesis for viruses. *Journal of virology* **81**: 2930–2939.
- Bull, J. J. and C. O. Wilke, 2008 Lethal mutagenesis of bacteria. *Genetics* **180**: 1061–1070.
- C W Birky, J. and J. B. Walsh, 1988 Effects of linkage on rates of molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* **85**: 6414–6418.
- Chao, L. and E. Cox, 2008 Competition Between High and Low Mutating Strains of *Escherichia coli*. *Evolution* **37**: 125–134.
- Charlesworth, B., 2009 Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**: 195–205.
- Charlesworth, B., 2013 Background Selection 20 Years on: The Wilhelmine E. Key 2012 Invitational Lecture. *Journal of Heredity* **104**: 161–171.
- Charlesworth, B., M. T. Morgan, and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- Cox, E. C. and T. C. Gibson, 1974 Selection for high mutation rates in chemostats. *Genetics* **77**: 169–184.
- Denver, D. R., Dolan, P. C., L. J. Wilhelm, W. Sung, J. I. Lucas-Lledo, D. K. Howe, S. C. Lewis, K. Okamoto, W. K. Thomas, M. Lynch, and C. F. Baer, 2009 A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proceedings of the National Academy of Sciences* **106**: 16310–16314.
- Desai, M., D. Fisher, and A. Murray, 2007 The speed of evolution and maintenance of variation in asexual populations. *Current Biology* **17**: 385–394.
- Desai, M. M. and D. S. Fisher, 2007 Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**: 1759–1798.
- Eigen, M., 1971 Selforganization of matter and the evolution of biological macromolecules. *Die Naturwissenschaften* **58**: 465–523.
- Eigen, M., 2000 Natural selection: a phase transition? *Biophysical Chemistry* **85**: 101–123.
- Eigen, M., 2002 Error catastrophe and antiviral strategy. *PNAS* **99**: 13374–13376.
- Eigen, M. and P. Schuster, 1977 The hypercycle. A principle of natural self-organization. Part A: Emergence of the hypercycle. *Die Naturwissenschaften* **64**: 541–565.
- Elena, S. F. and R. Sanjuán, 2005 Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *Journal of virology* **79**: 11555–11558.
- Felsenstein, J., 1974 The evolutionary advantage of recombination. *Genetics* **78**: 737–756.
- Fontanari, J. F., A. Colato, and R. S. Howard, 2003 Mutation accumulation in growing asexual lineages. *Physical Review Letters* **91**: 218101.
- Gentile, C. F., S. Yu, S. Serrano, P. J. Gerrish, and P. D. Sniegowski, 2011 Competition between high- and higher-mutating strains of *Escherichia coli*. *Biology letters* **7**: 422–424.
- Gerrish, P. J., A. Colato, A. S. Perelson, and P. D. Sniegowski, 2007 Complete genetic linkage can subvert natural selection. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 6266–6271.
- Gerrish, P. J., A. Colato, and P. D. Sniegowski, 2013 Genomic mutation rates that neutralize adaptive evolution and natural selection. *Journal of The Royal Society Interface* **10**: 20130329–20130329.
- Gerrish, P. J. and R. E. Lenski, 1998 The fate of competing beneficial mutations in an asexual population. *Genetica* **102–103**: 127–144.
- Gerrish, P. J. and P. D. Sniegowski, 2012 Real time forecasting of near-future evolution. *Journal of the Royal Society, Interface / the Royal Society*.
- Gessler, D. D., 1995 The constraints of finite size in asexual populations and the rate of the ratchet. *Genetical research* **66**: 241–253.
- Gillespie, J., 1991 *The Causes of Molecular Evolution*. Oxford Univ. Press, Oxford.
- Good, B. H. and M. M. Desai, 2014 Deleterious Passengers in Adapting Populations. *Genetics* **198**: 1183–1208.
- Gordo, I. and B. Charlesworth, 2001 Genetic linkage and molecular evolution. *Current Biology* **11**: R684–R686.
- Goyal, S., D. J. Balick, E. R. Jerison, R. A. Neher, B. I. Shraiman, and M. M. Desai, 2012 Dynamic mutation-selection balance as an evolutionary attractor. *Genetics* **191**: 1309–1319.
- Haigh, J., 1978a The accumulation of deleterious genes in a population–Muller’s ratchet. *Theoretical Population Biology* **14**: 251–267.
- Haigh, J., 1978b The accumulation of deleterious genes in a population–Muller’s ratchet. *Theoretical Population Biology* **14**: 251–267.
- Haldane, J., 1927 A mathematical theory of natural and artificial selection. Part V. Selection and mutation. *Proc. Camb. Philos. Soc.* **23**: 838–844.
- Hartfield, M., S. Otto, and P. D. Keightley, 2010 The Role of Advantageous Mutations in Enhancing the Evolution of a Recombination Modifier. *Genetics* **184**: 1153–1164.
- Hautphenne, S., G. Latouche, and G. Nguyen, 2013 Extinction probabilities of branching processes with countably infinitely many types. *Adv. in Appl. Probab.* **45**: 1068–1082.
- Jiang, X., Z. Xu, J. Li, Y. Shi, W. Wu, and S. Tao, 2011 The Influence of Deleterious Mutations on Adaptation in Asexual Populations. *PLoS One* **6**: e27757–8.
- Johnson, T., 1999a Beneficial mutations, hitchhiking and the evolution of mutation rates in sexual populations. *Genetics* **151**: 1621–1631.
- Johnson, T., 1999b The approach to mutation-selection balance in an infinite asexual population, and the evolution of mutation rates. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **266**: 2389–2397.
- Johnson, T. and N. H. Barton, 2002a The effect of deleterious alleles on adaptation in asexual populations. *Genetics* **162**: 395–411.
- Johnson, T. and N. H. Barton, 2002b The effect of deleterious alleles on adaptation in asexual populations. *Genetics* **162**: 395–411.
- Kaiser, V. B. and B. Charlesworth, 2009 The effects of deleterious mutations on evolution in non-recombining genomes. *Trends in Genetics* **25**: 9–12.
- Keightley, P. D. and S. P. Otto, 2006 Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**: 89–92.
- Kesten, H. and B. P. Stigum, 1967 Limit theorems for decomposable multi-dimensional Galton-Watson processes. *J. Math. Anal. Appl.* **17**: 309–338.
- Kim, Y. and W. Stephan, 2000 Joint effects of genetic hitchhiking and background selection on neutral variation. *Genetics* **155**: 1415–1427.
- Matic, I., M. Radman, F. Taddei, B. Picard, C. Doit, E. Bingen, E. Denamur, and J. Elion, 1997 Highly variable mutation rates in commensal and pathogenic *E. coli*. *Science (New York, NY)*



- 277: 1833–1834.
- McDonald, M., T. F. Cooper, H. Beaumont, and P. Rainey, 2011 The distribution of fitness effects of new beneficial mutations in *Pseudomonas fluorescens*. *Biology letters* **7**: 98–100.
- McFarland, C. D., K. S. Korolev, G. V. Kryukov, S. R. Sunyaev, and L. A. Mirny, 2013 Impact of deleterious passenger mutations on cancer progression. *Proceedings of the National Academy of Sciences* **110**: 2910–2915.
- McFarland, C. D., L. A. Mirny, and K. S. Korolev, 2014 Tug-of-war between driver and passenger mutations in cancer and other adaptive processes. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 15138–15143.
- M’Gonigle, L. K., J. J. Shen, and S. Otto, 2009 Mutating away from your enemies: The evolution of mutation rate in a host–parasite system. *Theoretical Population Biology* **75**: 301–311.
- Muller, H., 1950 Our load of mutations. *Am. J. Hum. Genet.* **2**: 111–176.
- Muller, H., 1964 The relation of recombination to mutational advance. *Mutation Research* **1**: 2–9.
- Nowak, M. and P. Schuster, 1989 Error thresholds of replication in finite populations mutation frequencies and the onset of Muller’s ratchet. *Journal of theoretical biology* **137**: 375–395.
- Orr, H., 2000 The rate of adaptation in asexuals. *Genetics* **155**: 961–968.
- Pal, C., M. Maciá, A. Oliver, I. Schachar, and A. Buckling, 2007 Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature* **450**: 1079–1081.
- Peck, J. R., 1994 A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* **137**: 597–606.
- Poon, A. and S. P. Otto, 2000 Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution* **54**: 1467–1479.
- Raynes, Y., M. R. Gazzara, and P. D. Sniegowski, 2011 Mutator dynamics in sexual and asexual experimental populations of yeast. *BMC evolutionary biology* **11**: 158.
- Rokyta, D. R., C. J. Beisel, P. Joyce, M. T. Ferris, C. L. Burch, and H. A. Wichman, 2008 Beneficial fitness effects are not exponential for two viruses. *J. Mol. Evol.* **67**: 368–376.
- Rozen, D. E., J. A. G. M. de Visser, and P. J. Gerrish, 2002 Fitness effects of fixed beneficial mutations in microbial populations. *Current biology : CB* **12**: 1040–1045.
- Sewastjanow, B. A., 1975 *Verzweigungsprozesse*. R. Oldenbourg Verlag, Munich-Vienna, Übersetzt aus dem Russischen von Walter Warmuth.
- Shaver, A. C., P. G. Dombrowski, J. Y. Sweeney, T. Treis, R. M. Zappala, and P. D. Sniegowski, 2002 Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* **162**: 557–566.
- Silander, O. K., O. Tenaillon, and L. Chao, 2007 Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS Biology* pp. 1–10.
- Smid, B. and A. J. Stam, 1975 Convergence in distribution of quotients of order statistics. *Stochastic Processes and their Applications* **3**: 287–292.
- Sniegowski, P. D. and P. J. Gerrish, 2010 Beneficial mutations and the dynamics of adaptation in asexual populations. *Phil. Trans. Roy. Soc. Ser. B* **365**: 1255–1263.
- Sniegowski, P. D., P. J. Gerrish, T. Johnson, and A. Shaver, 2000 Evolution of mutation rates: Separating causes from consequences. *BioEssays* **22**: 1057–1066.
- Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski, 1997 Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* **387**: 703–705.
- Söderberg, R. J. and O. G. Berg, 2011 Kick-starting the ratchet: the fate of mutators in an asexual population. *Genetics* **187**: 1129–1137.
- Solé, R., 2004 An error catastrophe in cancer? *Journal of Theoretical Biology* .
- Springman, R., T. Keller, I. J. Molineux, and J. J. Bull, 2009 Evolution at a High Imposed Mutation Rate: Adaptation Obscures the Load in Phage T7. *Genetics* **184**: 221–232.
- Stephan, W., 2010 Genetic hitchhiking versus background selection: the controversy and its implications. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* **365**: 1245–1253.
- Sturtevant, A. H., 1937 Essays on evolution. I. On the effects of selection on mutation rate. *Q. Rev. Biol.* **12**: 467–477.
- Tejero, H., A. Marín, and F. Montero, 2011 The relationship between the error catastrophe, survival of the flattest, and natural selection. *BMC evolutionary biology* **11**: 2.
- Wichman, H., 2005 Adaptive Molecular Evolution for 13,000 Phage Generations: A Possible Arms Race. *Genetics* **170**: 19–31.
- Wiehe, T., 1997 Model dependency of error thresholds: the role of fitness functions and contrasts between the finite and infinite sites models. *Genetical Research* **69**: 127–136.
- Zucca, F., 2011 Survival, extinction and approximation of discrete-time branching random walks. *J. Stat. Phys.* **142**: 726–753.

## Supporting Information

### Beneficial lineages in an initially homogeneous population: multitype branching process model.

Our stochastic model is a discrete-time multi-type branching process with infinite set of types  $\mathbb{N}$ , where a type  $i \in \mathbb{N}$  corresponds to the number of acquired deleterious mutations. Since we do not assume back mutations, the branching process is irreducible. We recall that the vector  $\mathbf{X}_t = (X_{t,0}, X_{t,1}, \dots) \in \mathbb{N}^{\mathbb{N}}$  describes the composition of the population at time  $t \in \mathbb{N}$ ,  $X_{t,i}$  being the number of individuals in the population at time  $t$  carrying  $i$  deleterious mutations. This branching process can be entirely described by its offspring generating function  $\mathbf{f} = (f_0, f_1, \dots)$ , defined for each  $\mathbf{r} = (r_0, r_1, \dots) \in [0, 1]^{\mathbb{N}}$  by  $f_i(\mathbf{r}) = \mathbb{E} \left[ r_0^{X_{1,0}} r_1^{X_{1,1}} \dots \mid \mathbf{X}_0 = \mathbf{e}_i \right]$ , where  $\mathbf{e}_i$  denotes the vector  $(\delta_{ij})_{j \in \mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$  and  $\delta$  stands for the Kronecker delta. By construction, for each  $i \in \mathbb{N}$  and each  $\mathbf{r} \in [0, 1]^{\mathbb{N}}$ ,  $f_i$  only depends on  $r_i, r_{i+1}, \dots$  and

$$f_i(\mathbf{r}) = \left( 1 - \frac{e^{-U} (1 + s_b) (1 - s_d)^i}{2} \sum_{k \in \mathbb{N}} \frac{(U (1 - s_d))^k}{k!} (1 - r_{i+k}) \right)^2. \quad (\text{S1})$$

#### Mean demographic dynamics of each sub-population

The mean matrix  $\mathbf{M} = [m_{ij}]_{i,j \in \mathbb{N}}$  of the branching process, defined such that  $m_{ij} = \mathbb{E} [X_{1,j} \mid \mathbf{X}_0 = \mathbf{e}_i]$  (mean number of individuals of type  $j$  produced by one individual of type  $i$ ) is upper triangular with, for each  $i, k \in \mathbb{N}$ ,

$$m_{i,i+k} = e^{-U} (1 + s_b) (1 - s_d)^i \frac{(U (1 - s_d))^k}{k!}. \quad (\text{S2})$$

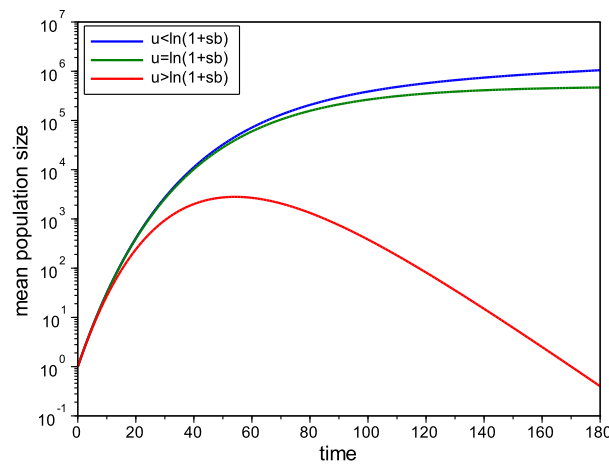
Moreover, for each  $t \in \mathbb{N}$ , the mean matrix of the process at time  $t$  corresponds to the  $t$ -th power of  $\mathbf{M}$ . We denote  $\mathbf{M}^t = [m_{ij}(t)]_{i,j \in \mathbb{N}}$ . It satisfies  $m_{ij}(t) = \mathbb{E} (X_{t,j} \mid \mathbf{X}_0 = \mathbf{e}_i)$  (mean number of individuals of type  $j$  at time  $t$  produced by one individual of type  $i$ ). The matrix is upper triangular and is such that for each  $i, k \in \mathbb{N}$ ,

$$m_{i,i+k}(t) = \left( e^{-U} (1 + s_b) (1 - s_d)^i \right)^t \left( \frac{U (1 - s_d)}{s_d} \right)^k \frac{(1 - (1 - s_d)^t)^k}{k!}. \quad (\text{S3})$$

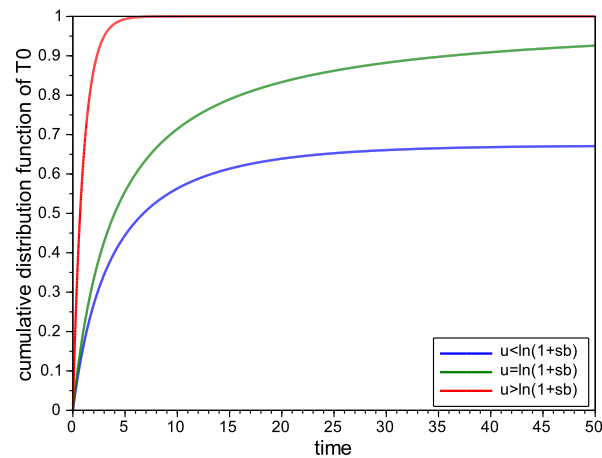
From this we deduce that the mean size of each sub-population at time  $t \in \mathbb{N}$  is given by

$$\mathbb{E} \left( \sum_{i \in \mathbb{N}} X_{t,i} \right) = N \sum_{i \in \mathbb{N}} m_{0,i}(t) = N e^{-Ut} e^{U(1-(1-s_d)^t)(1-s_d)/s_d}, \quad (\text{S4})$$

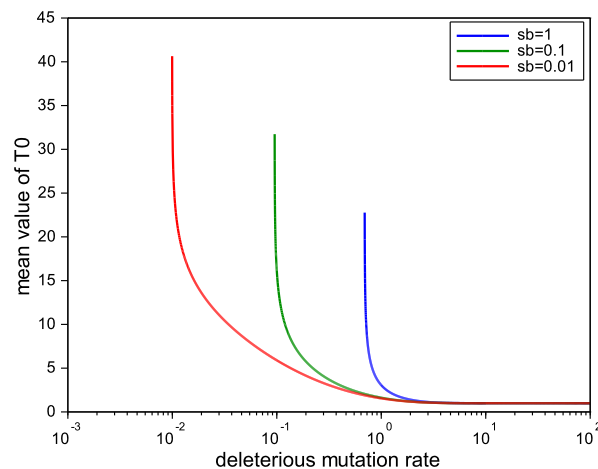
$$\mathbb{E} \left( \sum_{i \in \mathbb{N}} X_{t,i}^b \right) = \sum_{i \in \mathbb{N}} m_{0,i}(t) = (1 + s_b)^t e^{-Ut} e^{U(1-(1-s_d)^t)(1-s_d)/s_d}. \quad (\text{S5})$$



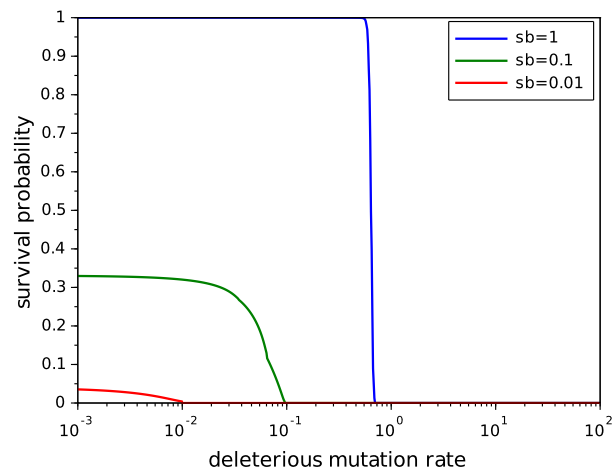
**Figure S1** Evolution over time of the mean population size stemming from one single beneficial lineage, with  $s_d = 0.03$ ,  $s_b = 0.5$ , and  $U = 0.4$  (resp.  $U = \ln(1 + 0.5)$ ,  $U = 0.5$ ). We use a large value for  $s_b$  to show that a beneficial lineage can grow to a large size before being driven extinct by lineage contamination (red curve).



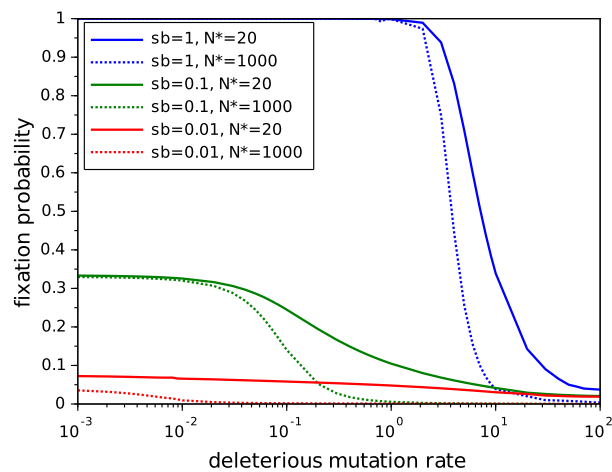
**Figure S2** Cumulative distribution function of the first least-loaded class extinction time  $T_0$ , for one single beneficial lineage, with  $s_b = 0.1$ ,  $s_d = 0.03$ , and  $U = 0.001$  (resp.  $U = \ln(1 + 0.1)$ ,  $U = 10$ ).



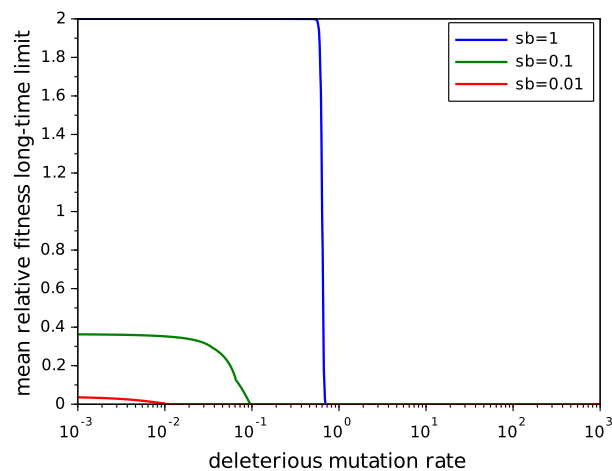
**Figure S3** Mean value of the first least-loaded class extinction time  $T_0$ , for one single beneficial lineage, with  $s_d = 0.03$ . For  $U \leq \ln(1 + s_b)$ , this mean value is infinit. ()



**Figure S4** Survival probability  $1 - p_{ext}^b$  of a single beneficial lineage, with  $s_d = 0.03$ , as a function of the deleterious mutation rate.



**Figure S5** Fixation probability  $p_{fix}$  of a single beneficial lineage in a wild-type population of initial size  $N$ , with  $s_d = 0.03$ . Here  $N^* = N + 1$  corresponds to the initial size of the total population.



**Figure S6** Long-term limit of the mean relative fitness of a single beneficial lineage in a large wild-type population, with  $s_d = 0.03$ .

## Extinction probabilities

### Proposition 1.

- For any  $U \geq 0$  and  $0 < s_b, s_d < 1$ , the two following extinction probabilities of a single beneficial lineage  $(\mathbf{X}_t^b)_{t \in \mathbb{N}}$  are equal:

$$p_{\text{ext}}^b = \mathbb{P}\left(\forall i \in \mathbb{N}, \lim_{t \rightarrow +\infty} X_{t,i}^b = 0\right) = \mathbb{P}\left(\lim_{t \rightarrow +\infty} \sum_{i \in \mathbb{N}} X_{t,i}^b = 0\right). \quad (\text{S6})$$

Moreover,

$$p_{\text{ext}}^b = 1 \iff U \geq \ln(1 + s_b). \quad (\text{S7})$$

- For any  $U \geq 0$ ,  $0 < s_d < 1$  and  $N \in \mathbb{N}$ , a neutral population  $(\mathbf{X}_t)_{t \in \mathbb{N}}$  of initial size  $N$  becomes almost surely (a.s.) extinct:

$$\mathbb{P}\left(\forall i \in \mathbb{N}, \lim_{t \rightarrow +\infty} X_{t,i} = 0\right) = \mathbb{P}\left(\lim_{t \rightarrow +\infty} \sum_{i \in \mathbb{N}} X_{t,i} = 0\right) = 1. \quad (\text{S8})$$

*Proof.* The following proof holds for any initial population size and any  $s_b \geq 0$ . In what follows the notation  $(\mathbf{X}_t)_{t \in \mathbb{N}}$  consequently stands indifferently for a neutral population of initial size  $N \geq 1$  with  $s_b = 0$ , or for a single beneficial lineage with  $s_b > 0$ . If not mentioned otherwise, the probabilities in this proof are computed conditionally on the event  $\{\mathbf{X}_0 = N\mathbf{e}_0\}$ .

The extinction of the whole population implies the extinction of every type, hence the obvious inclusion  $B := \{\lim_{t \rightarrow +\infty} \sum_{i \in \mathbb{N}} X_{t,i} = 0\} \subseteq \{\forall i \in \mathbb{N}, \lim_{t \rightarrow +\infty} X_{t,i} = 0\} =: A$ . However, the simultaneous survival of the whole population and extinction of all the types imply the existence of at least one infinite line of descent with one descendant whose type is indefinitely increasing. We thus obtain by the Markov property that

$$\mathbb{P}(A \setminus B) \leq N \sum_{\substack{0=i_0 < i_1 < i_2 < \dots \\ i_j \in \mathbb{N}}} \mathbb{P}(X_{1,i_1} > 0 \mid \mathbf{X}_0 = \mathbf{e}_{i_0}) \mathbb{P}(X_{1,i_2} > 0 \mid \mathbf{X}_0 = \mathbf{e}_{i_1}) \dots \leq N \sum_{0=i_0 < i_1 < i_2 < \dots} \prod_{k \in \mathbb{N}} m_{i_k, i_{k+1}}.$$

Since for each strictly increasing sequence  $(i_k)_{k \in \mathbb{N}}$  with  $i_0 = 0$  we have  $i_k \geq k$ , it comes by (S2) that

$$\lim_{n \rightarrow +\infty} \prod_{k=0}^n m_{i_k, i_{k+1}} \leq \lim_{n \rightarrow +\infty} e^{-nUs_d} (1 + s_b)^n (1 - s_d)^{\sum_{k=0}^n i_k} \leq \lim_{n \rightarrow +\infty} e^{-nUs_d} (1 + s_b)^n (1 - s_d)^{\frac{n(n+1)}{2}} = 0.$$

Hence  $\mathbb{P}(A) = \mathbb{P}(B)$ , which in turn implies (S6) and the first equality in (S8).

Moreover,  $\mathbb{P}(A) < 1$  if and only if there exists some  $i \in \mathbb{N}$  such that  $\mathbb{P}(\limsup_{t \in \mathbb{N}} X_{t,i} > 0) > 0$ . For each  $i \in \mathbb{N}$ , let us denote  $\tau_i := \inf\{t \in \mathbb{N} : X_{t,i} > 0\}$ . By the Markov property and thanks to the independence of the initial lineages we can write

$$\mathbb{P}(\limsup_{t \in \mathbb{N}} X_{t,i} > 0) = \sum_{k \in \mathbb{N}} \mathbb{P}(X_{\tau_i, i} = k) \left(1 - \left(1 - \mathbb{P}(\limsup_{t \in \mathbb{N}} X_{t,i} > 0 \mid \mathbf{X}_0 = \mathbf{e}_i)\right)^k\right). \quad (\text{S9})$$

Since a mutant of type 0 can produce one mutant of type  $i \in \mathbb{N}$  with positive probability, one can find for each  $k, i \in \mathbb{N}$  an event of the form  $\{\mathbf{X}_0 = N\mathbf{e}_0, \dots, \mathbf{X}_{t-1} = k\mathbf{e}_0, \mathbf{X}_t = k\mathbf{e}_i\} \subseteq \{X_{\tau_i, i} = k\}$  occurring with positive probability, hence implying that  $\mathbb{P}(X_{\tau_i, i} = k) > 0$ . Consequently, (S9) implies that  $\mathbb{P}(A) < 1$  if and only if there exists some  $i \in \mathbb{N}$  such that  $\mathbb{P}(\limsup_{t \in \mathbb{N}} X_{t,i} > 0 \mid \mathbf{X}_0 = \mathbf{e}_i) > 0$ . Seeing the infinite-type branching process  $(\mathbf{X}_t)_{t \in \mathbb{N}}$  as a branching random walk on  $\mathbb{N}$  as in Zucca (2011), the last assertion means that there exists  $i \in \mathbb{N}$  such that so-called local survival of the branching walk at  $i$  occurs with positive probability, starting with one individual at  $i$ . By Theorem 4.1 in Zucca (2011), local survival at  $i$  occurs with positive probability if and only if  $\limsup_{t \in \mathbb{N}} m_{i,i}(t)^{1/t} > 1$ , which by (S3) is equivalent to  $e^{-U}(1 + s_b)(1 - s_d)^i > 1$ . Hence  $\mathbb{P}(A) < 1$  if and only if  $e^{-U}(1 + s_b) > 1$ , which proves (S7) and the second equality in (S8).  $\square$

**Numerical computation of the extinction probability.** We use here the fact that the extinction probability  $p_{\text{ext}}$  can be obtained as the limit of the extinction probability of a multi-type branching process with a finite number of types, when the number of types tends to infinity. We show moreover that this limit value is actually reached for a finite number of types, which consequently leads to an exact computation (see (S11)) rather than an numerical approximation.

First note that if  $U \geq \ln(1 + s_b)$ , then  $p_{\text{ext}} = 1$  by Proposition 1. We thus assume from now on that  $U < \ln(1 + s_b)$ . For each  $D \in \mathbb{N}$ , we shall associate with  $(\mathbf{X}_t^b)_{t \in \mathbb{N}}$  the process  $(\mathbf{X}_t^{b,(D)})_{t \in \mathbb{N}}$  in which none of the individuals with more than  $D$  acquired deleterious mutations are counted. Then  $(\mathbf{X}_t^{b,(D)})_{t \in \mathbb{N}}$  is a multi-type branching process with finite set of types  $\{0, \dots, D\}$  and with offspring generating function  $\mathbf{f}^{(D)}$  given for each  $\mathbf{r} \in [0, 1]^{D+1}$  and  $i \in \{0, \dots, D\}$  by

$$f_i^{(D)}(\mathbf{r}) := f_i(r_0, \dots, r_D, 1, 1, \dots). \quad (\text{S10})$$

Let us define its extinction probability vector  $\mathbf{q}^{(D)} = (q_0^{(D)}, \dots, q_D^{(D)}) \in [0, 1]^{D+1}$  where for each  $i \in \{0, \dots, D\}$ ,

$$q_i^{(D)} := \mathbb{P}\left(\forall j \in \{0, \dots, D\}, \lim_{t \rightarrow +\infty} X_{t,j}^{b,(D)} = 0 \mid \mathbf{X}_0^{b,(D)} = \mathbf{e}_i\right).$$



Let us moreover define  $D_0 := \max\{i \in \mathbb{N} : e^{-U} (1 + s_b) (1 - s_d)^i > 1\}$ . We then have

$$p_{ext} = q_0^{(D_0)}. \quad (S11)$$

Indeed, we know from Hautphenne *et al.* (2013) that  $\mathbf{q}^{(D)}$  converges as  $D \rightarrow +\infty$  to the "partial" extinction probability of  $(\mathbf{X}_t^b)_{t \in \mathbb{N}}$ , which by (S6) is equal to its "global" extinction probability. Hence in particular  $p_{ext}^b = \lim_{D \rightarrow +\infty} q_0^{(D)}$ . Moreover, since for each  $i > D_0$  we have  $m_{ii} < 1$ , a process starting from one mutant of type  $i$  almost surely becomes extinct, which implies that for each  $D \geq D_0$ ,  $\mathbf{q}^{(D)} = (q_0^{(D)}, \dots, q_{D_0}^{(D)}, 1, \dots, 1)$ . Moreover, it is known (see e.g. Satz V.1.4 in Sewastjanow (1975)) that for each  $D \in \mathbb{N}$ ,  $\mathbf{q}^{(D)}$  is the smallest non-negative fixed point of  $\mathbf{f}^{(D)}$ . From what precedes this implies that for each  $D \geq D_0$ ,  $(q_0^{(D)}, \dots, q_{D_0}^{(D)})$  is the smallest non-negative solution of the system

$$r_i = f_i^{(D)}(r_0, \dots, r_{D_0}, 1, \dots, 1) = f_i^{(D_0)}(r_0, \dots, r_{D_0}), \quad 0 \leq i \leq D_0,$$

which is known to be  $\mathbf{q}^{(D_0)}$ . Consequently, for each  $D \geq D_0$ ,  $\mathbf{q}^{(D)} = (q_0^{(D_0)}, \dots, q_{D_0}^{(D_0)}, 1, \dots, 1)$ , leading to the desired result (S11).

### Fixation probabilities

Similarly as for the extinction probability we approximate  $p_{fix}$  by the fixation probability of one single beneficial lineage in a neutral population of initial size  $N$ , when none of the individuals carrying more than  $D$  deleterious mutations are counted. For this purpose we consider for each  $D \in \mathbb{N}$ ,  $(\mathbf{X}_t^{b,(D)})_{t \in \mathbb{N}}$  and  $(\mathbf{X}_t^{(D)})_{t \in \mathbb{N}}$  the multi-type branching processes with offspring generating function (S10) (with  $s_b = 0$  for the process  $(\mathbf{X}_t^{(D)})_{t \in \mathbb{N}}$ ), and such that  $\mathbf{X}_0^{b,(D)} = \mathbf{e}_0$ ,  $\mathbf{X}_0^{(D)} = N\mathbf{e}_0$ . Let  $T_{ext}^{b,(D)}$  and  $T_{ext}^{(D)}$  their respective extinction times. Then, according to Hautphenne *et al.* (2013), for  $t \in \mathbb{N}$  fixed and  $D \rightarrow +\infty$ ,  $\mathbf{X}_t^{b,(D)}$  (resp.  $\mathbf{X}_t^{(D)}$ ) almost surely monotonically converges to  $\mathbf{X}_t^b$  (resp.  $\mathbf{X}_t$ ). Hence  $\lim_{D \rightarrow +\infty} (T_{ext}^{b,(D)}, T_{ext}^{(D)}) \stackrel{a.s.}{=} (T_{ext}^b, T_{ext})$  and in particular

$$p_{fix} = \lim_{D \rightarrow +\infty} \mathbb{P}(T_{ext}^{b,(D)} > T_{ext}^{(D)}). \quad (S12)$$

Let us describe how to compute the probability  $\mathbb{P}(T_{ext}^{b,(D)} > T_{ext}^{(D)})$  for any fixed  $D \in \mathbb{N}$ . First, it can be shown by standard arguments that the multi-type branching process with finite number of types  $(\mathbf{X}_t^{(D)})_{t \in \mathbb{N}}$  almost surely becomes extinct, i.e.  $T_{ext}^{(D)} \stackrel{a.s.}{<} +\infty$ . Consequently,

$$\mathbb{P}(T_{ext}^{b,(D)} > T_{ext}^{(D)}) = \sum_{t \in \mathbb{N}^*} \mathbb{P}(T_{ext}^{(D)} = t) \mathbb{P}(T_{ext}^{b,(D)} > t).$$

Note for instance that for each  $t \in \mathbb{N}$ ,  $\mathbb{P}(T_{ext}^{(D)} \leq t) = \mathbb{P}(|\mathbf{X}_t^{(D)}| = 0) = \mathbb{E}(0^{|\mathbf{X}_t^{(D)}|})$ , and that by independence of the  $N$  lineages,  $\mathbb{P}(T_{ext}^{(D)} \leq t) = \mathbb{E}(0^{|\mathbf{X}_t^{(D)}|} \mid \mathbf{X}_0^{(D)} = \mathbf{e}_0)^N$ . Furthermore (see e.g. Athreya and Ney (2004) Section V.1), the generating function  $\mathbf{r} \mapsto \mathbb{E}(r_0^{X_{t,0}^{(D)}} \dots r_D^{X_{t,D}^{(D)}} \mid \mathbf{X}_0^{(D)} = \mathbf{e}_0)$  of the branching process at time  $t \in \mathbb{N}$  is given by the  $t$ -th iterate of  $\mathbf{f}^{(D)}$  defined by (S10), with  $s_b = 0$ . We denote  $\mathbf{f}^{(D),0} = id$  and for each  $t \in \mathbb{N}$ ,  $\mathbf{f}^{(D),t+1} = \mathbf{f}^{(D)} \circ \mathbf{f}^{(D),t}$ . Therefore,  $\mathbb{P}(T_{ext}^{(D)} \leq t \mid \mathbf{X}_0^{(D)} = \mathbf{e}_0) = f_0^{(D),t}(0, \dots, 0) := a_{D,t}$ , where  $s_b = 0$  in (S1). Similarly, taking  $s_b > 0$  in (S10),  $\mathbb{P}(T_{ext}^{b,(D)} \leq t) = f_0^{(D),t}(0, \dots, 0) := b_{D,t}$ . Hence we finally obtain that

$$\mathbb{P}(T_{ext}^{b,(D)} > T_{ext}^{(D)}) = \sum_{t \in \mathbb{N}^*} (a_{D,t}^N - a_{D,t-1}^N) (1 - b_{D,t}). \quad (S13)$$

### Fitness dynamics of the beneficial lineages within a population

**Large population approximation.** By construction, the population  $(\mathbf{X}_t)_{t \in \mathbb{N}}$  can be written as the sum of  $N$  independent and identically distributed copies of a branching process with offspring generating function (S1) (with  $s_b = 0$ ), and initial state  $\mathbf{e}_0$ , which we write  $\mathbf{X}_t = \sum_{j=1}^N \mathbf{Y}_t^{(j)}$ . The strong law of large numbers combined with (S3) then implies that for each  $t \in \mathbb{N}$ ,  $t < T_{ext}$ ,

$$\begin{aligned} \lim_{N \rightarrow +\infty} W(\mathbf{X}_t) &= \lim_{N \rightarrow +\infty} \frac{\sum_{i \in \mathbb{N}} (1 - s_d)^i \sum_{j=1}^N Y_{t,i}^{(j)}}{\sum_{i \in \mathbb{N}} \sum_{j=1}^N Y_{t,i}^{(j)}} \stackrel{a.s.}{=} \frac{\sum_{i \in \mathbb{N}} (1 - s_d)^i \mathbb{E}(Y_{t,i}^{(1)})}{\sum_{i \in \mathbb{N}} \mathbb{E}(Y_{t,i}^{(1)})} \\ &= \frac{\sum_{i \in \mathbb{N}} (1 - s_d)^i m_{0i}(t)}{\sum_{i \in \mathbb{N}} m_{0i}(t)} = e^{-U(1-s_d)(1-(1-s_d)^t)}. \end{aligned}$$

The same limit would be obtained for  $W(\mathbf{X}_t + \mathbf{X}_t^b)$ , hence the approximation

$$\lim_{N \rightarrow +\infty} \frac{W(\mathbf{X}_t^b)}{W(\mathbf{X}_t + \mathbf{X}_t^b)} \stackrel{a.s.}{=} e^{U(1-s_d)(1-(1-s_d)^t)} W(\mathbf{X}_t^b). \quad (S14)$$

**Long-time limit of the mean relative-fitness.** Let us first assume that  $U \geq \ln(1 + s_b)$ . By Proposition 1,  $T_{ext}^b \stackrel{a.s.}{<} +\infty$ , which approximating the relative fitness by its almost sure limit (S14) leads to

$$\begin{aligned} \lim_{t \rightarrow +\infty} \mathbb{E} \left[ \frac{W(\mathbf{X}_t^b)}{W(\mathbf{X}_t + \mathbf{X}_t^b)} \mathbf{1}_{t < T_{ext}^b} \right] &= \lim_{t \rightarrow +\infty} e^{U(1-s_d)(1-(1-s_d)^t)} \mathbb{E} \left[ W(\mathbf{X}_t^b) \mathbf{1}_{t < T_{ext}^b} \right] \\ &\leq \lim_{t \rightarrow +\infty} e^{U(1-s_d)(1-(1-s_d)^t)} \frac{1+s_b}{s_d} \mathbb{P}(t < T_{ext}^b) = 0, \end{aligned}$$

Since  $p_{ext}^b = 1$ , we have proven the desired result

$$\lim_{t \rightarrow +\infty} \mathbb{E} \left[ \frac{W(\mathbf{X}_t^b)}{W(\mathbf{X}_t + \mathbf{X}_t^b)} \mathbf{1}_{t < T_{ext}^b} \right] = (1 + s_b) (1 - p_{ext}^b). \quad (\text{S15})$$

Let us now assume that  $U < \ln(1 + s_b)$ . In this part of the proof we approximate  $(\mathbf{X}_t^b)_{t \in \mathbb{N}}$  by the multi-type branching process  $(\mathbf{X}_t^{b,(D)})_{t \in \mathbb{N}}$  with finite set of types  $\{0, \dots, D\}$ ,  $D \in \mathbb{N}$ , introduced previously (see (S10)). We assume  $\mathbf{X}_0^{b,(D)} = \mathbf{e}_0$ . Note that its mean matrix corresponds to the truncated matrix  $[m_{i,j}]_{0 \leq i,j \leq D}$ , with largest eigenvalue  $m_{00} > 1$ . Hence  $(\mathbf{X}_t^{b,(D)})_{t \in \mathbb{N}}$  is not irreducible, and we thus make use of Theorem 2.1 in Kesten and Stigum (1967) for reducible branching processes with largest eigenvalue greater than 1 to deduce the existence of some one-dimensional random variable  $Z^{(D)}$  with  $\mathbb{P}(Z^{(D)} = 0) = q_0^{(D)}$  such that for each  $i \in \{0, \dots, D\}$ ,

$$\lim_{t \rightarrow +\infty} \frac{X_{t,i}^{b,(D)}}{e^{-tU} (1 + s_b)^t} \stackrel{a.s.}{=} \frac{1}{i!} \left( \frac{U(1-s_d)}{s_d} \right)^i Z^{(D)}.$$

Defining the absolute fitness  $W(\mathbf{X}_t^{b,(D)})$  and the extinction time  $T_{ext}^{b,(D)}$  similarly as before we thus obtain that

$$\lim_{t \rightarrow +\infty} W(\mathbf{X}_t^{b,(D)}) \mathbf{1}_{t < T_{ext}^{b,(D)}} \stackrel{a.s.}{=} (1 + s_b) \frac{\sum_{i=0}^D \frac{1}{i!} \left( \frac{U(1-s_d)^2}{s_d} \right)^i}{\sum_{i=0}^D \frac{1}{i!} \left( \frac{U(1-s_d)}{s_d} \right)^i} \mathbf{1}_{Z^{(D)} > 0}.$$

By the dominated convergence theorem, the previous convergence also holds in  $L^1$ . Consequently, using the fact that  $p_{ext}^b = \lim_{D \rightarrow +\infty} q_0^{(D)}$ ,

$$\begin{aligned} \lim_{D \rightarrow +\infty} \lim_{t \rightarrow +\infty} \mathbb{E} \left[ W(\mathbf{X}_t^{b,(D)}) \mathbf{1}_{t < T_{ext}^{b,(D)}} \right] &= \lim_{D \rightarrow +\infty} (1 + s_b) \frac{\sum_{i=0}^D \frac{1}{i!} \left( \frac{U(1-s_d)^2}{s_d} \right)^i}{\sum_{i=0}^D \frac{1}{i!} \left( \frac{U(1-s_d)}{s_d} \right)^i} (1 - q_0^{(D)}) \\ &= (1 + s_b) e^{-U(1-s_d)} (1 - p_{ext}^b). \end{aligned}$$

Using again the approximation (S14), we finally obtain

$$\begin{aligned} \lim_{D \rightarrow +\infty} \lim_{t \rightarrow +\infty} \mathbb{E} \left[ \frac{W(\mathbf{X}_t^{b,(D)})}{W(\mathbf{X}_t^{b,(D)} + \mathbf{X}_t)} \mathbf{1}_{t < T_{ext}^{b,(D)}} \right] &= \lim_{D \rightarrow +\infty} \lim_{t \rightarrow +\infty} e^{U(1-s_d)(1-(1-s_d)^t)} \mathbb{E} \left[ W(\mathbf{X}_t^{b,(D)}) \mathbf{1}_{t < T_{ext}^{b,(D)}} \right] \\ &= (1 + s_b) (1 - p_{ext}^b). \end{aligned} \quad (\text{S16})$$

**Upper and lower bounds of the mean relative fitness.** Our goal is to provide upper and lower bounds of the mean relative fitness for each  $t \in \mathbb{N}$ , which as detailed below are both numerically computable by iteration of suitable generating functions. Defining  $|\mathbf{u}| := \sum_{i \in \mathbb{N}} u_i$  for any  $\mathbf{u} \in \mathbb{R}^{\mathbb{N}}$  we have for each  $t \in \mathbb{N}$ , using (S14),

$$\begin{aligned} \mathbb{E} \left[ \frac{W(\mathbf{X}_t^b)}{W(\mathbf{X}_t + \mathbf{X}_t^b)} \mathbf{1}_{t < T_{ext}^b} \right] &= \mathbb{E} \left[ e^{U(1-s_d)(1-(1-s_d)^t)} \frac{\sum_{i \in \mathbb{N}} (1 + s_b) (1 - s_d)^i X_{t,i}^b}{\sum_{i \in \mathbb{N}} X_{t,i}^b} \mathbf{1}_{|\mathbf{X}_t^b| > 0} \right] \\ &\leq e^{U(1-s_d)(1-(1-s_d)^t)} (1 + s_b) \mathbb{P}(|\mathbf{X}_t^b| > 0), \end{aligned} \quad (\text{S17})$$

and, by Cauchy-Schwarz inequality,

$$\begin{aligned} \mathbb{E} \left[ \frac{W(\mathbf{X}_t^b)}{W(\mathbf{X}_t + \mathbf{X}_t^b)} \mathbf{1}_{t < T_{ext}^b} \right] &= e^{U(1-s_d)(1-(1-s_d)^t)} (1 + s_b) \left( \mathbb{P}(|\mathbf{X}_t^b| > 0) - \sum_{i \in \mathbb{N}} (1 - (1 - s_d)^i) \mathbb{E} \left[ \frac{X_{t,i}^b}{|\mathbf{X}_t^b|} \mathbf{1}_{|\mathbf{X}_t^b| > 0} \right] \right) \\ &\geq e^{U(1-s_d)(1-(1-s_d)^t)} (1 + s_b) \left( \mathbb{P}(|\mathbf{X}_t^b| > 0) - \mathbb{E} \left( \frac{1}{|\mathbf{X}_t^b|^2} \mathbf{1}_{|\mathbf{X}_t^b| > 0} \right)^{\frac{1}{2}} \sum_{i \in \mathbb{N}} (1 - (1 - s_d)^i) \mathbb{E} \left( (X_{t,i}^b)^2 \right)^{\frac{1}{2}} \right). \end{aligned} \quad (\text{S18})$$

**Numerical approximation of the mean relative fitness bounds.** Similarly as before, let us for each  $D \in \mathbb{N}$  consider the branching process  $(X_t^{b,(D)})_{t \in \mathbb{N}}$  with offspring generating function  $f^{(D)}$  defined by (S10) (where  $s_b > 0$ ), and let us denote by  $f^{(D),t}$  its  $t$ -th iterate. Then the almost sure monotonous non-decreasing convergence of  $X_t^{b,(D)}$  to  $X_t^b$ , for  $t \in \mathbb{N}$  fixed, entails that the terms appearing in (S17)-(S18) can be for each  $t \in \mathbb{N}$  obtained as follows. First,

$$\mathbb{P}(|X_t^b| > 0) = \lim_{D \rightarrow +\infty} \mathbb{P}(|X_t^{b,(D)}| > 0) = \lim_{D \rightarrow +\infty} (1 - f_0^{(D),t}(0, \dots, 0)).$$

Second,

$$\sum_{i \in \mathbb{N}} (1 - (1 - s_d)^i) \mathbb{E}((X_{t,i}^b)^2) = \lim_{D \rightarrow +\infty} \sum_{i=0}^D (1 - (1 - s_d)^i) \mathbb{E}((X_{t,i}^{b,(D)})^2).$$

By construction, for each  $D \in \mathbb{N}$  and each  $0 \leq i \leq D$ ,  $X_{t,i}^{b,(D)} = X_{t,i}^b$ . Hence, denoting for each  $r \in [0, 1]$

$$g_i^{D,t}(r) := \mathbb{E}(r^{X_{t,i}^{b,(D)}}) = f_0^{(D),t}(1, \dots, r, \dots, 1),$$

we obtain that

$$\sum_{i \in \mathbb{N}} (1 - (1 - s_d)^i) \mathbb{E}((X_{t,i}^b)^2) = \lim_{D \rightarrow +\infty} \sum_{i=0}^D (1 - (1 - s_d)^i) \left[ \frac{d^2 g_i^{D,t}}{dr^2}(1) + \frac{d g_i^{D,t}}{dr}(1) \right].$$

Finally, the dominated convergence theorem implies that

$$\mathbb{E}\left(\frac{1}{|X_t^b|^2} \mathbf{1}_{|X_t^b| > 0}\right) = \lim_{D \rightarrow +\infty} \mathbb{E}\left(\frac{1}{|X_t^{b,(D)}|^2} \mathbf{1}_{|X_t^{b,(D)}| > 0}\right) = \lim_{D \rightarrow +\infty} \sum_{k=1}^{2^t} \frac{1}{k^2} \mathbb{P}(|X_t^{b,(D)}| = k),$$

where for each  $k \in \mathbb{N}$ ,  $\mathbb{P}(|X_t^{b,(D)}| = k)$  is the coefficient of  $r^k$  in the polynomial function

$$r \mapsto \mathbb{E}(r^{|X_t^{b,(D)}|}) = f_0^{(D),t}(r, \dots, r)$$

of degree  $2^t$ . This coefficient can be for instance obtained with a software such as Scilab via the tools `pol` and `coef`.

### Mutational meltdown of a beneficial lineage

We recall that in this section we model the random accumulation of deleterious mutations via the continuous-time analog  $(X_t^b)_{t \geq 0}$  of the branching process  $(X_t^b)_{t \in \mathbb{N}}$  used until now. In this setting, each individual has a random lifespan which is exponentially distributed with parameter 1 (instead of a deterministic lifespan of one time-unit), and produces offspring at the end of the lifespan as previously. This modification allows us to benefit from the fact that the generating functions of continuous-time branching processes are solutions of differential equations, which will in our model lead to a computation of the cumulative distribution function of the fittest class extinction time. The following proposition provides the sequence  $\mathbb{E}_{\mathbf{n}_0}(T_0)$ ,  $\mathbb{E}_{\mathbf{n}_1}(T_1)$ , ... of the mean extinction times of the least loaded-classes, where  $(\mathbf{n}_i)_{i \in \mathbb{N}}$  reflects the mean evolution of  $(X_t^b)_{t \geq 0}$ , that is to say  $\mathbf{n}_0 = (1, 0, 0, \dots)$ , and  $\mathbf{n}_{i+1} = \text{round}(\mathbb{E}_{\mathbf{n}_i}(X_{T_i}^b))$  for each  $i \in \mathbb{N}$ . For each  $i \in \mathbb{N}$ , we define  $a_i := e^{-U}(1 + s_b)(1 - s_d)^i / 2$ .

#### Proposition 2.

- If  $U < \ln(1 + s_b)$ ,

$$\begin{cases} \mathbb{P}_{\mathbf{n}_0}(T_0 \leq t) = 1 - \frac{(1 - 2a_0)e^{-(1-2a_0)t}}{(1 - a_0)^2 - a_0^2 e^{-(1-2a_0)t}}, \\ \mathbb{E}_{\mathbf{n}_0}(T_0) = +\infty. \end{cases} \quad (\text{S19})$$

- If  $U = \ln(1 + s_b)$ ,

$$\begin{cases} \mathbb{P}_{\mathbf{n}_0}(T_0 \leq t) = 1 - \frac{4}{4 + t}, \\ \mathbb{E}_{\mathbf{n}_0}(T_0) = +\infty. \end{cases} \quad (\text{S20})$$

- If  $U > \ln(1 + s_b)$ ,

$$\begin{cases} \mathbb{P}_{\mathbf{n}_0}(T_0 \leq t) = 1 - \frac{(1 - 2a_0)e^{-(1-2a_0)t}}{(1 - a_0)^2 - a_0^2 e^{-(1-2a_0)t}}, \\ \mathbb{E}_{\mathbf{n}_0}(T_0) = \frac{1}{a_0^2} \ln\left(\frac{(1 - a_0)^2}{1 - 2a_0}\right). \end{cases} \quad (\text{S21})$$

More generally, for each  $i \in \mathbb{N}$  and  $j > i$ ,

$$\begin{cases} \mathbb{P}_{\mathbf{n}_i}(T_i \leq t) = \left(1 - \frac{(1-2a_i)e^{-(1-2a_i)t}}{(1-a_i)^2 - a_i^2 e^{-(1-2a_i)t}}\right)^{n_{ii}}, \\ \mathbb{E}_{\mathbf{n}_i}(T_i) = \frac{1}{1-2a_i} \int_0^1 \left(1 - \frac{(1-2a_i)u}{(1-a_i)^2 - a_i^2 u}\right)^{n_{ii}} u^{-1} du, \\ \mathbb{E}_{\mathbf{n}_i}(X_{T_{i,j}}) = \sum_{\substack{i \leq k \leq j \\ 0 \leq l \leq j-k}} \frac{(-1)^l n_{ik} (1-2a_{k+l}) (U(1-s_d) s_d^{-1})^{j-k}}{(1-2a_i) l! (j-k-l)!} \int_0^1 u^{\frac{1-2a_{k+l}}{1-2a_i}-1} \left(1 - \frac{(1-2a_i)u}{(1-a_i)^2 - a_i^2 u}\right)^{n_{ii}} du. \end{cases} \quad (\text{S22})$$

Note that this prove in particular that  $\lim_{U \rightarrow +\infty} \mathbb{E}_{\mathbf{n}_0}(T_0) = 1$ , as we can see in Figure S3.

*Proof.* Let  $i \in \mathbb{N}$  and  $\mathbf{n}_i = (0, \dots, 0, n_{ii}, n_{i,i+1}, \dots) \in \mathbb{N}^{\mathbb{N}}$ . Note that under  $\mathbb{P}_{\mathbf{n}_i}$ , the law of the extinction time  $T_i$  does not depend on  $n_{ij}$ ,  $j > i$ , since none of the individuals of type  $j$  can produce individuals of type  $i$ . Under  $\mathbb{P}_{\mathbf{n}_i}$ , the extinction time  $T_i$  actually has the same distribution as the extinction time of a single-type branching process  $(Y_t)_{t \in \mathbb{R}_+}$  in which only the individuals of type  $i$  are counted, conditionally on  $\{Y_0 = n_{ii}\}$ . By construction, the offspring generating function of  $(Y_t)_{t \in \mathbb{R}_+}$  is given for each  $r \in [0, 1]$  by  $g(r) = f_i(r, 1, 1, \dots) = (1 - a_i + a_i r)^2$ . For each  $t \in \mathbb{R}_+$ , let us define  $p_i(t) := \mathbb{P}(Y_t = 0 \mid Y_0 = 1)$ . Then for each  $t \in \mathbb{R}_+$ ,

$$\mathbb{P}_{\mathbf{n}_i}(T_i \leq t) = \mathbb{P}_{\mathbf{n}_i}(X_{t,i} = 0) = \mathbb{P}(Y_t = 0 \mid Y_0 = n_{ii}) = p_i(t)^{n_{ii}}. \quad (\text{S23})$$

Moreover (see for instance III.3. in [Athreya and Ney \(2004\)](#)),  $p_i(t) = \mathbb{P}(Y_t = 0 \mid Y_0 = 1)$  is solution of the differential equation

$$y' = g(y) - y = (1 - a_i)^2 + (2a_i(1 - a_i) - 1)y + a_i^2 y^2, \quad (\text{S24})$$

with the boundary condition  $p_i(0) = 0$ . Note that (S24) is a Riccati equation for which one solution ( $y(t) = 1$ ) is known.

If  $U < \ln(1 + s_b)$  then  $2a_0 > 1$  and solving (S24) leads to

$$p_0(t) = 1 - \frac{(2a_0 - 1)e^{(2a_0-1)t}}{a_0^2 e^{-(2a_0-1)t} - (1 - a_0)^2},$$

which combined with (S23) entails (S19). In particular,  $\mathbb{P}_{\mathbf{n}_0}(T_0 < +\infty) = 1 - \frac{2a_0-1}{a_0^2} < 1$  and thus  $\mathbb{E}_{\mathbf{n}_0}(T_0) = +\infty$ .

If  $U = \ln(1 + s_b)$  then  $2a_0 = 1$  and (S24) leads to

$$p_0(t) = 1 - \frac{4}{4 + t},$$

leading to (S20) and immediately implying that  $\mathbb{E}_{\mathbf{n}_0}(T_0) = +\infty$ .

Finally, if  $U > \ln(1 + s_b)$  then for each  $i \in \mathbb{N}$ ,  $2a_i < 1$ , and solving (S24) leads to

$$p_i(t) = 1 - \frac{(1-2a_i)e^{-(1-2a_i)t}}{(1-a_i)^2 - a_i^2 e^{-(1-2a_i)t}},$$

hence the first equality in (S21) and (S22). Since  $T_i$  is a positive random variable, its first moment is given by  $\mathbb{E}_{\mathbf{n}_i}(T_i) = \int_{\mathbb{R}_+} (1 - \mathbb{P}_{\mathbf{n}_i}(T_i \leq t)) dt$ . Integrating by substitution we obtain the second equality in (S21) and (S22). Note that a more explicit but rather bulky formula could be found for  $\mathbb{E}_{\mathbf{n}_i}(T_i)$ , but for the sake of simplicity we choose to keep the definite integral form which is directly computable with a software such as Scilab. We moreover deduce from the first equality in (S22) that for each  $j > i$ ,

$$\mathbb{E}_{\mathbf{n}_i}(X_{T_{i,j}}) = \int_{\mathbb{R}_+} \mathbb{E}_{\mathbf{n}_i}(X_{t,j}) (p_i(t)^{n_{ii}})' dt. \quad (\text{S25})$$

In the continuous-time setting, the mean composition  $\mathbb{E}_{\mathbf{x}_0}(\mathbf{X}_t)$  of the multi-type branching process at time  $t \in \mathbb{R}_+$ , conditionally on  $\{\mathbf{X}_0 = \mathbf{x}_0\}$ ,  $\mathbf{x}_0 \in \mathbb{N}^{\mathbb{N}}$ , is given by the vector  $\mathbf{x}_0 e^{(\mathbf{M}-\mathbf{I})t}$  (see for instance [Athreya and Ney \(2004\)](#)), where  $e$  denotes the matrix exponential. Hence by (S3) we obtain that for each  $t \in \mathbb{R}_+$ ,

$$\begin{aligned} \mathbb{E}_{\mathbf{n}_i}(X_{t,j}) &= \sum_{k=i}^j n_{ik} \left[ e^{(\mathbf{M}-\mathbf{I})t} \right]_{kj} = \sum_{k=i}^j n_{ik} e^{-t} \sum_{n \in \mathbb{N}} \frac{t^n}{n!} m_{kj}(n) \\ &= \sum_{k=i}^j n_{ik} e^{-t} \sum_{n \in \mathbb{N}} \frac{(2a_k t)^n}{n!} \frac{(U(1-s_d) s_d^{-1})^{j-k}}{(j-k)!} \sum_{l=0}^{j-k} \binom{j-k}{l} (-1)^l (1-s_d)^{nl} \\ &= \sum_{k=i}^j \sum_{l=0}^{j-k} \binom{j-k}{l} \frac{(-1)^l n_{ik} (U(1-s_d) s_d^{-1})^{j-k}}{(j-k)!} e^{-(1-2a_{k+l})t}. \end{aligned}$$

Integrating by part (S25) next leads to

$$\mathbb{E}_{n_i} \left( X_{T_i, j} \right) = \sum_{k=i}^j \sum_{l=0}^{j-k} \binom{j-k}{l} \frac{(-1)^l n_{ik} (1 - 2a_{k+l}) \left( U (1 - s_d) s_d^{-1} \right)^{j-k}}{(j-k)!} \int_{\mathbb{R}^+} e^{-(1-2a_{k+l})t} p_i(t)^{n_{ii}} dt,$$

Finally, an integration by substitution entails the third equality in (S22), which is a definite integral easily computable with any suitable software.  $\square$

## Beneficial lineages in an evolving population

### Incorporating background selection

In the absence of beneficial mutations, an idealized (Wright-Fisher) population will over time reach a state in which individuals within the population carry a Poisson-distributed number of deleterious mutations with mean  $\theta = U/\bar{s}$ , where  $U$  is deleterious mutation rate and  $\bar{s}$  is geometric mean effect of deleterious mutations (Orr 2000; Haigh 1978b; Johnson 1999b). This now classical result breaks down, however, at sufficiently high mutation rate,  $U$ . Gessler (1995) points out that Poisson classes with frequency less than  $1/N$ , where  $N$  is population size, (i.e., fitness classes in the tails of the Poisson distribution) will not exist: their frequency will be zero. While this observation is of little consequence where the right tail of the distribution is concerned, it is nevertheless quite consequential to the evolution of the population where the left (high fitness) tail of the distribution is concerned.

In fact the fittest  $k$  classes may disappear deterministically (left tail may be truncated) if their expected frequency is less than  $1/N$ . Intuitively, this number may be computed as:

$$k = \min \left\{ x | N e^{-\theta} \theta^x / x! \geq 1, x \in \{0, 1, 2, \dots, n\} \right\}$$

As the  $k$  fittest classes disappear due to mutation pressure, the evolving distribution of numbers of deleterious mutations tends toward a shifted Poisson:

$$P(X = x) = \frac{\lambda^{x-k}}{(x-k)!} e^{-\lambda}, \quad x \in \{k, k+1, k+2, \dots\}$$

where  $\lambda = \theta - k$ . Notably, this distribution has mean  $\theta$  and variance  $\theta - k$ , meaning that it is “under-dispersed” relative to the Poisson distribution:  $V(X) < E(X)$ , where  $X$  is a random variable denoting number of deleterious mutations.

Gessler’s key insight was the perhaps counterintuitive subtlety that there is no guarantee, from the above derivations, that the zero class of the shifted Poisson distribution exists, i.e., that  $e^{-\lambda} \geq 1/N$ . The condition used to find  $k$  is that it be the smallest class in the Poisson distribution with parameter  $\theta$  whose expected frequency exceeds  $1/N$ . This condition, however, does not guarantee that this same class exists (has frequency greater than  $1/N$ ) in the new shifted-Poisson distribution with parameter  $\lambda = \theta - k$ . Put differently, the fittest class in the new shifted distribution is not necessarily the class that carries  $k$  deleterious mutations.

Instead, to find the fittest, or ‘best’, class in the new distribution, we must find the smallest class whose expected frequency is greater than  $1/N$  in a new truncated Poisson:

$$b = \min \{ x | N f_k(x) \geq 1, x \in \{k, k+1, k+2, \dots, \theta\} \}$$

where  $f_k(x)$  is a shifted Poisson truncated on the left at  $k$ :

$$\begin{aligned} f_k(x) &= P(X = x) / P(X \geq x) = \frac{\lambda^{x-k}}{(x-k)!} \left( \sum_{i=x-k}^{\infty} \frac{\lambda^i}{i!} \right)^{-1} \\ &= \frac{\lambda^{x-k}}{(x-k)!} e^{-\lambda} \left( 1 - \frac{\Gamma(x-k, \lambda)}{\Gamma(x-k)} \right)^{-1}. \end{aligned}$$

(The final expression is our own simplification of Gessler.)

### Survival probabilities

In direct analogue to (2), we can immediately write down the survival probability of a beneficial mutation arising in a heterogeneous population. If the beneficial mutation in question has selective  $s_b$  and it arises on a genetic background carrying  $j$  deleterious mutations, then the initial growth rate of the beneficial lineage formed by this beneficial mutation will be:

$$w_j = (1 + s_b)(1 - \bar{s}_d)^j.$$

The survival probability of this beneficial mutation is bounded by:

$$p_{svl} = \begin{cases} K \frac{e^U}{w_j} \left( 1 - \frac{e^U}{w_j} \right), & \bar{s}_d \geq \bar{s}_b \\ K \frac{1}{w_j} \left( 1 - \frac{1}{w_j} \right), & \bar{s}_d \rightarrow 0, \end{cases} \quad (\text{S26})$$



If we do not know the selective advantage of the beneficial mutation in question, and we do not know how many deleterious mutations are in the background upon which the beneficial mutation arises, we can employ the distributions of these quantities to compute ensemble survival probabilities of beneficial mutations as:

$$\langle p_{svl} \rangle = \begin{cases} K \mathbb{E} \left[ \frac{e^U}{W_l} \left( 1 - \frac{e^U}{W_l} \right) \right], & \bar{s}_d \geq \bar{s}_b \\ K \mathbb{E} \left[ \frac{1}{W_u} \left( 1 - \frac{1}{W_u} \right) \right], & \bar{s}_d \rightarrow 0, \end{cases} \quad (\text{S27})$$

where  $K = 4$  for binary fission, and  $K \approx 2$  for Poisson-distributed offspring;  $W_l = (1 - \bar{s}_d)^{J_l} (1 + S_b)$  and  $J_l \sim \mathcal{P}_{k,\beta}(j) \forall j \in [\beta, \hat{j}_l]$ ;  $W_u = (1 - \bar{s}_d)^{J_u} (1 + S_b)$  and  $J_u \sim \mathcal{P}_{k,\beta}(j) \forall j \in [\beta, \hat{j}_u]$ ;  $S_b$  is also a random variable and, for each value of  $S_b$ ,  $\hat{j}_l = \max\{j | W_j > e^U\}$  and  $\hat{j}_u = \max\{j | W_j > 1\}$ , and  $W_j$  is defined as  $W_j = (1 + S_b)(1 - \bar{s}_d)^j$ .

**Monte Carlo integration** Monte Carlo integration of (S27) was achieved by first drawing  $S_b$  from an exponential distribution with mean  $\bar{s}_b$ , then computing  $\hat{j}_u$  and  $\hat{j}_l$ , and finally computing a value for  $p_{svl}$ , and repeating. The average of the values computed for  $p_{svl}$  in this way was then taken to be the ensemble average  $\langle p_{svl} \rangle$ .

**Distribution of beneficial mutational effects** We will rely on statistical arguments put forth by Gillespie (1991), after Smid and Stam (1975), that selection coefficients of beneficial mutations are exponentially distributed. While this assumption, and the statistical arguments upon which it is based, have been questioned and in some cases rejected based on different data analyses (Rokyta et al. 2008), we employ it here because of its far-reaching basin of attraction, for its simplicity, and because it is likely to hold approximately in the short term even if it does not ultimately hold exactly in the long term.

**Refining the Haldane survival probability** Haldane derived what he called the “fixation probability” of a beneficial mutation of selective advantage  $s_b$  to be approximately  $2s_b$ . In our terminology, this quantity is more accurately called the “survival probability”, as it does not take into account finite population size. Because we will be comparing our results to very large numbers of simulations, we would like to refine Haldane’s result a bit.

*General expression.* Haldane’s derivation starts with a result from single-type branching process theory, namely, that extinction probability of a Galton-Watson process is the smallest positive  $x$  that satisfies  $x = h(x)$ , where  $h(x)$  denotes the probability generating function associated with the distribution of numbers of offspring produced by individuals in the population. To compute survival probability – the complement of extinction probability – we simply replace  $x$  with  $y = 1 - x$ . Survival probability is thus the largest  $y \in [0, 1]$  satisfying  $1 - y = h(1 - y)$ . Expansion yields

$$\begin{aligned} 1 - y &\approx h(1) - y h'(1) + \frac{1}{2} y^2 h''(1) \\ &= 1 - y m_h + \frac{1}{2} y^2 (\sigma_h^2 + m_h^2 - m_h), \end{aligned}$$

where  $m_h = h'(1)$  is the mean number of offspring, and  $\sigma_h^2 = h''(1) - [h'(1)]^2 + h'(1)$  is the variance in number of offspring. Rearranging thus yields a general approximation for survival probability in the single-type case:

$$p_{svl} \approx \frac{2(m_h - 1)}{\sigma_h^2 + m_h^2 - m_h} \quad (\text{S28})$$

In this single-type case the mean number of offspring does not change over time and is given by  $m_h = 1 + s_b$ , resulting in:

$$p_{svl} \approx \frac{2s_b}{\sigma_h^2 + (1 + s_b)^2 - (1 + s_b)}.$$

If it is the case that the quantity  $\sigma_h^2 + (1 + s_b)^2 - (1 + s_b)$  is close to one, then the survival probability may be approximated as  $p_{svl} \approx 2s_b$ , recovering Haldane’s classical result.

*Poisson-distributed offspring.* When numbers of offspring have a Poisson distribution, then we have  $\sigma_h^2 = m_h$ , and (S28) becomes:

$$p_{svl} \approx \frac{2(m_h - 1)}{m_h^2} = \frac{2s_b}{(1 + s_b)^2}. \quad (\text{S29})$$

*Bacteria (binary fission).* After reproduction and subsequent population culling, or sampling, the number of offspring a bacterium produces has a Binomial distribution with parameters  $n = 2$  and  $p = m_h/2$ , i.e.,  $h(x) = [(1 - m_h/2) + (m_h/2)x]^2$ . For this particular case, no approximation is necessary:  $1 - y = [(1 - m_h/2) + (m_h/2)(1 - y)]^2$  may be solved exactly for  $y$ , yielding:

$$p_{svl} = \frac{4(m_h - 1)}{m_h^2} = \frac{4s_b}{(1 + s_b)^2}. \quad (\text{S30})$$

*Expression used in main text.* Results in the main text are presented in a way that can accommodate either of the foregoing two cases by implementing the following expression for single-type survival probability:

$$p_{svl} = \frac{Ks_b}{(1 + s_b)^2} ,$$

where  $K = 2$  for Poisson-distributed offspring and  $K = 4$  for bacteria (binary fission). When  $s_b$  is replace with average selective advantage,  $\bar{s}_b$ , we introduce a new variable  $\hat{s}_b = \bar{s}_b / (1 + \bar{s}_b)^2$ , so that single-type survival probability may be written as:

$$p_{svl} = K\hat{s}_b .$$

## Approximation

*Strategy and rationale.* Rationale for our approximation strategy is based on the observation that, as  $s_b$  increases, survival probability jumps from zero to something close to Haldane's classical survival probability at  $U = \ln W$ , where  $W$  is the relative fitness of the beneficial lineage; this is especially true for small  $\bar{s}_d$ .

Thus, we approximate survival probabilities as the product of the probabilities that, for a randomly chosen beneficial mutation: 1) the mutation has a selective advantage large enough to overcome the effects of lineage contamination, and 2) the mutation survives. For such beneficial mutations, the probability of 1) is the probability that  $U < \ln W$ ; and, the probability of 2) is our refined Haldane survival probability  $K\hat{s}_b$ .

*Framework.* To facilitate the following developments, we introduce a new function,  $s_b^*(j)$ , defined as follows. If a beneficial mutation arises on a background that contains  $j$  deleterious mutations, it has zero probability of survival unless its selective advantage exceeds a critical value, which is given by the function  $s_b^*(j)$ , defined as:

$$s_b^*(j) = \begin{cases} e^{U(1 - \bar{s}_d)^{-j}} - 1 , & \bar{s}_d \geq \bar{s}_b \\ (1 - \bar{s}_d)^{-j} - 1 , & \bar{s}_d \rightarrow 0 . \end{cases}$$

This function comes from the observation that zero probability of survival in (S26) is achieved when  $1 - e^{U/w_j} \leq 0$  (i.e., when  $e^U \geq w_j$ ) for the case  $\bar{s}_d \geq \bar{s}_b$ , or when  $1 - 1/w_j \leq 0$  (i.e., when  $1 \geq w_j$ ) for the case  $\bar{s}_d \rightarrow 0$ . It is of interest, and will be used later on, to note that the ensemble average critical selective advantage is:

$$\langle s_b^*(j) \rangle = \sum_{j=0}^{\infty} s_b^*(j) \frac{(U/\bar{s}_d)^j}{j!} e^{-U/\bar{s}_d} = \begin{cases} e^{U(\frac{2-\bar{s}_d}{1-\bar{s}_d})} - 1 , & \bar{s}_d \geq \bar{s}_b \\ e^{U(\frac{1}{1-\bar{s}_d})} - 1 , & \bar{s}_d \rightarrow 0 . \end{cases}$$

Based on the rationale outlined above for our approximation strategy, survival probability is thus:

$$p_{svl} \approx K\hat{s}_b e^{-s_b^*(j)/\bar{s}_b} \quad (\text{S31})$$

where  $\hat{s}_b = \bar{s}_b / (1 + \bar{s}_b)^2$ , as derived above. Ensemble survival probability is thus:

$$\langle p_{svl} \rangle \approx \sum_{j=0}^{\infty} K\hat{s}_b e^{-s_b^*(j)/\bar{s}_b} \frac{(U/\bar{s}_d)^j}{j!} e^{-U/\bar{s}_d} \quad (\text{S32})$$

*Lower bounds on ensemble survival probability.* To derive lower bounds on survival probability, we take the logarithm of (S31):

$$\ln p_{svl} \approx \ln(K\hat{s}_b) - s_b^*(j)/\bar{s}_b$$

and write corresponding ensemble log survival probability:

$$\begin{aligned} \langle \ln p_{svl} \rangle &\approx \ln(K\hat{s}_b) - \langle s_b^*(j) \rangle / \bar{s}_b \\ &= \begin{cases} \ln(K\hat{s}_b) - (1/\bar{s}_b)(e^{U(\frac{2-\bar{s}_d}{1-\bar{s}_d})} - 1) , & \bar{s}_d \geq \bar{s}_b \\ \ln(K\hat{s}_b) - (1/\bar{s}_b)(e^{U(\frac{1}{1-\bar{s}_d})} - 1) , & \bar{s}_d \rightarrow 0 . \end{cases} \end{aligned}$$

Jensen's inequality provides:

$$\langle p_{svl} \rangle \geq e^{\langle \ln p_{svl} \rangle}$$

and thus yields exact minimums on both upper and lower bounds for the ensemble survival probability:

$$\langle p_{svl} \rangle \geq \begin{cases} K\hat{s}_b e^{-(1/\bar{s}_b)(e^{U(\frac{2-\bar{s}_d}{1-\bar{s}_d})} - 1)} , & \bar{s}_d \geq \bar{s}_b \\ K\hat{s}_b e^{-(1/\bar{s}_b)(e^{U(\frac{1}{1-\bar{s}_d})} - 1)} , & \bar{s}_d \rightarrow 0 , \end{cases} \quad (\text{S33})$$

This expression is a bound-of-bounds and thus of questionable utility. Comparison with simulations, however, reveals the upper bound (for  $\bar{s}_d \rightarrow 0$ ) to be quite accurate; nevertheless, the lower bound (for  $\bar{s}_d \geq \bar{s}_b$ ) appears to be overly conservative.

*Approximate ensemble survival probability.* We now employ a different approach that avoids the use of Jensen's inequality by making approximations that, at first blush, may appear somewhat crude; nevertheless, comparison with simulations shows this approach to work spectacularly well. We first note that typical values of  $\bar{s}_d$  (even when larger than  $\bar{s}_b$ ) are likely to be small enough to permit the following approximations:

$$s_b^*(j) = \begin{cases} e^U(1 - \bar{s}_d)^{-j} - 1 \approx U + j\bar{s}_d, & \bar{s}_d \geq \bar{s}_b \\ (1 - \bar{s}_d)^{-j} - 1 \approx j\bar{s}_d, & \bar{s}_d \rightarrow 0. \end{cases}$$

(S32) now becomes:

$$\langle p_{svl} \rangle \approx \sum_{j=0}^{\infty} K\hat{s}_b e^{-(c_1 + j\bar{s}_d)/\bar{s}_b} \frac{(U/\bar{s}_d)^j}{j!} e^{-U/\bar{s}_d}, \quad (S34)$$

where  $c_1 = U$  when  $\bar{s}_d \geq \bar{s}_b$ , and  $c_1 = 0$  when  $\bar{s}_d$  is small. This gives rise to approximate survival probabilities:

$$\langle p_{svl} \rangle \approx \begin{cases} K\hat{s}_b e^{-U/\bar{s}_b} e^{-U/\bar{s}_d(1 - e^{-\bar{s}_d/\bar{s}_b})}, & \bar{s}_d \geq \bar{s}_b \\ K\hat{s}_b e^{-U/\bar{s}_b}, & \bar{s}_d \rightarrow 0. \end{cases} \quad (S35)$$

From here, it is apparent that the smallest value of  $\langle p_{svl} \rangle$  is achieved when  $\bar{s}_d = \bar{s}_b$ , so that survival probability is bounded as:

$$K\hat{s}_b e^{-\varepsilon U/\bar{s}_b} \lesssim \langle p_{svl} \rangle \lesssim K\hat{s}_b e^{-U/\bar{s}_b}, \quad (S36)$$

where  $\varepsilon = 2 - e^{-1} \approx 1.63$ . Remarkably, the foregoing bounds on survival probability are independent of  $\bar{s}_d$ .

**Mutation rate that maximizes production of surviving beneficial mutations** We first note that the rate of production of surviving beneficial mutations will be proportional to  $U\langle p_{svl} \rangle$ . This is the only information we need. The genomic mutation rate that maximizes this rate of production is found by setting  $\partial_U U\langle p_{svl} \rangle = 0$  and solving for  $U$ , we find this maximum production rate occurs at mutation rate  $\hat{U}$ , given by:

$$\hat{U} = \begin{cases} [1/\bar{s}_b + (\varepsilon - 1)/\bar{s}_d]^{-1}, & \bar{s}_d \geq \bar{s}_b \\ \bar{s}_b, & \bar{s}_d \rightarrow 0. \end{cases} \quad (S37)$$

The smallest value of  $\hat{U}$  is achieved when  $\bar{s}_d = \bar{s}_b$ , resulting in the bounds:

$$\frac{1}{\varepsilon} \bar{s}_b \leq \hat{U} < \bar{s}_b. \quad (S38)$$

**Clonal interference threshold** As delineated in [Sniegowski and Gerrish \(2010\)](#), the clonal interference regime is entered when a second, alternative beneficial mutation is likely to be produced on the ancestral background before the first, or focal, beneficial mutation becomes fixed. If the focal mutation has selective advantage  $s_b$  then, given that it survives stochastic sampling in the first few generations (genetic drift), the dynamics of its frequency,  $x(t)$ , is described by the differential equation,

$$x'(t) = s_b x(t)(1 - x(t)).$$

Conditioning on surviving drift ([Barton 1995](#)), the initial condition is:

$$x(0) = \frac{1}{Ns_b},$$

yielding solution:

$$x(t) = [e^{-s_b t}(Ns_b - 1) + 1]^{-1}.$$

At this point, we refine our definition of "fixation" of the focal beneficial mutation as being the point at which half of the population carries the mutation. The rationale for this refinement is that beyond this point, when more than half of the population carry the mutation, it becomes more likely that the next beneficial mutation to arise will appear not on the wildtype background but on the background carrying the focal beneficial mutation. Fixation, defined in this way, will occur  $\tau$  generations after the appearance of the focal mutations, where  $x(\tau) = 1/2$ . Thus,

$$\tau = \frac{\ln(Ns_b - 1)}{s_b}.$$

We are interested in knowing whether an alternative beneficial mutation will arise in the interval  $(0, \tau)$  on the wildtype background. Letting  $y(t)$  denote wildtype frequency at time  $t$ , we have that

$$y(t) = 1 - x(t).$$

The expected number of alternative beneficial mutations produced in the wildtype subpopulation is:

$$\mu N \int_0^\tau y(t) dt ,$$

where  $\mu = cU$  is beneficial mutation rate, and:

$$\begin{aligned} \int_0^\tau y(t) dt &= \frac{1}{s_b} [\ln(Ns_b) + \ln(Ns_b - 1) - \ln(2(Ns_b - 1))] \\ &\approx \frac{\ln(Ns_b/2)}{s_b} . \end{aligned}$$

The expected number of alternative beneficial mutations *produced* on the wildtype background is thus:

$$\mu N \frac{\ln(Ns_b/2)}{s_b} .$$

The expected number of alternative beneficial mutations produced, however, is different from the number of *competing* beneficial mutations, because many of those produced will not survive the effects of genetic drift and the accelerated accumulation of deleterious mutations (lineage contamination). Letting random variable  $M$  denote the number of *competing* beneficial mutations, then  $M$  has a Poisson distribution with expectation:

$$\mathbb{E}(M) = \mu N \frac{\ln(Ns_b/2)}{s_b} \langle p_{svl} \rangle .$$

A population is in the clonal interference regime when at least one competing beneficial will arise on average; this criterion is thus met when:

$$\mathbb{E}(M) \geq 1 .$$

The only remaining issue is to determine the value of  $s_b$  to be used in the foregoing expressions. We will use  $\tilde{s}_b$  denote this value. To let  $\tilde{s}_b = \bar{s}_b$  would be erroneous, because beneficial mutations that fix will typically have higher-than-average selective advantages. And to survive the effects of lineage contamination, we have established that the selective advantage must be at least  $e^U - 1$ . In previous work (Rozen *et al.* 2002; Gerrish and Lenski 1998), we find that selective advantages of beneficial mutations that survive drift will be Gamma distributed with shape parameter 2. Fortunately, on the border between clonal interference and periodic selection regimes, we expect only a single competing beneficial mutation, and we therefore do not have to account for the effects of clonal interference on selective advantages of beneficial mutations that fix. In light of the foregoing observations, we have:

$$\tilde{s}_b = \frac{\int_{e^U-1}^\infty \alpha s^2 e^{-\alpha s} ds}{\int_{e^U-1}^\infty \alpha s e^{-\alpha s} ds} = \bar{s}_b + e^U - 1 + \frac{\bar{s}_b^2}{\bar{s}_b + e^U - 1} \approx \bar{s}_b + U + \frac{\bar{s}_b^2}{\bar{s}_b + U} .$$

It is readily verified that, as  $U \rightarrow 0$ ,  $\tilde{s}_b \rightarrow 2\bar{s}_b$ , as expected.

Finally, we note that there will typically be two solutions to the equation  $\mathbb{E}(M) = 1$ , which defines the borders between clonal interference and periodic selection regimes. The smaller of these two solutions defines the border, as low mutation rate increases, from a periodic selection regime into a clonal interference regime. The larger of these two solutions defines the border, as already high mutation rate increase further, from a clonal interference regime back into a periodic selection regime. This second transition is the direct result of lineage contamination and is what we have called the “clonal interference threshold”.

**Fixation threshold** The probability that a randomly chosen beneficial mutation does *not* survive is:

$$\langle p_{ext} \rangle = 1 - \langle p_{svl} \rangle = \begin{cases} 1 - K\hat{s}_b e^{-(U/\bar{s}_b)} e^{-(U/\bar{s}_d)(1-e^{-\bar{s}_d/\bar{s}_b})} , & \bar{s}_d \geq \bar{s}_b \\ 1 - K\hat{s}_b e^{-(U/\bar{s}_b)} , & \bar{s}_d \rightarrow 0 . \end{cases}$$

where  $\kappa \in (1, \varepsilon]$ . The probability that  $n$  beneficial mutations do not survive is thus:

$$\langle p_{ext} \rangle^n = [1 - \langle p_{svl} \rangle]^n ,$$

which may be approximated by:

$$\langle p_{ext} \rangle^n \approx \begin{cases} e^{-nK\hat{s}_b e^{-(U/\bar{s}_b)} e^{-(U/\bar{s}_d)(1-e^{-\bar{s}_d/\bar{s}_b})}} , & \bar{s}_d \geq \bar{s}_b \\ e^{-nK\hat{s}_b e^{-(U/\bar{s}_b)}} , & \bar{s}_d \rightarrow 0 . \end{cases}$$

The fixation threshold is obtained by finding the maximum  $U$  that satisfies:

$$\langle p_{svl} \rangle^n = p_c$$

where  $p_c$  is the probability with which one wishes to ensure that no fixations occur. The “fixation threshold” is exceeded when no beneficial mutation produced by a population has a selective advantage strong enough to survive the effects of lineage contamination.

Concretely, in a given interval of time  $\tau$ , we suppose a population produces a total of  $n$  beneficial mutations; then, the fixation threshold is defined as the mutation rate that ensures extinction of even the fittest of these mutations. This critical mutation rate, which defines the fixation threshold and which we will denote by  $U_f$ , thus ensures the extinction of all  $n$  beneficial mutations produced with specified probability  $p_c$ ; it is given by:

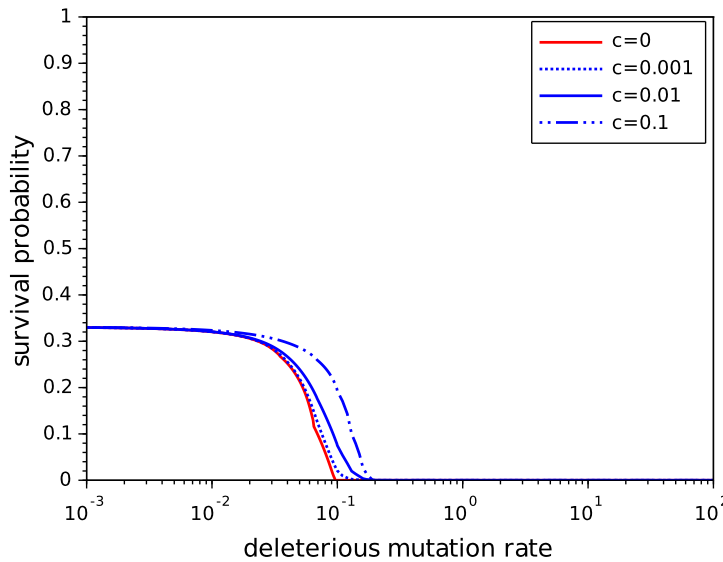
$$U_f = \begin{cases} \max\{U \mid U = -[1/\bar{s}_b + (\varepsilon - 1)/\bar{s}_d]^{-1} \ln\left(-\frac{\ln p_c}{n(U)}\right)\}, & \bar{s}_d \geq \bar{s}_b \\ \max\{U \mid U = -\bar{s}_b \ln\left(-\frac{\ln p_c}{n(U)}\right)\}, & \bar{s}_d \rightarrow 0 \end{cases} \quad (\text{S39})$$

More compactly, this may be rewritten as:

$$U_f = \max\{U \mid U = -\hat{U} \ln\left(-\frac{\ln p_c}{n(U)}\right)\},$$

where  $\hat{U}$  is defined and derived above, and  $n(U) = NUcK\hat{s}_b\tau$ , and  $\tau$  denotes the relevant time period; for example, to compute the mutation rate at which, over a time period of 5000 generations, all fixations will be suppressed with probability 95%, we set  $\tau = 5000$  and  $p_c = 0.95$ .

**Multiple beneficial mutations** Our multitype branching process model assumes that beneficial mutations occur infrequently enough that acquiring a second beneficial mutation in linkage with the focal beneficial mutation is improbable in the time required for the focal mutation to either survive or go extinct. In reality, it might be the case that multiple beneficial mutations arise on the same background and sweep to fixation, collectively overcoming the lineage contamination effect. Figure S7 plots survival probabilities computed from an extension of our multitype branching process model that allows for the occurrence of a second beneficial mutation within the lineage formed by the first, or focal, beneficial mutation.



**Figure S7** Survival probability of a single beneficial lineage, potentially acquiring a second beneficial mutation with probability  $1 - e^{-cU}$ . Here  $s_b = 0.1$  and  $s_d = 0.03$ .

**Simulations** Our simulations describe the evolution of a population whose size at time  $t$  is  $N_t$ . At time  $t$ , the  $i^{th}$  individual in the population has fitness  $W_{t,i}$ , so simulations model the evolution of the vector  $\mathbf{W}_t = (W_{t,1}, W_{t,2}, \dots, W_{t,N_t})$ . To compute vector  $\mathbf{W}_{t+1}$ , the first step is replication: each individual is granted a number of offspring,  $Y_{t,i}$ , proportional to that individual's fitness:  $Y_{t,i}$  is a Poisson random variable with mean  $W_{t,i}/\bar{W}_t$ , where  $\bar{W}_t = \frac{1}{N_t} \sum_{j=1}^{N_t} W_{t,j}$ . If individual  $j$  in the population at time  $t+1$  is an offspring of individual  $i$  in the parent population (at time  $t$ ), then its fitness is computed by applying  $W_{t+1,j} = W_{t,i} \prod_{k=1}^M (1 - S_d(k))$ , where  $i = 1, 2, \dots, N_t$ , each  $S_d(k)$  is an exponentially-distributed random variable with mean  $s_d$ , and  $M$  is a Poisson-distributed random variable with mean  $U$ .

*Single beneficial lineage in an otherwise homogeneous population.* To study the effects of lineage contamination in isolation, we simulate the growth of a beneficial lineage in an initially homogeneous population by imposing the initial conditions:  $W_{0,1} = (1 + s_b)$  and  $W_{0,i} = 1$  for  $i = 2, 3, \dots, N_0$ . These simulations closely approximate the process modeled by our branching process approach described above, but they allow us to study the effects of relaxing certain assumptions.



*Single beneficial lineage in a heterogeneous population initially at mutation-selection balance.* In other simulations, we analyze the growth of a single beneficial lineage occurring at random (on any background) in a population initially at mutation-selection balance. Here, initial conditions are:  $W_{0,i} = \prod_{j=1}^X (1 - S_d(j))$ , where  $X$  is a Poisson-distributed random variate with mean  $U/s_d$ , and the  $S_d(j)$  are exponential random variates with mean  $s_d$ , and  $W_{0,K} = W_{0,K}(1 + s_b)$  where  $K$  is an integer randomly chosen in the interval  $[1, N_0]$ .

*Multiple beneficial mutations in a heterogeneous population.* These simulations assume a deleterious genome of infinite length and a binary (bit-string) beneficial genome of length 96 (three binary “genes” each of length 32). Each of the 96 positions on the beneficial genome was assigned a fixed fitness effect of mutation at that position. Initially, this set of selection coefficients was created by drawing at random from an exponential distribution with mean 0.03. For purposes of consistent comparison, we used the same set of available beneficial mutations for all such simulations.

*Isolating lineage contamination and background selection processes.* In simulations, we isolated the effects of the lineage contamination and background selection processes by allowing one of these processes to operate while suppressing the other. To examine the effects of lineage contamination alone, we suppressed background selection by not allowing deleterious mutations to arise during the replication of an individual with zero beneficial mutations. To examine the effects of background selection alone, we suppressed lineage contamination by not allowing deleterious mutations to arise during the early growth of beneficial lineages. And of course, to examine the combined effects of lineage contamination and background selection, we allowed equally indiscriminate mutation (we suppressed neither process).