Supporting Information

On the accumulation of deleterious mutations during range expansions

S. Peischl[1,2,3], I. Dupanloup[1,3], M. Kirkpatrick[2], and L. Excoffier[1,3]

[1] Institute of Ecology and Evolution, University of Berne, 3012 Berne, Switzerland
[2] Section of Integrative Biology, University of Texas, Austin Texas 78712, USA
[3] Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

Analytical approximations under a simpler one-dimensional expansion model

Even though our model is relatively simple, it does not lead to the derivation of an analytical expression for the probability that a mutation becomes fixed at the edge of the expansion. A standard way to model the dynamics of populations spreading into new territories, or, analogously, the spread of beneficial mutations through spatially extended populations, is by a continuous space approximation (Fisher, 1937). The resulting partial differential equation has received a lot of attention in the mathematics and physics literature and more recently also in the context of expanding populations (e.g. Hallatschek, Nelson, 2008). Results obtained for this class of models often require knowledge about the exact speed and shape of the expanding wave front, which is usually not known (but see Hallatschek, 2011; Mueller *et al.*, 2011). This makes the derivation of simple closed form approximations challenging and simulations are required to evaluate formulae for quantities of interest such as effective populations sizes or fixation probabilities (Barton *et al.*, 2013; Hallatschek, Nelson, 2008).

Here we pursue a different approach, exploiting the fact that the profile of the wave front will be very steep if $m < r$. Then, individuals that colonize new territories will only be recruited from a few demes at the edge of the range, with most individuals originating from the deme at the very edge of the expansion. We therefore model the expansion process by separating migration from population

growth such that migration only occurs once carrying capacity is reached. .First, neighboring demes exchange migrants such that $Km/2$ individuals move from deme $d_f(t)$ to deme $d_f(t)+1$, where $d_f(t)$ denotes the deme at the wave front in generation t. Then all demes reproduce and the population in the newly colonized deme grows exponentially until it reaches carrying capacity. We assume that selection acts through differential viability only during this growth phase.

In this model, the deme at the edge of the range evolves independently from demes in the range core, since migration rate *m* is used here only to define the number of founders in the newly colonized deme. We can thus study the dynamics of the wave front by simply tracking the dynamics of allele frequencies in deme d_f. The resulting model is equivalent to a model of repeated bottlenecks in a single population (Wahl, Gerrish, 2001), and similar to a recent modeling of a range expansion as a series of founder effects (Slatkin, Excoffier, 2012).

Probability of fixation

We use a diffusion approximation to calculate the probability of fixation of a new mutation (Kimura, 1964). Note that fixation in our model means that the mutation becomes fixed at the front of the expansion wave, i.e., we do not consider the fate of mutations that fall back into the wake of the wave.

First, we consider a mutation that is introduced when demes are at carrying capacity and before migrants colonize a new deme. We then generalize our results to mutations that occur at an arbitrary stage of the colonization process. The selective advantage of an individual carrying a single copy of the mutation is denoted s. Let Δx denote the change in mutant allele frequency x during the colonization of a new deme and let $E[\Delta x]$ and $V[\Delta x]$ denote its mean and variance, respectively. The probability of fixation of the mutant allele is then given by

$$p = \frac{\int_0^{x_0} g(x)\ dx}{\int_0^1 g(x)\ dx},$$

where x_0 is the initial frequency of the mutation and $g(x) = \exp\left(-\int \frac{2E[\Delta x]}{V[\Delta x]} dx \right)$.

During exponential growth, selection deterministically changes the mutant frequency to

$$x' = \frac{x \exp[sT]}{x \exp[sT] + (1-x)},$$

where $T \approx \log\left(2\ m^{-1}\right)/r$ is the length of the growth phase, r denotes the growth rate of the population, and *m* is the migration rate (i.e., $m/2$ is the fraction of a deme`s population that colonizes a new deme). Binomial sampling of individuals during migration does not change the expected frequency of the mutation but it increases sampling variance around *x*'. Hence, the expected change in allele frequency is simply

$$\mathrm{E}[\Delta x] = x - x'$$

and the variance of Δx is

$$V\left[\Delta x\right] = \frac{1}{2K\frac{m}{2}}\ \ x'\ \ (1-x'),$$

where *Km* is the number of genes founding the new deme on the wave front. Under weak selection (i.e, $|s| \ll 1$), we find that the fixation probability is

$$p \approx \frac{\exp(-4s_e K x_0) - 1}{\exp(-4s_e K) - 1},$$

where $s_e = s\frac{m}{2}\frac{\log(2/m)}{r}$ can be considered as an effective selection coefficient for a mutation occurring on the wave front. If the mutation appears when the deme is at carrying capacity, $x_0 = \frac{1}{2K}$ and we find that the fixation probability is

$$p \approx \frac{\exp(-2s_e) - 1}{\exp(-4s_e K) - 1}. \qquad \text{[S1]}$$

Note that Eq. [2] is equivalent to the probability of fixation of a mutation with effect s_e in a single panmictic population of constant size K.

(Kimura, 1964). Hence, if $|Ks_e| \ll 1$, random genetic drift will dominate selection at the range margin. Comparison of Eq. [2] with individual based simulations shows that our approximation is very accurate for a broad range of parameters (Figure S7).

We next consider a mutation with effect s that occurs t. generations after the initial colonization of a new deme and denote the probability of fixation of this mutation by $p(s,t)$. At the edge of the range, the number of individuals in a deme t generations after it has been colonized is $N(t) = (m/2)K\exp(rt)$. If the mutation appears in generation $t < T$, its expected frequency in generation T (i.e., when the deme at the edge is at carrying capacity, and before colonization of a new deme) will be

$$x_0(t) = \frac{\frac{1}{2N(t)}\exp[s(T-t)]}{\frac{1}{2N(t)}\exp[s(T-t)] + (1-\frac{1}{2N(t)})},$$

and using Eq. [2] we immediately get the probability of fixation

$$p(s,t) \approx \frac{\exp(-2s_e 2Kx_0(t)) - 1}{\exp(-4s_e K) - 1}.$$

Change in mean fitness at the wave front

We next calculate an approximation for the expected relative change in mean fitness per generation. We assume that the fitness on the wave front changes upon the successive fixation of new mutations, and that new mutations are either lost or fixed at any locus before the next mutation arises. In addition, we assume that we can treat loci independently (e.g., that recombination between loci is high).

We first consider the establishment of new mutations at a single locus. Let u denote the genome wide mutation rate and n the number of loci. Because mutations occur uniformly across the

genome, the mutation rate per locus is u/n. If this mutation rate is sufficiently small, the probability that a new mutation occurs in generation t is approximately $2N(t)un$. Let $\varphi(s)$ denote the probability that a mutation has effect s. Then the joint probability that a mutation with effect s occurs in generation t and then goes to fixation is

$$P(s,t) = 2N(t)\frac{u}{n}\varphi(s)p(s,t),$$

where $p(s,t)$ is again the probability of fixation of a mutation that occurs in generation t. A mutation that becomes fixed changes mean fitness by a factor $(1+s)(1+s) = 1+2s+s^2$. Because we consider only contributions to mean fitness by mutations that become fixed (i.e., we ignore the contributions of transient polymorphism), the expected effect on fitness of a mutation that occurs in generation t is

$$\mu_s(t) = \sum_{\sigma}(2\sigma+\sigma^2)P(\sigma,t),$$

where σ runs over all mutation effect sizes. If the distribution of fitness effects is continuous rather than discrete, the sum needs to be replaced by an integral, and $\varphi(s)$ then denotes the probability density of mutations with effect s. We then calculate the average value of $\mu_s(t)$ over the time *T* it takes for a deme to reach carrying capacity (and during which selection acts) as

$$\bar{\mu}_s = \frac{1}{T}\int_0^T \mu_s(\tau)d\tau.$$

We measure the change in mean fitness by the logarithm of the relative change in mean fitness in consecutive generations:

$$\omega = \log\left(\frac{\bar{w}_f(t+1)}{\bar{w}_f(t)}\right),$$

where the subscript *f* indicates that mean fitness is measured at the wave front.

Because fitness is multiplicative and loci evolve independently, the expected relative change in mean fitness per generation for n loci is simply $(1+\bar{\mu}_s)^n$ and

$$\omega = \log\left[(1+\bar{\mu}_s)^n\right] = n\log(1+\bar{\mu}_s) \approx n\bar{\mu}_s. \qquad \text{[S2]}$$

Eq. [3] is a key quantity to study the dynamics of expansion load: if $w<0$, load will increase over time at a rate proportional to the absolute value of w. Furthermore, the expected mean fitness at the wave front evolves according to

$$E[\bar{w}_f(t+1)\,|\,\bar{w}_f(t)] = \bar{w}_f(t)(1+\bar{\mu}_s)^n. \qquad \text{[S3]}$$

Here, $E[\bar{w}_f(t+1)\,|\,\bar{w}_f(t)]$ denotes the expected mean fitness at generation $t+1$ conditioned on the mean fitness at generation t.

To proceed further, we need to make assumptions about the distribution of fitness effects (DFE). Unless otherwise stated (see Figure S1), we henceforth assume that mutations have effects $\pm s$ and denote the fraction of mutations that are deleterious (i.e., the ones with effect $-s$) by φ_d. Numerical investigations reveal that $\bar{\mu}_s$ is well approximated by $\mu_s(T)$ if selection is weak. Setting $\bar{\mu}_s$ = $\mu_s(T)$ in Eq. [3], one can show that mean fitness on the wave front will decrease if φ_d is larger than a critical threshold value

$$\varphi_{d,c} = \frac{1}{1+\exp(-4Ks_e)} \approx \frac{1}{2} + Ks_e, \qquad \text{[S4]}$$

where s_e is again the effective selection coefficient in our model. Figure S15 shows $\varphi_{d,c}$ as a function of $2Ks_e$.

Supporting Figures

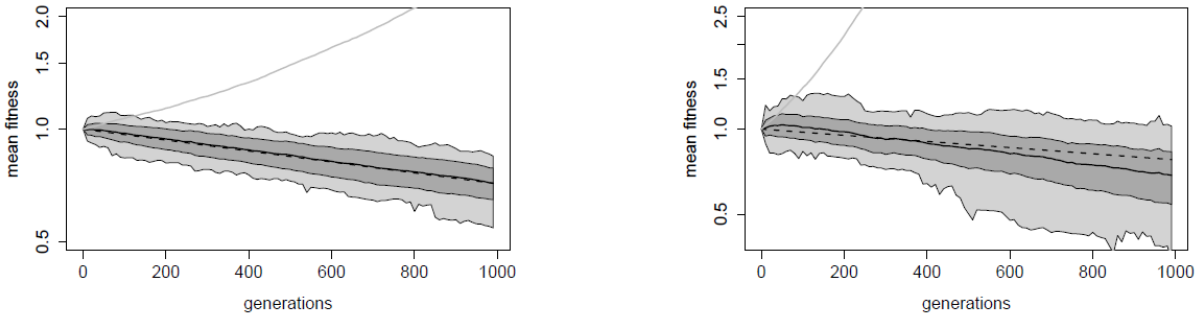

Figure S1. Changes in mean fitness on the wave front of an expanding population for different distribution of fitness effects. In each graph, the solid black line shows the simulated mean fitness at the wave front, dark gray area indicates two standard deviations of mean fitness, and light gray area the total range of observed simulated values. The solid gray line shows the average mean fitness in the core of the population, also obtained by simulations. Averages are taken over 50 simulations. Left: with probability 0.9 a mutation's effect is drawn from an exponential distribution with mean $s = -0.005$, and with probability 0.1 it is drawn from an exponential distribution with mean $s = 0.005$. Right: discrete DFE where mutations with effect $s = -0.05$, -0.005, 0.005, and 0.05 occur with probability $\varphi(s) = 0.09,\ 0.81,\ 0.081$, and 0.009, respectively. Results are for individuals with $n = 20$ freely recombining regions. Other parameter values are $K = 100$, $r = \log(2)$, $m = 0.05$, and $u = 0.05$.

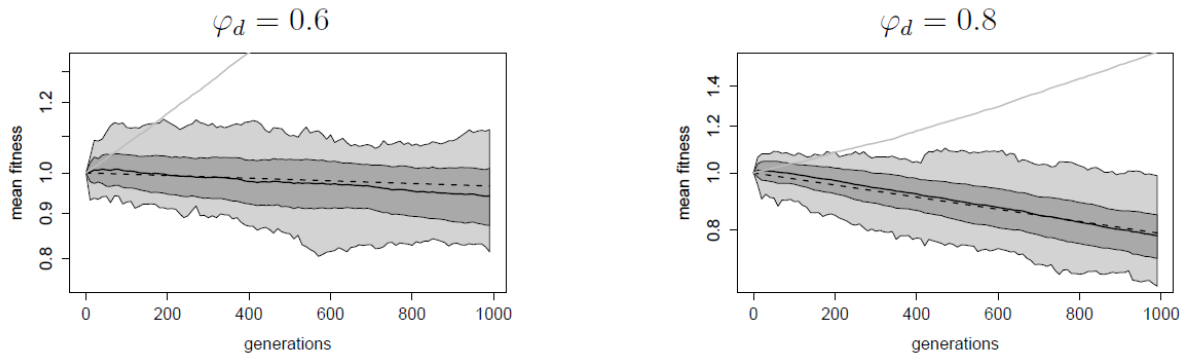

Figure S2. Changes in mean fitness on the wave front of an expanding population for different ratios of deleterious and beneficial mutations. The color code is as in Figure S1. Results are for individuals with $n = 20$ freely recombining regions. Other parameter values are $s = \pm 0.005$, $K = 100$, $r = \log(2)$, $m = 0.05$, and $u = 0.05$.

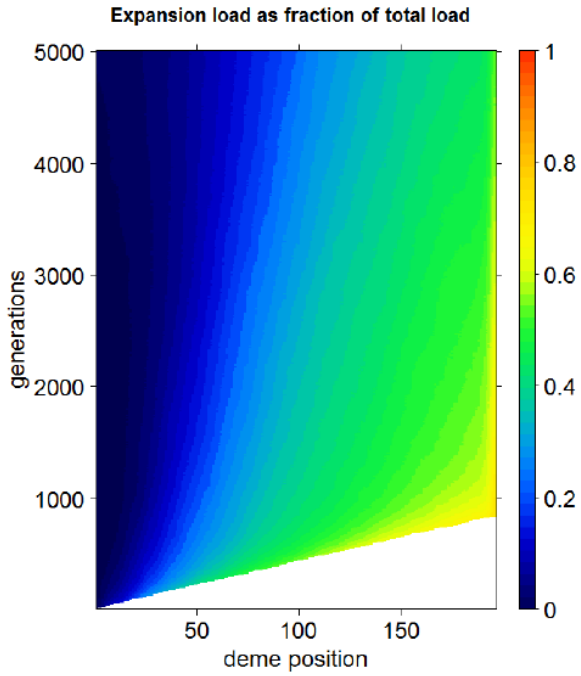

Figure S3. Fraction of total load that originated from the wave front (expansion load) for a finite range. Parameter values are as in Figure 1. Average is taken over 50 simulations.

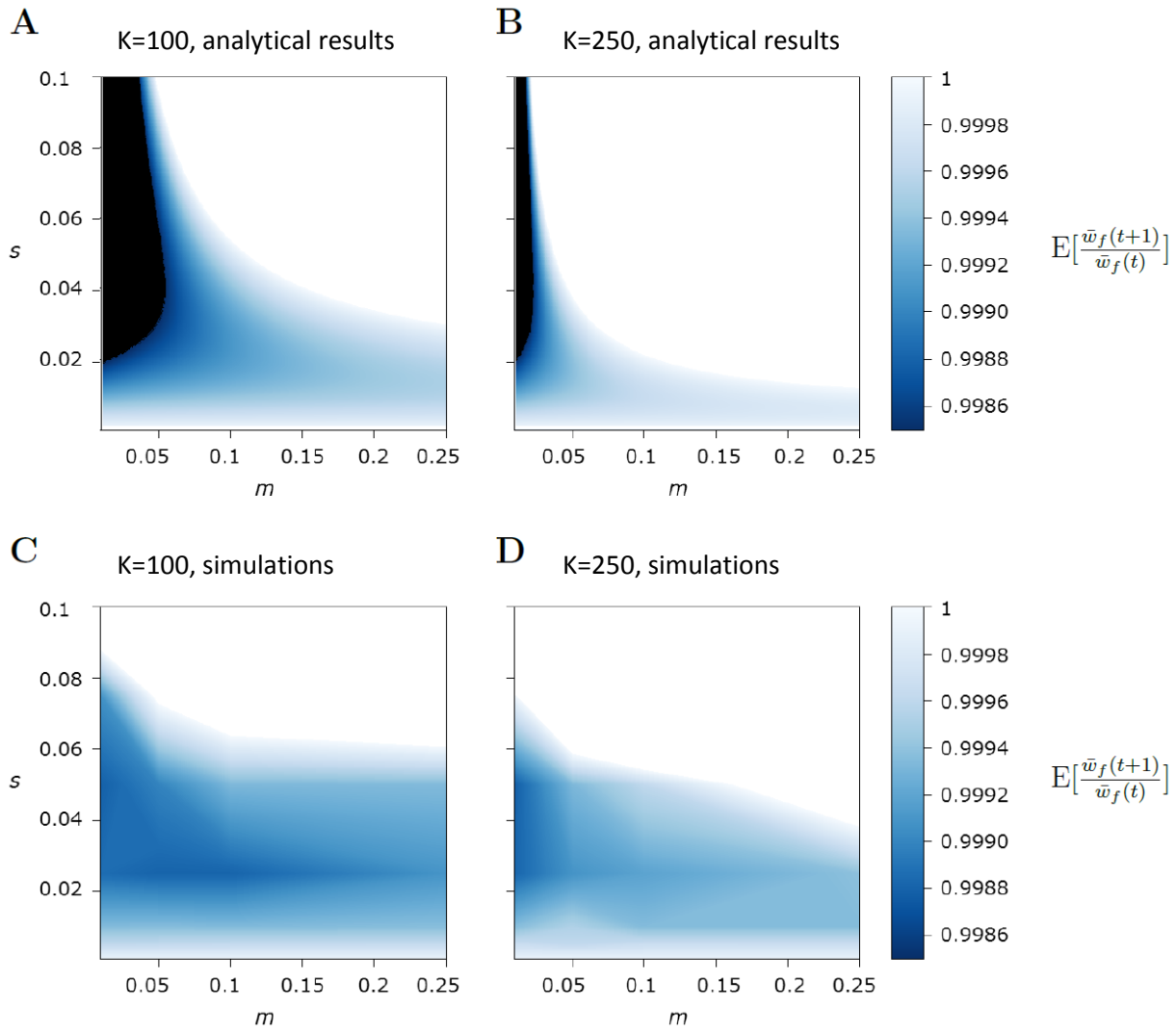

Figure S4. Change in mean fitness on the wave front. The contour plot shows the expected change in mean fitness at the wave front, $E[\bar{w}_f(t+1)/\bar{w}_f(t)]$, as a function of m and s. Mutations have effect $\pm s$. Panels A and B are obtained from the analytical approximation [4], and panels C and D are obtained from simulation of the more complex expansion model. In the white area, $E[\bar{w}_f(t+1)/\bar{w}_f(t)]>1$, and the black area indicates values smaller than the minimum of the color key. Remaining parameter values are $\varphi_d = 0.9, r = \log(2),$ and $u = 0.05.$

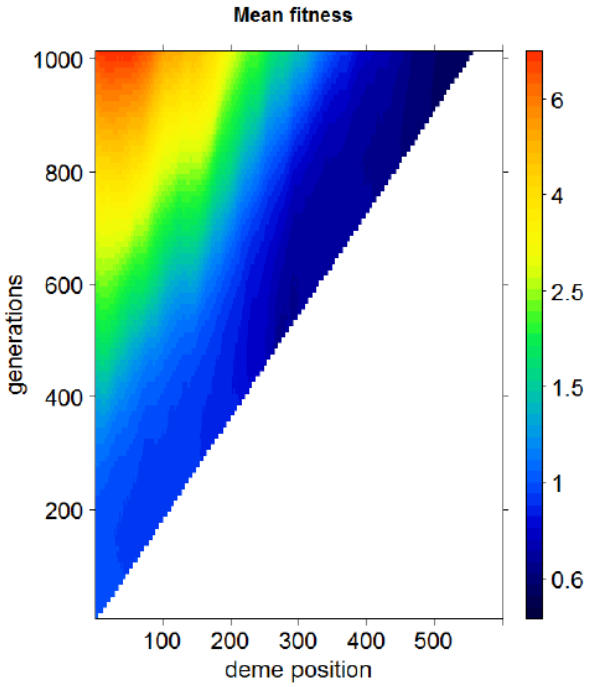

Figure S5. Evolution of population mean fitness with large local carrying capacities. Results are for individuals with $n = 20$ freely recombining regions. Parameter values are $s = \pm 0.01$, $K = 1000$, $r = \log(2)$, $m = 0.25$, $u = 0.05$, and $\varphi_d = 0.9$.

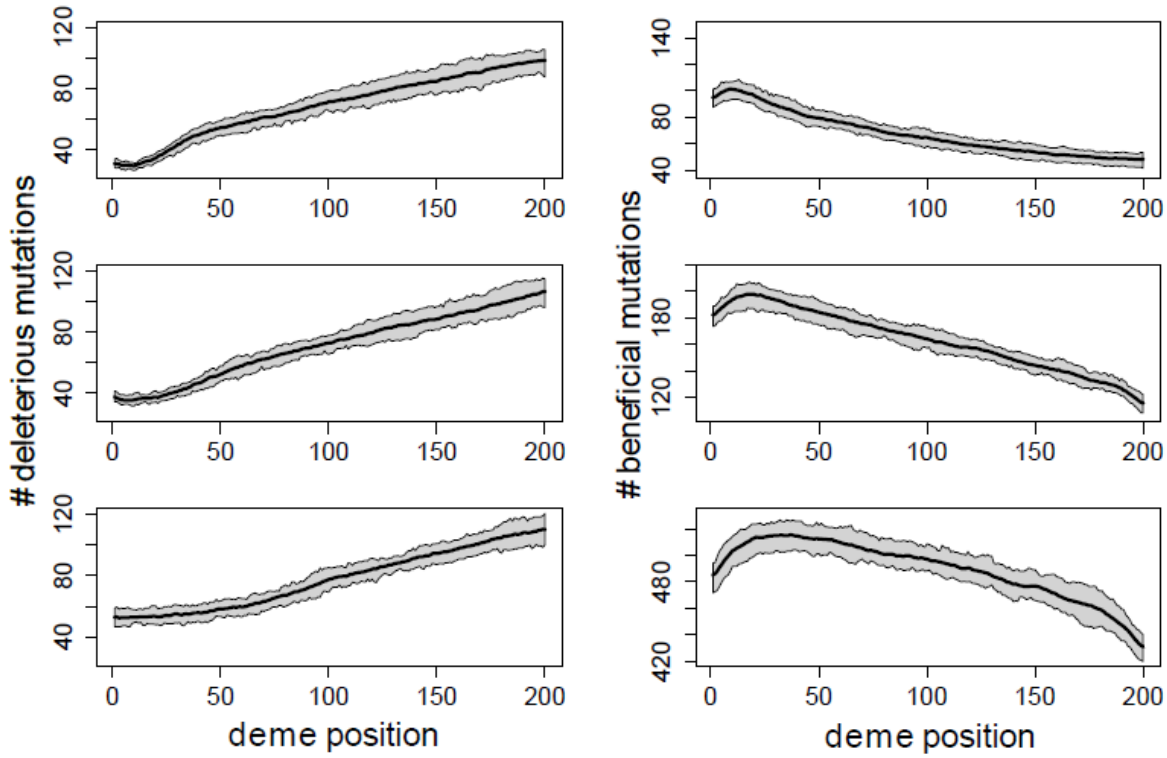

Figure S6. Spatial distribution of mutations under after a range expansion in a linear habitat restricted to 200 demes. The left and right panels show the number of deleterious and beneficial mutations in the 200 demes of the species range, respectively. Top: 1000 generations after the onset of the expansion, Middle: 2000 generations after the onset of the expansion. Bottom: 5000 generations after the onset of the expansion. Solid black line shows mean number of mutations and grey areas indicate the lower and upper quartiles. Average is taken over 50 simulations. Parameter values are as in Figure 1.

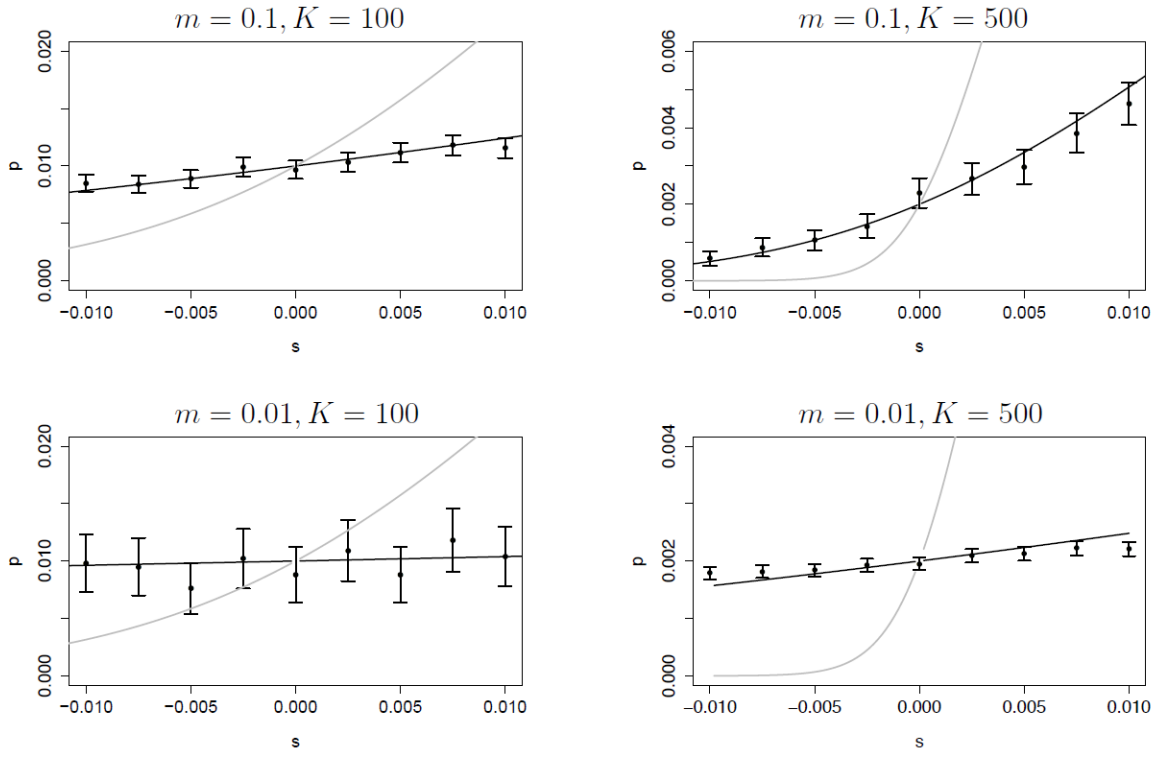

Figure S7. Probability of fixation of mutations at the wave front. Solid black line shows the analytical prediction [S1], and the whisker plots show results from simulations of the more complicated expansion model described in the main text. For comparison, the solid gray line shows the probability of fixation of a mutation in a single panmictic population. In the simulations the five leftmost demes were at carrying capacity and all other demes were empty before the onset of the expansion. A single copy of the mutation was present in the deme at the edge of the expansion, i.e., $x_0 = 1/2K$. In all cases $r = \log(2)$.

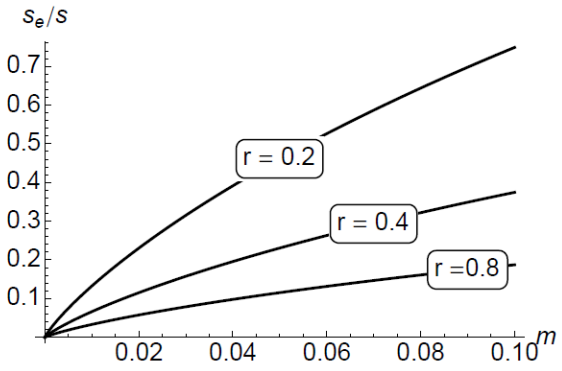

Figure S8. Ratio of the effective selection coefficent at the wave front, s_e, and the actual selection coefficient, s.

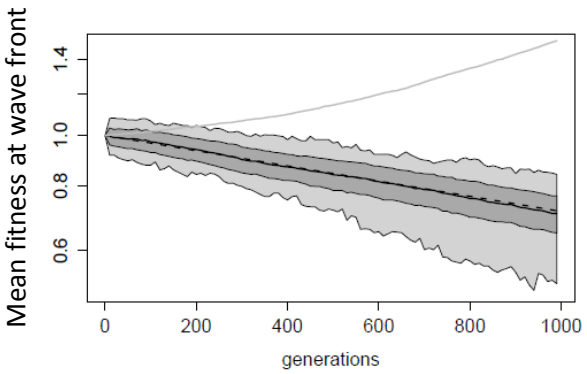

Figure S9. Changes in mean fitness on the wave front of an expanding population with long distance dispersal. Dispersal distances are drawn from a shifted exponential distribution such that the mean dispersal distance is 5 demes. The color code is as in Figure S1. Results are for individuals with n = 20 freely recombining regions. Other parameters are $s = \pm 0.005,$, $K = 100,$ $r = \log(2),$ $m = 0.05,$ $\varphi_d = 0.9$ and $u = 0.05$.

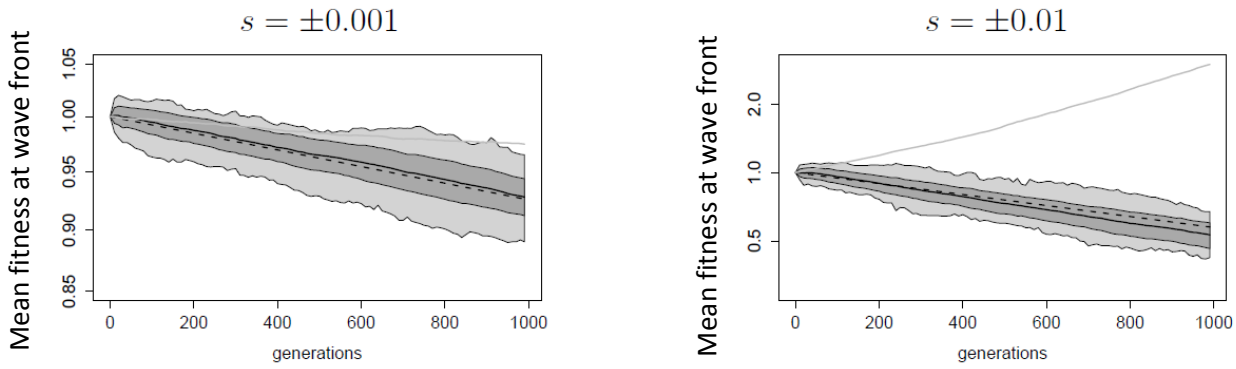

Figure S10. Changes in mean fitness on the wave front of an expanding population. The color code is as in Figure S1. In each graph, the solid black line shows the simulated mean fitness at the wave front, dark gray area indicates two standard deviations of mean fitness, and light gray area the total range of observed simulated values. The solid gray line shows the average mean fitness in the core of the population, also obtained by simulations. Averages are taken over 50 simulations. The dashed line shows the analytical approximation [4]. Results are for individuals with $n = 20$ freely recombining regions. Other parameters are $m = 0.05, K = 100, r = \log(2), \varphi_d = 0.9,$ and $u = 0.05$.

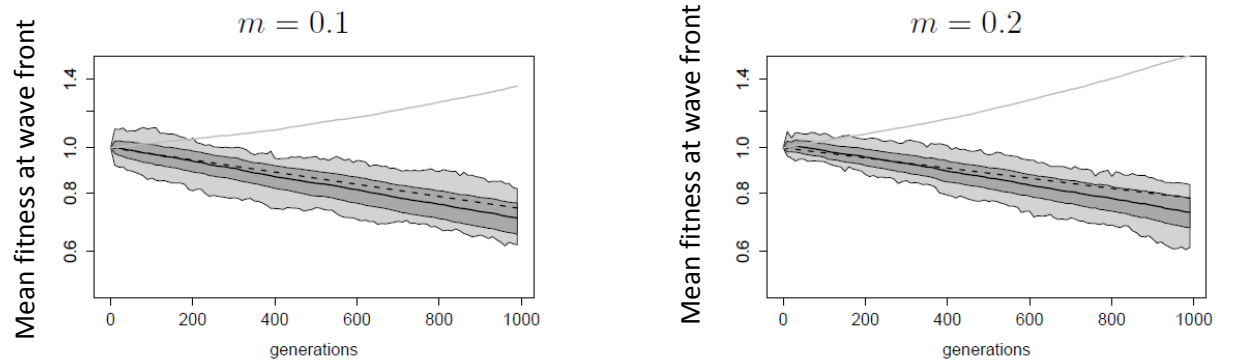

Figure S11. Changes in mean fitness on the wave front of an expanding population for different migration rates. The color code is as in Figure S1. Results are for individuals with $n = 20$ freely recombining regions. Other parameters are $s = \pm 0.005, K = 100, r = \log(2), \varphi_d = 0.9,$ and $u = 0.05$.

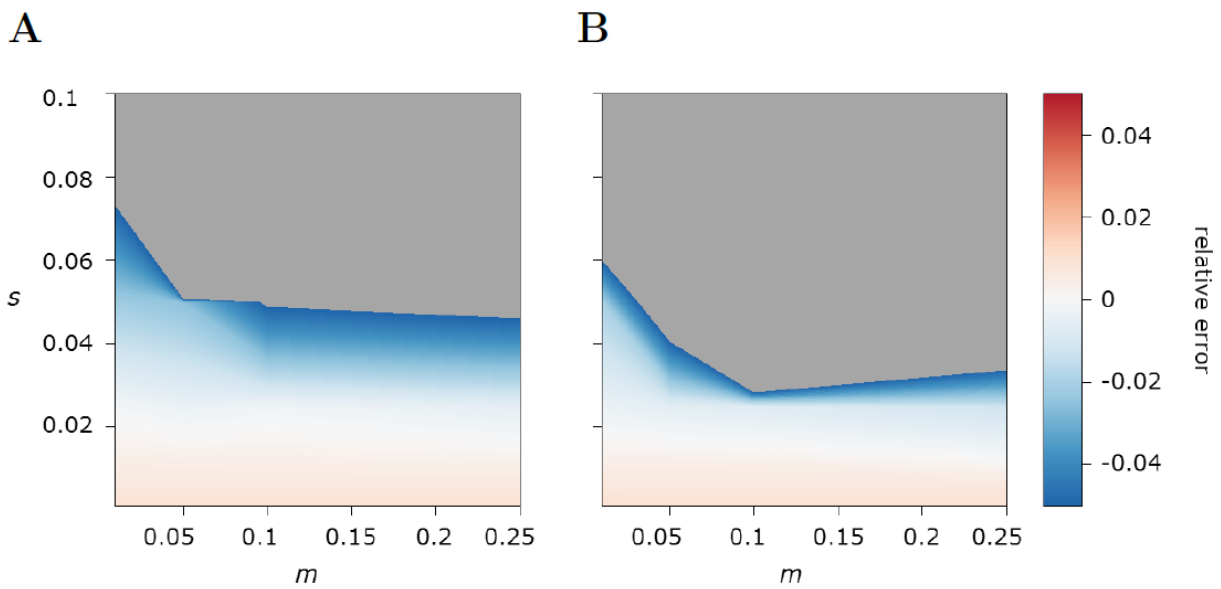

Figure S12. Relative error of the analytical approximation [4] as a function of m and s for $K = 100$ (A) and $K = 250$ (B). Positive values (red) mean that the analytical approximation overestimates the decline of mean fitness at the wave front, negative values (blue) mean that the analytical approximation underestimates the decline of mean fitness at the wave front. In the gray-shaded area the relative error is larger than 5%. Parameter values are as in Figure S4.

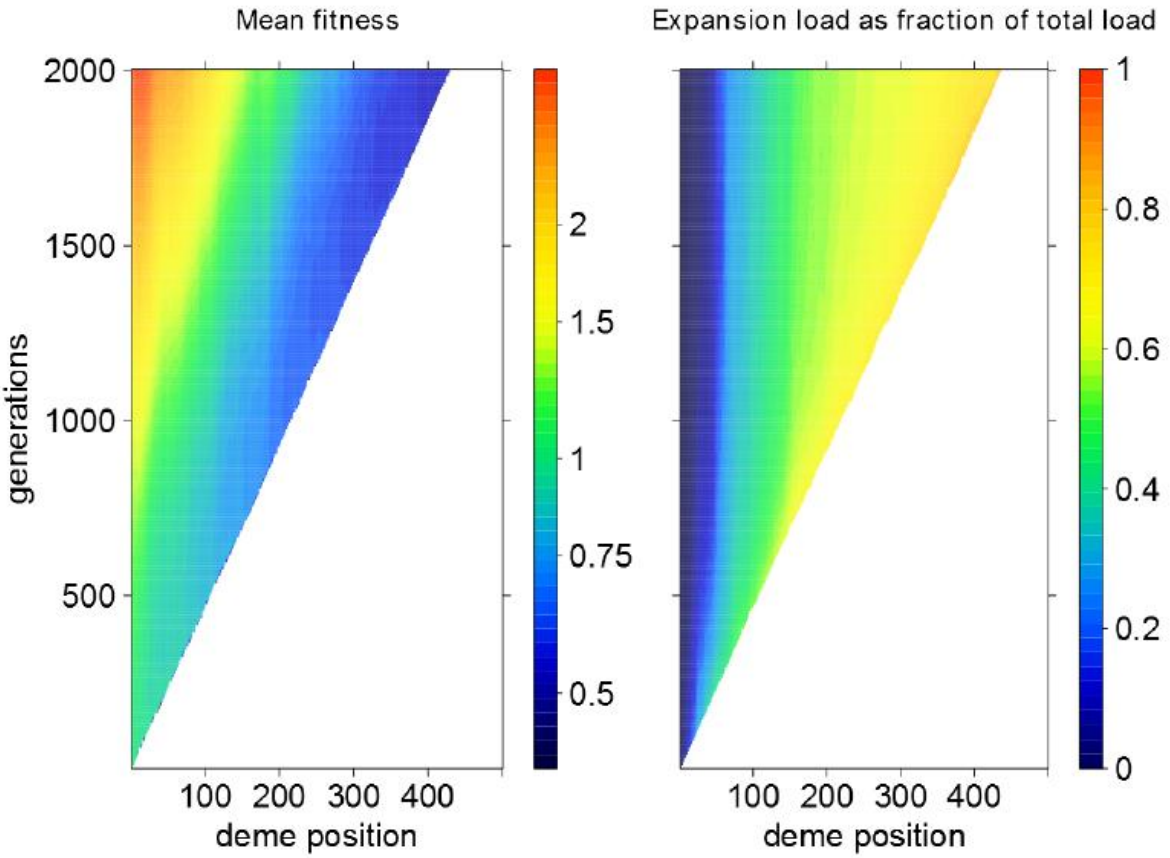

Figure S13. Evolution of population mean fitness and expansion load during a range expansion in a two-dimensional habitat. A: mean population fitness (normalized to 1 at the onset of the expansion). B: fraction of the total load due to mutations originating on the wave front. A mutation is considered to have originated at the front if it first appeared in an individual living in the deme currently at the edge of the expansion or one deme behind it. The habitat is 10 demes wide. The plots show averages over the width of the habitat. Results are for individuals with $n = 20$ freely recombining regions. Parameter values are $s = \pm 0.005, K = 100$, $r = \log(2)$, $m = 0.05$, $u = 0.05$, $\varphi_d = 0.9$.

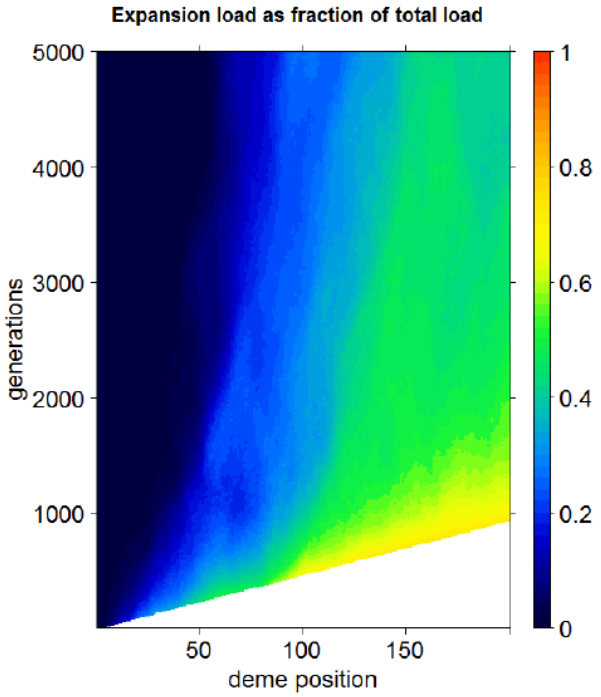

Figure S14. Fraction of total load that originated from the wave front (expansion load) for a linear range expansion in a finite two-dimensional habitat. The habitat is 20x200 demes. The plot shows the average over the width of the habitat. Results are for individuals with n = 20 freely recombining regions. Parameter values are $s = \pm 0.005$, $K = 100$, $r = \log(2)$, $m = 0.05$, $u = 0.05$, $\varphi_d = 0.9$.

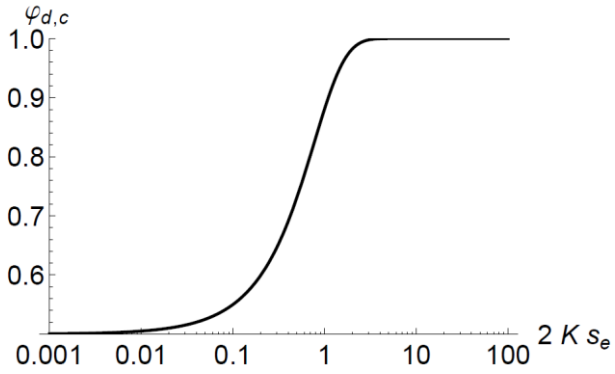

Figure S15. Critical value of φ_d as a function of $2Ks_e$. Mean fitness at the wave front decreases for values of φ_d above the solid lines.

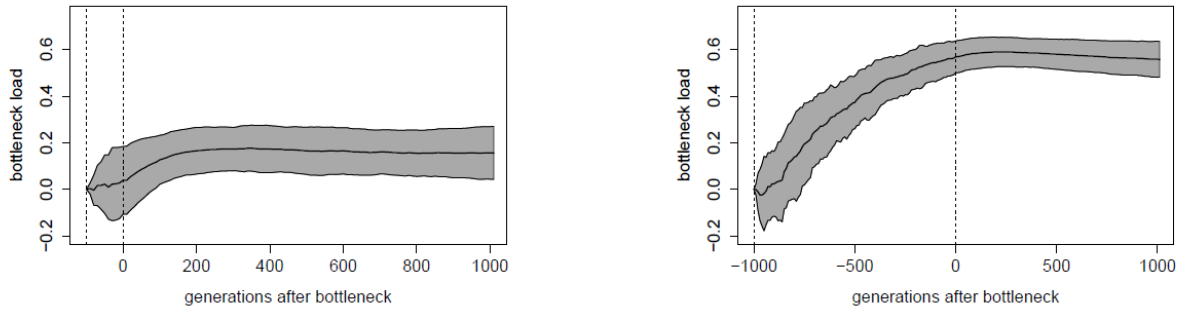

Figure S16. Fraction of the total load that is established during a single bottleneck. Thick black line shows the average over 50 realizations and gray area indicates two standard deviations. Vertical dashed lines indicate beginning and end of the bottleneck. Results are for $n = 20$ freely recombining regions. Parameter values are $s = \pm 0.005$, $u = 0.05$, $\varphi_d = 0.9$, $N = 10000$, $N_B = 100$, $T = 100$ (left), and $T = 1000$ (right).

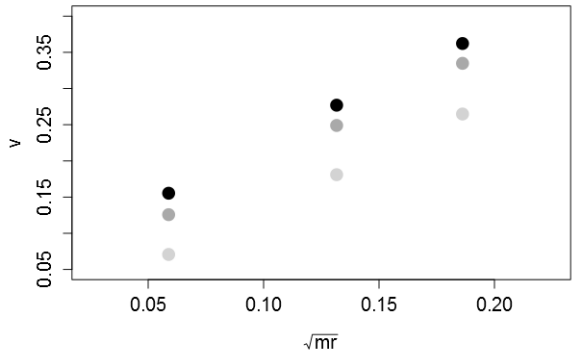

Figure S17. Speed (*v*) of a linear 1D expansion as a function of $\sqrt{mr}$ (Skellam, 1951). Parameter values are $r = \log(2)$, $m = 0.01, 0.05$, and 0.1. Carrying capacities are $K = 25$ (light gray), $K = 100$ (gray), and $K = 250$ (black).

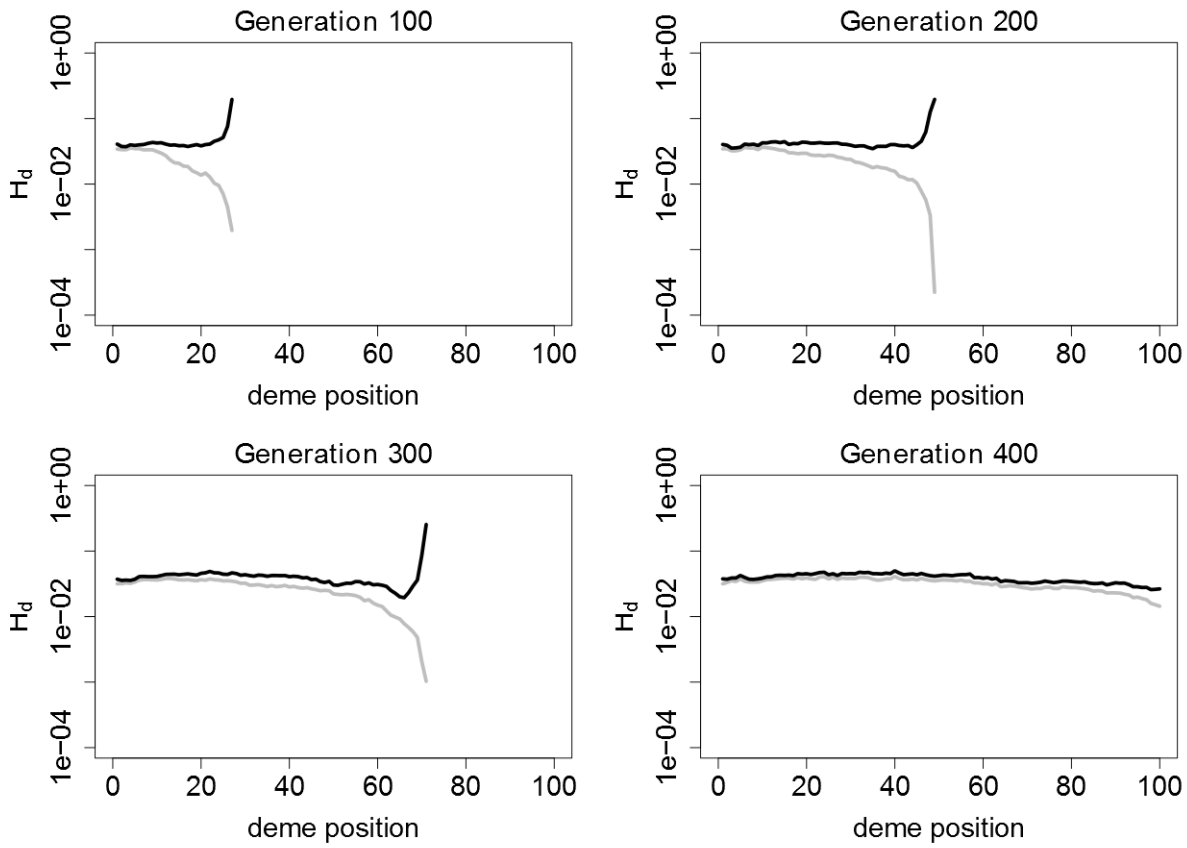

Figure S18. Heterozygosity at deleterious sites (H_d) during a 2D range expansion on a 10x100 grid. Gray lines show the average expected heterozygosity within demes along the expansion axis. Heterozygosity is calculated for each deme individually and then averaged over columns of demes. Black lines show the expected heterozygosity along the expansion axis if allele frequencies are pooled for each column of demes. Parameter values are as in Figure 1.

References

Barton N, Etheridge A, Kelleher J, Véber A (2013) Genetic hitchhiking in spatially extended populations. *Theoretical population biology*.

Fisher RA (1937) The wave of advance of advantageous genes. *Ann Hum Genet* **7**, 355-369.

Hallatschek O (2011) The noisy edge of traveling waves. *Proceedings of the National Academy of Sciences* **108**, 1783-1787.

Hallatschek O, Nelson DR (2008) Gene surfing in expanding populations. *Theor Popul Biol* **73**, 158-170.

Kimura M (1964) Diffusion models in population genetics. *Journal of Applied Probability* **1**, 177-232.

Mueller C, Mytnik L, Quastel J (2011) Effect of noise on front propagation in reaction-diffusion equations of KPP type. *Inventiones mathematicae* **184**, 405-453.

Skellam JG (1951) Random dispersal in theoretical populations. *Biometrika* **38**, 196-218.

Slatkin M, Excoffier L (2012) Serial founder effects during range expansion: a spatial analog of genetic drift. *Genetics* **191**, 171-181.

Wahl LM, Gerrish PJ (2001) The probability that beneficial mutations are lost in populations with periodic bottlenecks. *Evolution* **55**, 2606-2610.