Mathematical Appendix

In this part, for the model described in the main text, we obtain analytically the average density of sequences with k mutant loci, f_k, and the average rate of accumulation of deleterious mutations (or fixation of advantageous mutations), dk_{av}/dt. We restrict our attention to the case $q\mu L \ll s \ll \mu L$. In the next three sections, we will make use of six approximations (numbered 1 - 6), whose validity will be verified analytically in a separate section.

Discrete semideterministic approach

The deterministic equation for $f_k(t)$ has a form (1)

$$f_k(t+1) = e^{-\mu L} \sum_{i,j \geq 0} f_{k+j-i}(t) e^{-s(k+j-i-k_{av})} \frac{[\mu L(1-q(k+j-i))]^i}{i!} \frac{[\mu L q(k+j-i)]^j}{j!}, \tag{1}$$

where k is the number of uncompensated mutant loci, t is time (the generation number), μ is the mutation rate per locus, s is the selection coefficient at each locus, L is the total number of loci, and $q - 1/L$ is the frequency of compensating loci per deleterious variant at a locus. The sum over the numbers of forward, i, and back or compensating, j, mutations is limited by the condition $0 \leq k+j-i \leq L$. The term $e^{-s(k+j-i-k_{av})}$ is fitness of a sequence with $k+j-i$ uncompensated mutant loci relative to the average fitness defined as $e^{-sk_{av}} \equiv \sum_k f_k e^{-sk}$. The distribution $f_k(t)$ is normalized according to $\sum_k f_k(t) = 1$.

As we show below, when the population size is not too large, $N \ll (1/s)\ln(1/\mu L q k_{av})$, we can neglect terms with $i+j > 1$ in the right-hand side of Eq. 1, which correspond to multiple mutations in one sequence in one generation step (Approximation 1). In the same interval of N, we replace the fitness term in Eq. 1 with its linear expansion in k (Approximation 2). Under these approximations, and assuming $\mu L \ll 1$, Eq. 1 takes a form

$$f_k(t+1) - f_k(t) = \mu L(1-(k-1)q)f_{k-1}(t) - [\mu L + s(k-k_{av})]f_k + \mu L(k+1)q f_{k+1}, \tag{2}$$
$$k = 1, 2, \ldots, L-1.$$

For $k = 0$ and $k = L$, the first and, respectively, third terms in the right-hand side of Eq. 2 are absent.

The treatment presented so far is purely deterministic. To introduce the effect of random genetic drift into the model and obtain dependence of the results on the population size N, we will treat the first nonempty group at the left edge of the distribution, denoted $k = k_0$, stochastically, using a one-locus-type diffusion equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{2N} \frac{\partial^2}{\partial f_{k_0}^2}(f_{k_0}\rho) + \frac{\partial}{\partial f_{k_0}}\left[(Sf_{k_0} - M(t))\rho\right], \quad [f_{k_0}\rho(f_{k_0})]_{f_{k_0}\to 0} = 0, \tag{3}$$

$$S \equiv \mu L + s(k_0 - k_{av}), \tag{4}$$

$$M(t) \equiv \mu L(k_0+1)q f_{k_0+1}(t), \tag{5}$$

where $\rho(f_{k_0}, t)$ is the probability density, S and M are the effective selection coefficient and the mutation rate, respectively, for the equivalent one-locus model. The term $-[Sf_{k_0} - M(t)]$ under the first derivative in Eq. 3 represents the average change in the frequency f_{k_0} per generation, as given by Eq. 2. Multiplying both sides of Eq. 3, first, by f_{k_0} and, second, by $f_{k_0}^2$ and integrating in f_{k_0} by parts, for the averages of f_{k_0} and $f_{k_0}^2$, we obtain

$$d\overline{f}_{k_0}/dt = M - S\overline{f}_{k_0}, \tag{6}$$

$$d\overline{f_{k_0}^2}/dt = \overline{f}_{k_0}(1/N + 2M) - 2S\overline{f_{k_0}^2}, \tag{7}$$

respectively. Multiplying Eq. 6 by $2\overline{f}_{k_0}$ and subtracting it from Eq. 7, for the variance $V_{k_0} = \overline{f_{k_0}^2} - (\overline{f}_{k_0})^2$, we get

$$dV_{k_0}/dt = \overline{f}_{k_0}/N - 2SV_{k_0}. \tag{8}$$

According to the Main Approximation, we treat the remaining groups, $k \geq k_0 + 1$, deterministically, as given by Eq. 2. (This approach, at this point, is analogous to that used in refs. (2-4) in the case of $q = 0$ and large N, when the accumulation rate is very small.) We declare the edge group k_0 to be lost due to random drift and start treating stochastically the next group to the right, $k_0 + 1$, with the initial condition $V_{k_0+1} = 0$, if the following conditions are met

$$\overline{f}_{k_0} = 0: \quad d\overline{f}_{k_0}/dt < 0, V_{k_0} > (\overline{f}_{k_0})^2, \tag{9}$$

which corresponds to Muller's ratchet. Conversely, we declare group k_0 to become "deterministic" and start treating stochastically the next group to the left, $k_0 - 1$, with the initial conditions $\overline{f}_{k_0-1} = V_{k_0-1} = 0$, if

$$f_{k_0}\text{deterministic}: \ d\overline{f}_{k_0}/dt > 0, V_{k_0} < \left(\overline{f}_{k_0}\right)^2, \tag{10}$$

which corresponds to overall fixation of advantageous mutations. In the present work, we use the discrete equation, Eq. 2, and the cutoff conditions given by Eqs. 6, 8-10 for numeric computation of the time-dependent density $f_k(t)$ (see examples in Fig. 1).

Continuous semideterministic approach and the traveling wave solution

The derivation that follows, unlike the discrete approach in the previous section, applies in a much broader interval of N, $\ln(N/N^*) \ll (1/s)^2 \ln^2(1/s\mu L\alpha)$, because it allows for multiple mutations per sequence per generation and does not rely on expanding fitness in k (Approximations 1 and 2 are not used). We divide both sides of Eq. 1 by $f_k(t)$. As we show in *Validity of Approximations*, at $s \ll \mu L$, the relative change in $\ln f_k(t)$ between k and $k+1$ and between t and $t+1$ is small, so we can replace $\ln[f_{k+j-i}(t)/f_k(t)]$ with its linear expansion, $(j-i)\partial \ln f_k(t)/\partial k$ (Approximation 3) and $\ln[f_k(t+1)/f_k(t)]$ with $\partial \ln f_k(t)/\partial t$ (Approximation 4). We emphasize that $f_k(t)$, unlike its logarithm, cannot be expanded in a Taylor series in k, because, as we show below, in an important region of k, we have $|f_k - f_{k-1}| \sim f_k$. Further, at $k_{\mathrm{av}} \gg \ln N/\ln(\sigma/q)$, the characteristic width of f_k in k is much less than k_{av} (Approximation 5). Hence, we can replace $(k+j-i)q$ in Eq. 1 with $\alpha \equiv qk_{\mathrm{av}}$ and the upper limit of the sum in i and j in Eq. 1 with infinity. Under these simplifications, the sum in i and j can be evaluated exactly, and Eq. 1 becomes

$$\frac{\partial \ln f_k}{\partial \tau} = (1-\alpha)e^{-\partial \ln f_k/\partial k} + \alpha e^{\partial \ln f_k/\partial k} - \sigma(k - k_{\mathrm{av}}) - 1, \tag{11}$$

where $\tau \equiv \mu Lt$, $\sigma \equiv s/(\mu L) \ll 1$, and we used the inequality $s \ll 1$. A partial solution of equation 11 is a traveling solitary wave

$$\ln f_k(t) = \phi(k - k_{av}(\tau)). \tag{12}$$

where $x = k - k_{av}(\tau)$, and $\int dx \, e^{\phi(x)} = 1$. Substituting Eq. 12 into Eq. 11, we get an equation for $\phi(x)$

$$\sigma x = (1-\alpha)e^{-\phi'(x)} + \alpha e^{\phi'(x)} + v\phi'(x) - 1. \tag{13}$$

where $v \equiv \partial k_{av}/\partial \tau$ is the "velocity" of the wave in units of μL, which can be either positive or negative. Note that, because $\alpha = qk_{av}$ and v depend on time, $\phi(x)$, strictly speaking, depends on time as well. When deriving Eq. 13, we assumed $|\partial \phi/\partial \tau| \ll |v\phi'(x)|$, i.e., that the change in the distribution with time occurs mostly due to its shift as a whole rather than to the change of its shape (Approximation 6).

At small x, $|\phi'|$ is also small, and we can replace the exponentials in Eq. 13 with their linear expansion in ϕ'. Integrating the resulting equation with respect to $\phi(x)$ yields

$$\phi(x) = \ln\sqrt{\frac{\sigma}{2\pi(1-2\alpha-v)}} - \frac{\sigma x^2}{2(1-2\alpha-v)}, \ |x| \ll (1-2\alpha-v)/\sigma. \tag{14}$$

where the first term is found from the normalization condition for $\phi(x)$. The normalization integral in x converges, if $v < 1 - 2\alpha$, which determines the maximum wave velocity. Using Eq. 14, we can express the wave velocity in terms of the standard deviation in x or k, as given by

$$\sigma(\mathrm{std}_k)^2 + 2\alpha + v = 1. \tag{15}$$

Thus, at a given α, a wave moving to the left ($v < 0$, overall fixation of advantageous mutations) is broader than a wave moving to the right ($v > 0$, Muller's ratchet). (Note, that the integral $\int e^{-sx+\phi(x)}dx$ calculated with ϕ from Eq. 14 is not exactly equal to 1, as it should, according to the definition of k_{av}, but has a correction on the order of $s \, \mathrm{std}_k$ which is small within the validity range of Approximation 4.)

The wave solution defined implicitly by Eq. 13 exists at any value of $v < 1 - 2\alpha$. In order to specify the value of v, we have to use the cutoff conditions at the left edge (Eqs. 8-10). Near the edge, the Gaussian formula, Eq. 14, does not apply, and we must use the more general Eq. 13. We can express v in terms of N, α and σ without finding the explicit form of $\phi(x)$. Let us consider ϕ' the independent variable and x the function. At $v < 1 - 2\alpha$ and $\alpha < 1$, from Eq. 13, we have $x(\phi') \to +\infty$ at $\phi' \to \pm\infty$. The divergence of x at negative ϕ' implies an asymptotic decay of the distribution $f = e^{\phi(x)}$ on its right slope, $x \to \infty$. At $\phi' > 0$, the function $x(\phi')$ has an absolute minimum at $\phi' \equiv \ln u$, where u is given by

$$u = \frac{1}{2\alpha}\left[-v + \sqrt{v^2 + 4\alpha(1-\alpha)}\right]. \tag{16}$$

The existence of the minimum implies that, at $x < 0$, the distribution $f = e^{\phi(x)}$ ends at the point

$$x_0 \equiv x(\ln u) = -(1/\sigma)(-2\alpha u - v \ln u + 1 - v). \tag{17}$$

At smaller $x < x_0$, we have $f(x) \equiv 0$.

We consider now the difference

$$\phi(0) - \phi(x_0) = \int_{x_0}^0 \phi' \, dx = -x_0 \ln u - \int_{\ln u}^0 x(\phi')d\phi'. \tag{18}$$

The value of $\phi(x_0)$ in this equation can be estimated (see next paragraph) from the cut-off conditions, Eqs. 8-10, as given by

$$\phi(x_0) = -\ln[\mu L N \xi(\alpha, v)], \quad \xi(\alpha, v) \sim \begin{cases} 1, & \alpha \sim |v| \sim 1 \\ v \ln(e/v), & \alpha = 0, 0 < v < 1 \\ \alpha \ln^2(|v|/\alpha), & v < 0, |v| \gg \sqrt{\alpha} \end{cases}. \tag{19}$$

The right-hand side of Eq. 18 can be derived using Eq. 13, integrating in ϕ' and then using Eqs. 13 and 16. Finally, substituting $\phi(0)$ from Eq. 14 into Eq. 18, we arrive at the relation between v, α, and N

$$\sigma \ln(N/N^*) = 1 - 2\alpha - v - (v/2)\ln^2 u - v \ln u - 2\alpha u \ln u, \quad v < 1 - 2\alpha, \tag{20}$$

$$N^* = \sqrt{2\pi} \mathrm{std}_k / [\mu L \xi(\alpha, v)], \tag{21}$$

where u and std_k are given by Eqs. 16 and 15, respectively. Graphic representation of this result is given on Fig. 2a and b. According to definitions of α and v, we have

$$\frac{d\alpha}{dt} = q\mu L v(\alpha), \tag{22}$$

which determines, together with Eq. 20, the time dependence of α. Calculated trajectories $\alpha(t)$ are shown on Fig. 2c for different fixed values of $\sigma \ln(N/N^*)$. (Because we assume $\sigma \ll 1$, the dependence of N^* on α and v, Eq. 21, has a small effect on the trajectories.) The dependence of the characteristic wave width (standard deviation of k) on α derived from Eqs. 15 and 20 is shown on Fig. 2d. Comparison of the ratchet rate at $\alpha = 0$ and of the fixation rate with the results of Monte-Carlo simulation is presented in Fig. 2e and f. Fig. 2e also shows earlier results for the ratchet rate obtained in the limits of small and large N (4, 5).

Cutoff condition, Eq. 19. In the case of Muller's ratchet at $\alpha = 0$, we chose the initial moment $t = 0$, so that the value of $f_{k0}(0)$ is much higher than the stochastic threshold, implying $V_{k_0}(0) = 0$. Setting $M = 0$ in Eqs. 6 and 8, we obtain

$$\overline{f}_{k_0}(t) = f_{k_0}(0)e^{-St},$$

$$V_{k_0}(t) = \frac{f_{k_0}(0)}{NS}(e^{-St} - e^{-2St}) \approx \frac{\overline{f}(t)}{NS}. \tag{23}$$

From Eqs. 23 and 9, we obtain that the stochastic threshold is reached at $\overline{f}_{k_0} \approx 1/(NS)$. Using Eqs. 4, 17, and 16 at $\alpha \to 0$, we arrive at the second estimate in Eq. 19. In the case of fixation, we choose $t = 0$ at the time of appearance of new stochastic group k_0. Solving Eqs. 6 and 8 with initial conditions $\overline{f}_{k_0}(0) = V_{k_0}(0) = 0$ yields

$$\overline{f}_{k_0}(t) = e^{-St} \int_0^t dt' e^{St'} M(t'),$$

$$V_{k_0}(t) = \frac{1}{N} e^{-2St} \int_0^t dt' e^{2St'} \overline{f}_{k_0}(t'), \tag{24}$$

where S and $M(t)$ are given by Eqs. 4 and 5. For the traveling wave solution, at small times t, the time dependence of $f_{k_0+1}(t)$ can be approximated by

$$f_{k_0+1}(t) = f_{k_0+1}(0)e^{\beta t}, \quad \beta \equiv -\mu L v \phi'(x_0)t = (-\mu L v \ln u)t, \tag{25}$$

where u is given by Eq. 16. Substituting Eq. 25 into Eq. 5 and the latter into Eqs. 24, we get

3

$$\overline{f}_{k_0}(t) = \frac{\mu L \alpha f_{k_0+1}(0)}{S+\beta}(e^{\beta t} - e^{-St}),$$

$$V_{k_0}(t) = \frac{\mu L \alpha f_{k_0+1}(0)}{N(S+\beta)}\left[\frac{e^{\beta t}}{2S+\beta} - \frac{e^{-St}}{S} + \frac{S+\beta}{S(2S+\beta)}e^{-2St}\right]. \tag{26}$$

The wave moves one notch in time $t \sim 1/\mu L|v|$. On this time scale, the right-hand sides of Eqs. 26 can be estimated as

$$\overline{f}_{k_0}(t) \sim \frac{\alpha f_{k_0+1}(0)}{|v|}e^{\beta t},$$

$$V_{k_0}(t) \sim \frac{\alpha f_{k_0+1}(0)}{\mu L N[v\ln(|v|/\alpha)]^2}e^{2\beta t}, \tag{27}$$

where we used Eqs. 4, 25, 17, and 16 and assumed $|v| \gg \sqrt{\alpha}$. From the cutoff condition 10 and Eq. 27, we obtain the third estimate in Eq. 19. Technically, this derivation based on Eqs. 6 and 8 applies in the case $\ln(N/N^*) \ll (1/s)\ln(1/\mu L\alpha)$, when we can neglect with multiple mutations and expand fitness in k (Approximations 1 and 2). Still, Eq. 19 can be used, as a rough estimate, at larger N as well, because the inaccuracy affects only the argument of a very large logarithm in the right-hand side of Eq. 20.

Particular cases

Three asymptotic limits of Eq. 20 are worth writing down separately. The first case is when back/compensating mutations are almost absent, i.e., α is small. Keeping the linear correction in α, from Eqs. 16, 17, and 20, we obtain

$$x_0 = -(1/\sigma)\left[1 - v + v\ln v - \alpha(1-v^2)/v\right], \tag{28}$$

$$\sigma \ln(N/N^*_{\alpha=0}) = 1 - v\left[\tfrac{1}{2}\ln^2 v - \ln v + 1\right] + \frac{\alpha(1-v)}{v}[1 - v + (1+v)\ln v], \tag{29}$$

$$\alpha \ll v^2 \ln^2 v.$$

At $\alpha = 0$, Eq. 29 corresponds to the fat line in Fig. 2a. Note that, in this case, v is always positive, because fixation requires the presence of back/compensating mutations. The Muller's ratchet rate v vanishes at the point in N such that $\sigma \ln(N/N^*) = 1$ (see a comment in Approximation 3 below).

The second case is the steady state. Setting $v = 0$ in Eqs. 16, 17, and 20, we obtain x_0 and the equilibrium position of the distribution center, $k_{\mathrm{av}} = \alpha/q$, as given by

$$x_0 = -(1/\sigma)\left[1 - 2\sqrt{\alpha(1-\alpha)}\right], \tag{30}$$

$$\sigma \ln(N/N^*_{v=0}) = 1 - 2\alpha - \sqrt{\alpha(1-\alpha)}\ln\frac{1-\alpha}{\alpha}, \quad v = 0. \tag{31}$$

Relation 31 is shown as the fat line in Fig. 2b. Steady state is possible at $\alpha < 1/2$ only.

The third case is a very large population, $\ln N \gg 1/\sigma$. In this interval, Muller's ratchet does not operate, and a steady state is always at $\alpha \approx 0$ (fat line in Fig. 2b). (More precisely, as explained below, steady state is at the value $\alpha = q/\sigma \ll 1$ predicted by the one-locus model.) The value of v determined by Eq. 20 is negative (fixation), and its absolute value is much larger than 1. Eqs. 17 and 20, in this interval of N, can be replaced by approximate formulae

$$x_0 = -\frac{|v|}{\sigma}\ln\frac{|v|}{\alpha} = -\frac{2\ln(N/N^*)}{\ln(\sigma \ln N/\alpha)}, \tag{32}$$

$$\sigma \ln(N/N^*) = \frac{|v|}{2}\ln^2\frac{|v|}{\alpha}, \quad N^* = \sqrt{2\pi|v|/\sigma} \tag{33}$$

$$\ln\frac{N}{N^*} \gg \frac{1}{\sigma}\ln\frac{1}{\alpha}, \quad k \gg \ln\frac{N}{N^*}/\ln\frac{\sigma}{q}.$$

We can also find the half-time of reversion of k mutant loci, $t_{1/2}$, defined as the time in which the wave center, k_{av}, travels from k to $k/2$

$$t_{1/2} \approx \frac{k}{2\mu L|v|} \approx \frac{k\ln^2\left(\frac{s\ln N}{\mu kqL}\right)}{4s\ln(N\ln^{-1/2}N)}. \tag{34}$$

The right-hand side of Eq. 34 differs from an estimate obtained by Maynard Smith, $t_{1/2} \sim (k/s)\ln(s/\mu qLk)$ (Eqs. 7 and 11 in ref. (6); in our notation, μqL is the effective back mutation rate per locus) by a factor inversely

proportional to $\ln N$. Due to this additional factor, at $\ln N \sim k \ln(\sigma/q)$, the right-hand side of Eq. 34 crosses over to the well-known result of the deterministic one-locus model, $t_{1/2} = (1/s) \ln(\sigma/q)$. At this point in $\ln N$, the left edge of the wave reaches the "wall" at $k = 0$, $|x_0| \sim k_{av}$ (Eq. 32), i.e., Approximation 5 ceases to apply. Beyond this point, the existence of a cutoff at the left edge is not important, because $f_{k=0} \gg 1/(\mu L N)$, and all the results become almost independent on N. Also, the quasilinear accumulation of advantageous mutations predicted by the moving wave ansatz is replaced by the exponential time dependence characteristic for a single locus, Eq. 49. The transition to the one-locus theory is intuitively expected: in the limit of large N, every genetic variant pre-exists in a population, and frequent mutations break down linkage disequilibrium. In agreement with this, models assuming infinite population size, generally, do not find any advantage of recombination for progressive evolution (7, 8). To obtain his estimate of the fixation time, Maynard Smith chose a scenario describing how an emerging subpopulation containing advantageous mutations passes through stochastic bottleneck ("becomes established") and how it spreads, due to selection, to the entire population. In particular, he assumed that clones that are established consecutively differ by one advantageous mutation and did not take into account a distribution over k. Our analysis shows that the distribution is broad, within the interval $\sim |x_0|$ (Eq. 32).

At $N \ll 1/\mu q L$, the one-locus fixation time is $t_{1/2} \sim 1/\mu q L N s$ (9). Assuming $\sigma = 0.01 - 0.1$ and using results for $1 - 2\alpha - v$ plotted in Fig. 2d, we obtain that the one-locus time is much shorter than the multilocus estimate, $t_{1/2} \sim k/\mu L |v|$, if $N \gg 1/s$, at any α. Thus linkage delays fixation (and, therefore, recombination confers advantage to population) in a broad interval of population sizes, given by

$$\ln(1/s) < \ln N \ll k \ln(\sigma/q). \tag{35}$$

Validity of approximations

In this section, we show analytically that Approximations 1 and 2 used in the discrete approach, Eq. 2, are valid if $\ln(N/N^*) \ll (1/s) \ln(1/\mu L \alpha)$. We also show that Approximations 3 - 6 used in the continuous approach are valid in a broader range of N, $\ln(N/N^*) \ll (1/s)^2 \ln^2(1/s \mu L \alpha)$, and at sufficiently large values of k_{av}, $k_{av} \gg \ln(N/N^*)/\ln(\sigma/q)$. For the parameter range typical for RNA viruses (main text), even at the maximum value of k_{av}, $k_{av} = 1/(2q)$, the second inequality is more restrictive with respect to N than the first one.

Approximation 1: Neglecting multiple mutations. The numbers of forward and back/compensating mutations i, j contributing most to the sum in Eq. 1 can be estimated as

$$i \sim \max(1, \mu L e^{-\partial f_k/\partial k}), \quad j \sim \max(1, \mu L \alpha e^{\partial f_k/\partial k}). \tag{36}$$

Here $\partial f_k/\partial k = \phi'(x)$ can be estimated from Eq. 13 in several representative intervals of x, as given by

$$e^{\phi'(x)} \sim \begin{cases} u, & x \approx x_0 \\ 1, & |x| \ll \text{ or } \sim (1 - 2\alpha - v)/\sigma \\ 1/(\sigma x), & x \gg (1/\sigma) \max(|v| \ln |v|, 1) \end{cases} \tag{37}$$

From Eq. 16, we estimate

$$u \sim \begin{cases} 1/\max(|v|, \sqrt{\alpha}), & v > 0 \\ \max(|v|, \sqrt{\alpha})/\alpha, & v < 0 \end{cases} \tag{38}$$

Because μL and α are less than 1, from Eqs. 36, 37, and 38, we obtain $j \sim 1$, i.e., multiple back/compensating mutations are negligible, unless $x \approx x_0$ and v is large negative, $|v| \gg 1/\mu L$. In the latter case, which corresponds to $\ln(N/N^*) \gg (1/s) \ln^2(1/\mu L \alpha)$ (Eq. 33), the most contributing j is given by $j \approx \mu L |v|$. We can neglect mutliple forward mutations, except in the far-right tail of the distribution, $x \gg 1/s$, in which region $i = sx$.

Approximation 2: Linear selection term. Replacing $\exp[-s(k - k_{av})]$ in Eq. 1 with its linear expansion is justified, if $s|x_0| \ll 1$. At $v > 0$, from Eqs. 17, 28, and 30, we find $s|x_0|$ to be less or on the order of μL, so the condition is met. At large negative v, from Eqs. 32 and 33, we obtain $\mu L |v| \ln(|v|/\alpha) \ll 1$, or $\ln N \ll (1/s) \ln(1/\mu L \alpha)$.

Approximation 3: Continuity in k. Expanding the difference $\ln f_{k+j-i}(t) - \ln f_k(t)$ in Eq. 1 linearly in $j - i$ is justified, if $|(i - j)^2 \partial^2 f_k/\partial k^2| \ll |(i - j) \partial f_k/\partial k|$. In terms of the traveling wave solution, Eq. 12, this condition can be expressed as

$$|i - j| \ll \left| \phi' \frac{dx}{d\phi'} \right|. \tag{39}$$

Using Eqs. 13, 14, 16, and 17, we can estimate the right-hand side of Ineq. 39 in different intervals of x, as given by

$$\left| \phi' \frac{dx}{d\phi'} \right| \sim \begin{cases} |x|, & |x| \ll \text{ or } \sim (1 - 2\alpha - v)/\sigma \\ (\ln u) \max(\sqrt{v}, \alpha^{1/4})/\sqrt{\sigma}, & x \approx x_0, x - x_0 \sim 1 \\ |x \ln(\sigma x)|, & x \gg (|v| \ln |v|)/\sigma \end{cases} \tag{40}$$

Based on Eq. 40 and the estimates for i, j obtained when verifying Aproximation 1, we obtain that the condition, Eq. 39, is met due to $\sigma \ll 1$, unless $x \approx x_0$ and either v and α are small, $v \sim \sqrt{\alpha} \sim \sigma$, or v is large negative, $|v| \gg (1/s\mu L) \ln^2(|v|/\alpha)$. In the latter case, which corresponds to (Eq. 33) $\ln(N/N^*) \gg (1/s^2) \ln^4(1/\alpha s\mu L)$, the relevant value of j is large. The authors (2-4) who studied the ratchet for $N \gg \exp(1/\sigma)$, $\alpha = 0$, obtained a finite, albeit small ratchet speed, $v \sim \sigma$, at $N \sim \exp(1/\sigma)$. The reason for the difference between this earlier result and our result, $v = 0$ at $\sigma \ln(N/N^*) > 1$, is that the continuous approximation employed here is not valid, when both v and α are very small.

Note that, at $x \approx x_0$ and large negative v, we have $j\phi' >> 1$. A more accurate condition for neglecting ϕ'' in this case would be $j^2 |\phi''| \ll 1$, which yields $\ln N \ll (1/s)^{4/3} \ln^2(|v|/\alpha)$ (Eqs. 37 and 40). However, even at larger N, the error from neglecting ϕ'' corresponds to a factor multiplying α in the second term in Eq. 20. Because α enters the final expressions, Eqs. 32 and 33, only in the argument of a large logarithm, these expressions remain sufficiently accurate in the interval of N specified in the previous paragraph.

Approximation 4: Continuity in time. Replacing the difference $\ln f_k(t + 1) - \ln f_k(t)$ in Eq. 1 with the time derivative is justified, if $|\partial^2 f_k(t)/\partial t^2| \ll |\partial f_k(t)/\partial t|$, or, for the traveling wave solution, if $\mu L|v| \ll |\phi'(dx/d\phi')|$. Using the estimates given by Eqs. 40 and 38, we obtain that the condition is met due to $\sigma \ll 1$, two cases except:

(i) $|x| \ll \mu L|v|$. The upper limit on $|x|$ is much less than the standard deviation in k or x, Eq. 15, and the restriction on $|x|$ is not important, unless $|v| \gg 1/(s\mu L)$, which corresponds to $\ln(N/N^*) \gg (1/s^2) \ln^2(1/\alpha s\mu L)$ (Eq. 33).

(ii) $|v| \gg 1/(s\mu L) \ln^2(|v|/\alpha)$, i.e., at $\ln(N/N^*) \gg (1/s^2) \ln^4(1/\alpha s\mu L)$ (Eq. 33). In this case, the Approximation is violated at the left edge.

Approximation 5: The distribution is far from the origin. To replace qk by qk_{av}, we assumed that the distribution is narrow compared with its distance from the origin, $|x_0| \ll k$. At $v > 0$ or $v < 0, |v| \sim 1$, as can be shown from Eqs. 17, 28, 30, this condition always holds, as long as $\alpha \gg q/\sigma$ (hence the restriction $q \ll \sigma$ stated in the beginning). At large negative v (Eqs. 32, 33), the condition $|x_0| \ll k$ is equivalent to $\ln N \ll k \ln(\sigma/q)$ (hence the inequality in Eq. 33).

Approximation 6: Slow change of the shape. To obtain Eq. 13, we neglected $\partial\phi/\partial t$, assuming it to be much less than the time derivative due to the shift, $\mu Lv\phi'(x)$. We verify this assumption at the most important point, the distribution edge $x = x_0$, where we have

$$\partial\phi/\partial t = \phi'(x_0)\frac{dx_0}{dt} = \mu Lqv \ln u \frac{dx_0}{d\alpha}, \tag{41}$$

u, x_0, and v are given by Eqs. 16, 17, and 20. From Eq. 41, the validity condition reads

$$|dx_0/d\alpha| \ll 1/q. \tag{42}$$

The form of the derivative in α can be found by calculating $dv/d\alpha$ from Eq. 20 at fixed σ and N and then using Eq. 17 to find $dx_0/d\alpha$. The general expression for $dx_0/d\alpha$ is rather complex; however, it can be estimated in particular cases. In the case $v > 0$, and either $v \sim 1$ or $\alpha \sim 1$, we get $dx_0/d\alpha \sim 1/\sigma$. If both $\alpha \ll 1$ and $0 < v \ll 1$, we have $dx_0/d\alpha \sim 1/[\sigma \max(\sqrt{\alpha}, v \ln v)] \ll 1/\sigma$. Because we assume $q \ll \sigma$, the validity condition, Eq. 42, holds everywhere in the plane $(\alpha, \ln N)$ except in a small vicinity of the point $(0, 1/\sigma)$ where the ratchet is very slow (the dots in Fig. 2a and b). Finally, at large negative v corresponding to the range of N, $(1/\sigma) \ln(1/\alpha) \ll \ln N$, we can use asymptotic Eqs. 32, 33 for x_0, v to calculate $dx_0/d\alpha$; inequality 42 becomes $q|v|/(\sigma\alpha) \ll 1$. At the smallest α allowed by Approximation 5, $\alpha \sim q|x_0|$, we obtain $\ln(\sigma/q) \gg 1$ which holds marginally. At larger α, the condition holds better.

Correction to the continuous-in-k approximation at $\alpha = 0$

At realistically small values of $\sigma = 0.1 - 0.01$, corrections to the accumulation rate due to discreteness of k may be noticeable, especially at small N. At $\alpha = 0$, we can efficiently correct for these effects solving the discrete Eq. 2 near the left edge of the distribution in k. Suppose, at $t = 0, f_{k-1}(t)$ have reached the cut-off value given by Eqs. 8 and 9 and vanished. From Eq. 2, we have

$$df_k/d\tau = -\gamma f_k$$
$$df_{k+i}/d\tau = -\gamma f_{k+i} + f_{k+i-1}, \quad i = 1, 2, \ldots, \tag{43}$$

where $\tau \equiv \mu Lt, \gamma \equiv 1 + x_0\sigma = v \ln(e/v)$ (Eqs. 16 and 17). We can write down periodic conditions

$$f_{k+i}(\delta) = f_{k+i-1}(0), \quad i = 1, 2, \ldots, \tag{44}$$

where $\delta \equiv 1/v$ is the interval of τ in which the wave moves by one unit in k. Solving, e.g., the first five of Eqs. 43, and using periodic conditions and the identity $e^{\gamma\delta} = e\delta$, we obtain

$$
\begin{aligned}
f_k(\tau) &= f_k(\delta)\delta e^{1-\gamma\tau}, \\
f_{k+1}(\tau) &= f_k(\delta)\delta e^{1-\gamma\tau}[(e-1)\delta + \tau], \\
f_{k+2}(\tau) &= f_k(\delta)\delta e^{1-\gamma\tau} \left[\frac{\tau^2}{2} + (e-1)\delta\tau + \left(e^2 - 2e + \frac{1}{2}\right)\delta^2\right], \\
f_{k+3}(\tau) &= f_k(\delta)\delta e^{1-\gamma\tau} \left[\frac{\tau^3}{6} + \frac{e-1}{2}\delta\tau^2 + \right. \\
& \qquad \left. \left(e^2 - 2e + \frac{1}{2}\right)\delta^2\tau + \left(e^3 - 3e^2 + 2e - \frac{1}{6}\right)\delta^3\right], \\
f_{k+4}(\tau) &= f_k(\delta)\delta e^{1-\gamma\tau} \left[\frac{\tau^4}{24} + \frac{e-1}{6}\delta\tau^3 + \frac{1}{2}\left(e^2 - 2e + \frac{1}{2}\right)\delta^2\tau^2 + \right. \\
& \qquad \left. \left(e^3 - 3e^2 + 2e - \frac{1}{6}\right)\delta^3\tau + \left(e^4 - 4e^3 + \frac{9}{2}e^2 - \frac{4}{3}e + \frac{1}{24}\right)\delta^4\right]. \tag{45}
\end{aligned}
$$

We compare the continuous result, Eq. 20, with the discrete solution, Eq. 45, at the time when the left end of the continuous distribution is half-integer in k. The resulting correction to Eq. 20 is equivalent to the replacement $N^* \to N^* N_{\mathrm{cor}}$, where N_{cor} is given by

$$\ln N_{\mathrm{cor}} = \ln\left[N f_{k+i}(\delta/2)\right] - (i + 1/2)\ln\delta. \tag{46}$$

From Eqs. 45 and 46, within a 5% accuracy, we extrapolate $N_{\mathrm{corr}} = 1.6 + 2.0i \to 0.6 + 2.0|x_0|$, where we used $i = |x_0| - 1/2$ and Eq. 28. The corrected expression for N^* has a form

$$N^* = (0.6 + 2.0|x_0|)[(2\pi/\sigma)(1-v))]^{1/2}/[\mu L v \ln(e/v)], \quad \alpha = 0. \tag{47}$$

Note that the correction to the Muller's ratchet speed due to discreteness of k is contributed from the entire left slope of the distribution, not just the left edge, and is therefore approximate. Still, the corrected formula 20 yields good accuracy when compared with the results of Monte-Carlo simulation (Fig. 2e).

Fit of the one-locus model to vesicular stomatitis virus data

In this section, we write down some predictions of the one-locus model (which applies in the limit of very strong recombination) to compare them with data on passaged vesicular stomatitis virus (10, 11) and with the corresponding predictions of the multilocus model, as shown in Fig. 4. We will denote by $w = \exp[-s(k_{\mathrm{av}} - k_1)]$ the average fitness of a population with respect to a reference sequence having k_1 uncompensated deleterious mutations. (In Fig. 4, k_1 is one of the three fitting parameters.)

In a steady-state population of a size N, the average relative fitness, w_{ss}, is given by

$$\ln w_{ss} = k_1 s - L\left[\mu + \frac{s}{1 + qLe^{2Ns}}\right], \tag{48}$$

where the expression in brackets, at $s \gg \mu$, closely approximates the "mutation load" per locus (decrease in the log fitness due to deleterious mutations) obtained by Kimura *et al* (12). We used notation $qL = \nu/\mu$ for the average ratio of the effective (including compensating mutations) back mutation rate to the forward mutation rate.

Suppose now we start from a very large monomorphic population ($N \gg 1/\mu$), comprised of the reference sequence. The population will increase its average relative fitness with time t, as given by a deterministic expression

$$
\begin{aligned}
\ln w(t) &= k_1 s[1 - f_{\mathrm{fix}}(t)] - (L - k_1)s f_{\mathrm{acc}}(t), \\
f_{\mathrm{fix}}(t) &= \frac{\mu}{s} + \frac{1 + (\mu/s)(qL - 2)}{1 + (\mu/s)(-1 + qLe^{st})}, \quad f_{\mathrm{acc}}(t) = \frac{\mu}{s}(1 - e^{-st}), \tag{49}
\end{aligned}
$$

where $f_{\mathrm{fix}}(t)$ and $f_{\mathrm{acc}}(t)$ are the frequencies of deleterious variants at loci that undergo fixation of advantageous mutations and accumulation of deleterious mutations, respectively, depending on their initial state (9).

References:

1. Haigh, J. (1978) *Theor. Popul. Biol.* **14**, 251-267.
2. Stephan, W., Chao, L., and Smale, J. G. (1993) *Genet. Res.* **61**, 225-31.
3. Charlesworth, B., and Charlesworth, D. (1997) *Genet. Res.* **70**, 63-73.
4. Gordo, I. (2000) *Genetics* **154**, 1379-1387.
5. Lande, R. (1998) *Genetica* **102/103**, 21-27.
6. Maynard Smith, J. M. (1971) *J. Theor. Biol.* **30**, 319-335.
7. Barton, N. H. (1995) *Genet. Res., Camb.* **65**, 123-144.
8. Felsenstein, J. (1974) *Genetics* **78**, 737-756.
9. Rouzine, I. M., Rodrigo, A., and Coffin, J. M. (2001) *Microbiol. Mol. Biol. Rev.* **65**, 151-85.
10. Novella, I. S., Quer, J., Domingo, E., and Holland, J. J. (1999) *J. Virol.* **73**, 1668-71.
11. Novella, I. S., Elena, S. F., Moya, A., Domingo, E., and Holland, J. J. (1995) *J. Virol.* **69**, 2869-2872.
12. Kimura, M., Maruyama, T., and Crow, J. F. (1963) *Genetics* **61**, 763-771.