File S1

Supporting Text

1 Approximation for pairwise interaction diagrams

In the two diagrams (b) and (c) of Figure 2, we made the approximation that the advantage for the target mutation can be expressed as an increase in the initial frequency. In case of diagrams, (b), the factor of increase is given by $1/(1-x')$, where x' is the frequency of the deleterious background mutation at time τ, while in diagram (c) the factor of increase is given by $1/x'$, where x' is the frequency of the beneficial driver substitution at time τ. We simulated this situation with a two-locus model in which the beneficial background allele is introduced with frequency x' and the target mutation with frequency $x_0 = 0.01$ *within* the subpopulation carrying the background allele. We then simulate a Wright-Fisher process and measure the fixation probability of the target mutation.

In Figure S1 we show results of a numerical simulation for the case of diagram (c). The validation for diagram (b) can then be read by substituting $x' \to 1 - x'$ in the x-axis. As can be seen, the model is accurate as long as the increase factor $1/x'$ is not too large. For values of $x' < 0.4$, the fixation probability is overestimated. In case of diagram (b) this condition is almost always fulfilled, as deleterious mutants have very small frequencies, i.e. $1 - x' \gg 0.4$. In case of diagram (c) the approximation overestimates the fixation probability if the driver substitution appeared shortly before the target mutation and hence still has small frequency.

In a previously described model [7] the fixation probability in an expanding subpopulation is computed explicitly, which yields a result that is comparable to our approach (see their equation 11). Their result is more accurate than the expression we provide above (expanding the initial frequency in Kimura's formula). In our case, modeling the effect of the expansion as an increase in the initial frequency is accurate enough, in particular since the dominant interference effect is provided by future interfering mutations and not by background mutations.

The effects of background selection have been subject to a large number of articles. Often, these studies find that background selection in fact retains a substantial net effect on the fixation probability of a target mutation [10, 1, 6]. These studies typically assume a mutation-selection balance of many deleterious mutations, with a constant deterministic influx of deleterious mutations. For this case, an argument from Fisher [3] shows that the fixation probability of a beneficial mutation is reduced (see [1]) by a factor $\exp(-U_\mathrm{d}/\sigma_\mathrm{d})$, where U_d is the rate and σ_d is the selection coefficient of deleterious mutations.

In our model, Fisher's argument does not hold, for two reasons: i) because of the presence of adaptive substitutions (selective sweeps), variance in the population is constantly removed, hence a stationary mutation-selection balance is never maintained; ii) because we consider an exponential distribution of selection coefficients, the number of deleterious mutations that are *stronger* in effect than the target mutation, are typically rare enough to be treated stochastically, as done in our pairwise interaction scheme using the diagrams of Figure 2.

2 Single site propagator

In the derivation of the fixation probability (see *Appendix B*, we use the time evolution of the mean allele frequency $M(t, x_0, \sigma) = \int_0^1 \mathrm{d}x\, x\, G_0(x, t, x_0, \sigma)$ for which we can derive an analytical expression using the

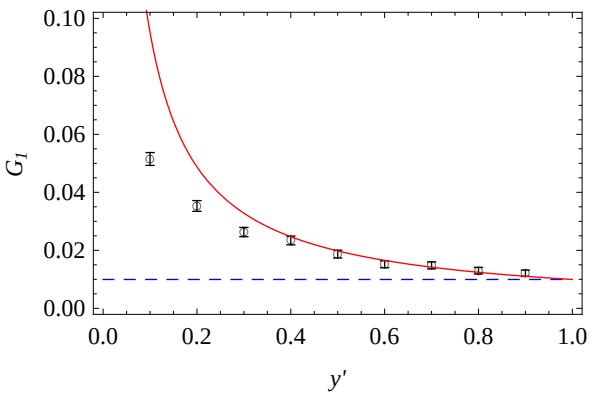

Figure S1: We plot the fixation probability of a beneficial target mutation ($2N\sigma = 10$) on the background of a strongly selected mutation ($2N\sigma' = 100$) initially present with frequency x'. The target mutation has initial frequency $x_0 = 0.01\%$. The black circles are simulation results, the blue dashed line is the expected fixation probability without background selection, the red line is the theory prediction by $G_0(x_0/x', \sigma)$. The population size is $N = 1000$.

diffusion approximation. The Fokker-Planck equation for a single site under drift and selection, using the selection coefficient σ and the population size N reads:

$$\partial_t G = \left[\frac{1}{2N}\partial_x^2(x(1-x)) - \sigma\partial_x(x(1-x))\right] G. \tag{S1}$$

To derive $M(t, x_0, \sigma)$, we multiply equation (S1) with x and integrate by parts, neglecting boundary terms:

$$\partial_t \int_0^1 \mathrm{d}x\, x\, G = \sigma \int_0^1 \mathrm{d}x\, x(1-x)G \tag{S2}$$

or

$$\partial_t M = \sigma(M - \int_0^1 \mathrm{d}x\, x^2 G). \tag{S3}$$

We introduce $M_2 = \sigma(\int_0^1 \mathrm{d}x\, x^2 G - M^2)$ to write

$$\partial_t M = \sigma M(1-M) - M_2. \tag{S4}$$

We identify two limits of the solution of this equation. First, for $t = 0$, the variance term M_2 must vanish and we see that M evolves logistically with initial value x_0. Secondly, if $\mu N \ll 1$, for times larger than the fixation time $\tau(\sigma) = (2\ln\sigma)/\sigma$ (see chapter *Fixation Time*), G becomes stationary and will consist of two delta peaks at 0 and 1 with weights reflecting the fixation probability G_0:

$$G(x, t, x_0, \sigma) \xrightarrow{t>\tau(\sigma)} (1-G_0)\delta(x) + G_0\delta(1-x), \tag{S5}$$

and

$$M_2 \xrightarrow{t>\tau(\sigma)} G_0(1-G_0). \tag{S6}$$

For the stationary mean allele frequency we get

$$\partial_t M_{\mathrm{stat}} = 0 \quad \Rightarrow \quad M_{\mathrm{stat}} = G_0. \tag{S7}$$

Motivated by the form of S4 and the known limit S7, we can try a logistic ansatz for beneficial and an exponential for deleterious mutations:

$$M_+(t, x_0, \sigma) = \frac{G_0(x_0, \sigma)}{1 + e^{-\hat{\sigma}t}(G_0(x_0, \sigma)x_0^{-1} - 1)} \tag{S8}$$

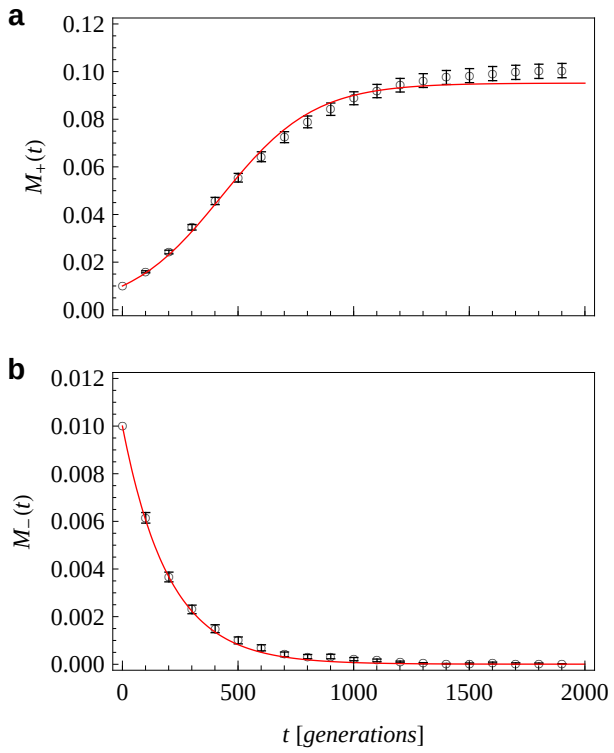

Figure S2: This plot shows the mean frequency of a beneficial (a) and deleterious (b) mutation as a function of time, measured in generations. Black circles are data obtained from simulations, red solid lines are the theory predictions of equation S8 and S9. Simulation data has been obtained from many trajectories started with different random seeds. Error bars indicate the standard error of the mean. Parameters: $N = 1000$, $2N\sigma = 10$, $x_0 = 10/N$.

and

$$M_-(t, x_0, \sigma) = x_0 e^{-\hat{\sigma} t} + (1 - e^{-\hat{\sigma} t}) G_0(x_0, -\sigma) \tag{S9}$$

with the standard fixation probability

$$G_0(x_0, \sigma) = \frac{1 - e^{-2N\sigma x_0}}{1 - e^{-2N\sigma}} \tag{S10}$$

and the regularized selection coefficient $\hat{\sigma}$ that has the two limits $\hat{\sigma} = 1/2N$ for $N\sigma \ll 1$ and $\hat{\sigma} = \sigma$ for $N\sigma \gg 1$. The exact form of this crossover is not important. Here we choose

$$\hat{\sigma} = \begin{cases} 1/2N & \text{for } N\sigma \leq 1 \\ \sigma & \text{for } N\sigma > 1 \end{cases} \tag{S11}$$

The predictions from equations S8 and S9 agree with simulations, as shown in Figure S2.

3 Fixation Times

The deterministic equation for the frequency of a mutation under selection σ is

$$x(t) = \frac{1}{1 + \exp(-\sigma'(\tau - \tau'))(x_0^{-1} - 1)} \tag{S12}$$

Since evolution is always stochastic at the boundaries, we define the fixation time as the time it takes to reach a frequency $1 - 1/(2N\sigma)$, starting from a frequency $x_0 = 1/(2N\sigma)$. This leads to

$$\tau_{\text{fix}}(\sigma) = \frac{2\log(2N\sigma)}{\sigma}. \tag{S13}$$

For small values of σ, the dynamics is not anymore deterministic and the above expression becomes wrong as can be seen by the limit value for neutral alleles, where we require $\tau_{\text{fix}} = 2N$. For values of $2N\sigma < 3$, we simply set

$$\tau_{\text{fix}}(\sigma) = 2N \quad 2N\sigma < 3. \tag{S14}$$

We typically consider average selection coefficients of $2N\bar{f} \gg 1$, and the average driver mutation has even larger selection coefficients. The exact setting of the above threshold is therefore not important.

4 Iterative Solution

As discussed in the main text, the following set of equations has to be solved:

$$p_{\text{drive}}(\sigma) = \exp(-\tau_{\text{fix}}(\sigma)V_>(\sigma)), \tag{S15}$$

$$V_{\text{drive}}(\sigma) = U(\sigma)G(\sigma)p_{\text{drive}}(\sigma), \tag{S16}$$

$$V_>(\sigma) = \int_\sigma^\infty \mathrm{d}\sigma'\, V_{\text{drive}}(\sigma'), \tag{S17}$$

$$U(\sigma) = LN\mu\lambda_{\text{d}}(\sigma)\rho(\sigma), \tag{S18}$$

$$\lambda_{\text{d}}(f) = \frac{G(-f) + \gamma/(\mu N)}{G(f) + G(-f) + 2\gamma/(\mu N)}, \tag{S19}$$

$$\begin{aligned} G(\sigma) = &\int_{-\infty}^{\tau} \mathrm{d}\tau' \int_\sigma^\infty \mathrm{d}\sigma'\, V_{\text{drive}}(\sigma') e^{-V_>(\sigma)(\tau-\tau')} \int_0^1 \mathrm{d}x'\, \delta(x' - x_{\text{det}}(\tau - \tau'; \sigma')) \\ &\times x' G_0\left(\frac{1}{Nx'}, \sigma\right) {}_2F_1\left[1, \frac{V_>(\sigma)}{\sigma}, 1 + \frac{V_>(\sigma)}{\sigma}; 1 - Nx' G_0\left(\frac{1}{Nx'}, \sigma\right)\right], \quad \sigma > 0 \end{aligned} \tag{S20}$$

and

$$\begin{aligned} G(\sigma) = &\int_{-\infty}^{\tau} \mathrm{d}\tau' \int_{|\sigma|}^\infty \mathrm{d}\sigma'\, V_{\text{drive}}(\sigma') e^{-V_>(|\sigma|)(\tau-\tau')} \int_0^1 \mathrm{d}x'\, \delta(x' - x_{\text{det}}(\tau - \tau'; \sigma')) \\ &\times \frac{1}{N(|\sigma| + V_>(|\sigma|))} \left(Nx' G_0\left(\frac{1}{Nx'}, \sigma\right) |\sigma| + V_>(|\sigma|)\right), \quad \sigma < 0, \end{aligned} \tag{S21}$$

where ${}_2F_1(a, b, c; z)$ is a hypergeometric function.

The selfconsistent solution $G(\sigma)$ of equations S16, S18, S19, S17, S20 and S21 is found by the following numerical procedure: We initialize the iteration by setting $V_{\text{drive}}(\sigma) \equiv 0$. We then iterate the following steps:

1. Use equation S17 to compute the cumulative rate of drivers $V_>(\sigma)$.
2. Use $V_>(\sigma)$ to compute the fixation probabilities $G(\pm\sigma)$ from equations S20 and S21.
3. Use $G(\pm\sigma)$ to compute the stationary state $\lambda_{\text{d}}(f)$ from equation S19 and hence the mutation rate $U(\sigma)$ and the driver rate $V_{\text{drive}}(\sigma)$ from equations S18 and S16.

Steps 1. and 2. involve standard numerical integration methods, such as the function *NIntegrate* in Wolfram Research's *Mathematica*. To speed up the iterations, we evaluated all of the involved functions at 40 equidistant discrete values of σ between 0 and $\sigma_{\max} = 10\bar{f}$ and linear interpolations between the evaluated points. We used 10 iterations and observe a quick convergence after six iterations of the above algorithm.

5 Comparison with Gerrish-Lenski theory

Gerrish and Lenski (GL) [4] computed the fixation probability of a beneficial mutation in a manner similar to how we compute the rate of driver mutations. In GL-theory, the fixation probability is

$$G_{\mathrm{GL}}(\sigma) = U(\sigma)G_0(\sigma)e^{-\frac{1}{2}\tau(\sigma)\int_\sigma^\infty U(\sigma)G_0(\sigma)}. \tag{S22}$$

The factor 1/2 in the exponent follows from counting only those stronger mutations that appear on the background *not carrying* the target mutation. Only these mutations decrease the fixation probability of the target mutation. Equation S22 is analogous to equation S16, in which we require that a driver mutation is not interfered with by any stronger driver mutation. There are, however, two important differences: i) Equation S16 has no factor 1/2, since we exclude from the driver rate those mutations that fix by positive interference (hitchhiking). In contrast, in GL-theory, the only mode of fixation are driver mutations that are not suffering negative interference. ii) Equation S16 reflects a self-consistent closure, in which each interfering driver must itself be free from even stronger interfering drivers. Therefore, apart from the factor 1/2, equation S22 can be seen as a first iteration loop of the rate of driver mutations. Taking into account hitchhiking as a positive outcome of interference dramatically enhances the fixation probability of weakly beneficial mutations and in particular allows us to compute the influence on deleterious alleles, a case which is not covered by GL-theory.

Our genomic model yields beneficial mutation rates and their distribution as an outcome, while GL-theory takes distribution and rate of beneficial mutations as an input. In GL-theory all stronger driver mutations in the exponent of equation S22 are assumed to be free of interference. In our model this corresponds to assuming the unlinked rate of beneficial mutations, using results from single site theory, to be

$$U_0(\sigma) = \lambda_{0,\mathrm{d}}(f)\rho(f) \tag{S23}$$

and

$$\lambda_{0,\mathrm{d}}(f) = \frac{G_0(-f) + \gamma/(\mu N)}{G_0(f) + G_0(-f) + 2\gamma/(\mu N)}. \tag{S24}$$

A comparison with GL-theory is shown in Figures 3a, 3c and S3b as a brown curve.

6 Computer simulations

We simulate the Wright Fisher model with N individuals (N fixed throughout). Each individual consists of a sequence of L alleles, denoted by bits $a_{ij} = \{1, 0\}$ with $i = 1 \dots N$ and $j = 1 \dots L$. At every generation, alleles at any individual can mutate from 1 to 0 and vice versa with probability μ. The fitness F_i of individual i is obtained by a sequence of additive selection coefficients f_j multiplied by the time-dependent direction of selection $\eta_j(t) = \{-1, 1\}$:

$$F_i = \sum_{j=1}^{L} a_{ij}\eta_j(t)f_j. \tag{S25}$$

The local selection coefficients $f_j \geq 0$ are kept constant throughout evolution and fixed at values:

$$f_j = \frac{\bar{f}}{\Gamma(1+1/\kappa)}\left(-\log\left(1 - \frac{j-1}{L}\right)\right)^{1/\kappa} \tag{S26}$$

The numbers f_j reflect an ordered set of random variates from the Weibull distribution (as suggested in [9]):

$$\rho(f) = \frac{\kappa}{\zeta}\left(\frac{f}{\zeta}\right)^{\kappa-1} e^{-(f/\zeta)^\kappa} \tag{S27}$$

with $\zeta = \bar{f}/\Gamma(1+1/\kappa)$ such that the mean of this distribution is $\bar{f}$. For $\kappa = 1$ this distribution is exponential.

The direction of selection $\eta_j(t)$ is a stochastic random variable that flips its sign on average every $1/\gamma$ generations with the autocorrelation

$$\langle \eta_j(t)\eta_k(t')\rangle = e^{-\gamma|t-t'|}\delta_{jk}. \tag{S28}$$

A flip of direction affects the fitnesses of all individuals in the population. Each generation is stochastically sampled from the previous generation, using a multinomial sampling process, in which the probability p_i of picking individual i is given by:

$$p_i = e^{F_i - \langle F\rangle}, \tag{S29}$$

with the mean fitness

$$\langle F\rangle = \frac{1}{N}\sum_{i=1}^{N} F_i. \tag{S30}$$

A substitution is observed if a site that was monomorphic at some allele becomes polymorphic and then fixed at the other allele. Each substitution can be categorized as "beneficial", if the new allele is the currently fitter, or "deleterious" if it is the currently less fit of the two alleles. We also keep track of the fixation state: Over a sufficiently large number of generations, we can compute λ_j of site j as the fraction of time at which the frequency of the currently fitter allele was below 0.5. Since we generally consider low local mutation rates $\mu N \ll 1$, this fraction yields approximately the fraction of time the population was *fixed* at the locally less fit allele. At every site, the origination rate of new beneficial mutations can be computed as $\mu\lambda$ and of new deleterious mutations as $\mu(1-\lambda)$. Fixation probabilities can be computed by dividing the rate of beneficial/deleterious substitutions by the rate of beneficial/deleterious originations.

For the stationary adaptation simulations, we initialize the program with a monomorphic population where all alleles are 0. In each run of the simulation, we let the population equilibrate for $\min(1/\mu, 1/\gamma)$ generations before initiating any measurement to ensure that the population is in the stationary state.

In case of the approach to equilibrium, we first run the above protocol for stationary adaptation (for some parameter γ as given in the particular presentation of the results) for sufficiently long time to ensure stationarity. Then we set the flip rate $\gamma = 0$ and start obtaining measurements as described above. All results are now time-dependent. To obtain averages, we therefore have to repeat this program many times and average over the full ensemble of simulations for each time point.

The theoretical derivations for our model suggest some simple scaling laws. As can be seen, the population size N and the genome length L are only relevant in the parameter combinations $N\mu L$, $N\gamma L$ and $N\bar{f}$. As long as we keep these parameter combinations fixed, we can use smaller values for N and L to speed up the simulations. This scaling holds up to the following conditions: i) $\mu L \ll 1$ to avoid two mutations in the same individual in the same generation, ii) $\mu N \ll 1$ so that sites follow substitution dynamics with short polymorphic times and iii) $N\gamma \ll 1$ so that the time between selection flips is larger than the time needed for a fixation. These conditions can always be fulfilled for given parameters $N\mu L$, $N\gamma L$ and $N\bar{f}$.

7 Supplementary Results

7.1 Mutation and Fixation rate

As seen in equation S18, the rate of beneficial and deleterious mutations in our model depends on the dynamics. In Figure S3a) we plot the rate of mutations compared to the expectation without interference from linkage. The rate of beneficial mutations is enhanced, while the rate of deleterious mutations is decreased in comparison to single site theory. This increase is reflecting the fact that the population under linkage is less adapted to its environment and many sites in the genome are not fixed at the locally fitter allele. Hence mutations at these maladapted sites emit more beneficial mutations. It turns out that for beneficial mutations, $\sigma > 0$, the distribution $U(\sigma)$ is exponential. In Figure S3b) we show the rate of fixations, which is a product of the mutation rate and the fixation probability. The comparison with GL-theory emphasizes the point made in section 5: the rate of weakly beneficial fixations is underestimated due to neglecting

hitchhiking. Figure S3c) shows the theory prediction of the substitution rate (see S3b) together with its partitioning in passengers (gray shading) and drivers (green shading).

7.2 Positive and negative fitness flux

Since in our model we explicitly include the dynamics of deleterious mutations, we can disentangle the fitness flux into a positive fitness flux, constituted by the beneficial mutations, and a negative part, caused by the fixation of deleterious mutations:

$$\Phi(f) = \Phi_+(f) - \Phi_-(f) = fV(f) - fV(-f). \tag{S31}$$

In figure S4 we show this decomposition. As can be seen, the two terms of the fitness flux have their main contributions coming from different parts of the spectrum of selection coefficients: While the positive flux is mainly carried by strongly beneficial mutations, the negative flux consists of weaker deleterious fixations. In total, the positive flux is always much larger than the negative one, but the decomposition reveals an interesting pattern: Remarkably, even very strongly selected genomic sites provide a significant contribution to the negative flux, reflecting ubiquitous (strongly) deleterious passenger mutations.

7.3 Relation to mutation based models

The self-consistency of genomic state and mutations is one feature that distinguishes our model from most previous studies of adaptation under linkage [4, 12, 2, 8]. These mutation-based models constrain the distribution of selection coefficients for beneficial mutations, $u(\sigma) = (1/U_b)U(\sigma)$, to a fixed shape and use the total rate of beneficial mutations, U_b, and their mean effect, $\sigma_b = \int_0^\infty \sigma u(\sigma) d\sigma$, as independent input parameters. This is a suitable setup to evaluate the speed of adaptation at stationarity, because Φ depends in good approximation only on the distribution of beneficial mutations (see Figure S5a). However, mutation-based approaches of this type cannot predict genomic quantities such as the average degree of adaptation, α, which arguably is the most appropriate measure of the efficiency of the adaptive process over long periods of time. Even at stationarity, the average degree of adaptation is not uniquely determined by U_b and σ_b, but depends on all three genomic parameters $\bar{f}$, γ, and L in a nontrivial way (see Figure S5b). In our model, the rates $U(\sigma)$ of beneficial mutations (and, hence, U_b and σ_b) are dependent quantities, which must be derived from the self-consistent solution of the genome dynamics described in the main text. Changing any of the genomic parameters, say L, will change U_b and σ_b, so that these parameters are not suitable as input if we want to evaluate the dependence of the model on L. Similarly, U_b and σ_b change with time in a non-stationary adaptive process.

7.4 Epistasis

In the model presented in the manuscript, we use a strictly additive fitness model without any epistatic interactions between genomic sites. To check how epistasis affects our results, we employ a simple extension to our model that explicitly realizes pairwise epistatic interactions. This extension is done in the spirit of the study presented in reference [5]. In addition to the strictly additive contribution to an individuals fitness S25, we add another term for the pairwise interactions:

$$F_i = \sum_{j=1}^{L} a_{ij}\eta_j(t)f_j + e\sum_{j<k}(2a_{ij}-1)(2a_{ik}-1)f_{jk}, \tag{S32}$$

with the matrix f_{jk} describing the pairwise epistatic interactions between sites. In the simplest case realized here, we use normal distributed random numbers with mean 0 and standard deviation 1, fixed throughout evolution. The impact of the epistatic interactions can be estimated by comparing the magnitude of the additive fitness effects with the magnitude of the epistatic effects. As can be seen from equation S32, the

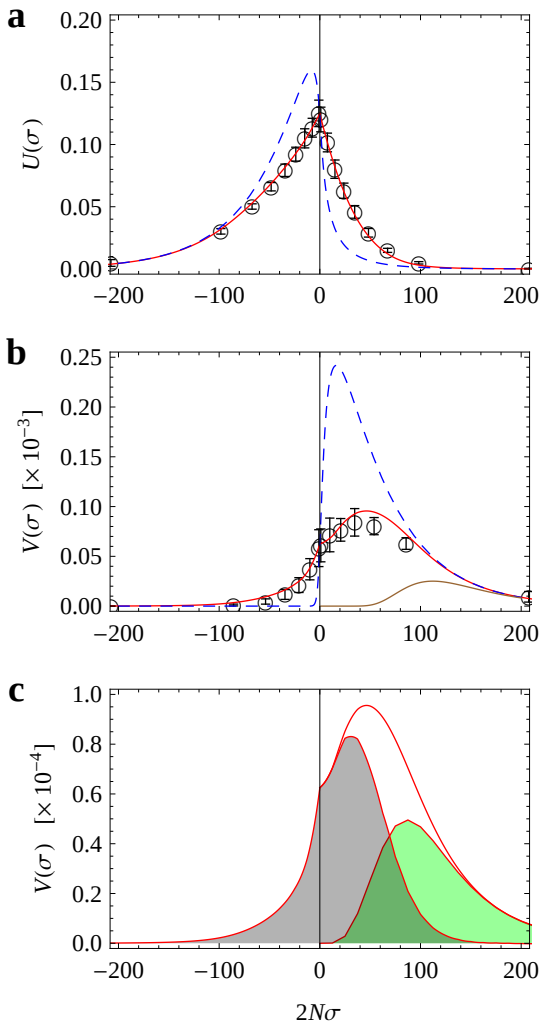

Figure S3: a) rate of mutations, b) rate of substitutions, c) theory prediction of substitutions with partitioning into passengers (gray) and drivers (green). Black circles: Simulation data, blue line: single site theory, red line: self-consistent theory, brown line: GL-theory. Parameters are $N = 2000$, $L = 1000$, $2N\gamma = 0.1$, $2N\mu = 0.025$, $2N\bar{f} = 50$.

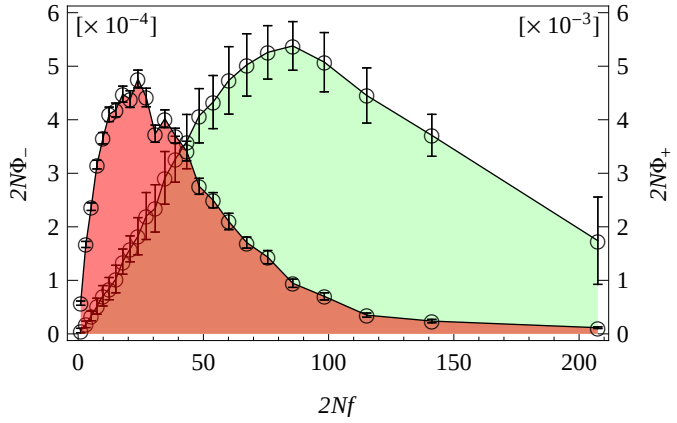

Figure S4: In green: Positive fitness flux, In red: negative fitness flux. For clarity, the negative fitness flux is shown with a ten-fold amplification to give a more direct comparison. Black circles: Simulation data ($N = 2000$, $L = 1000$, $2N\gamma = 0.1$, $2N\mu = 0.025$, $2N\bar{f} = 50$).

epistatic term scales as L^2 (double sum), whereas the additive term scales as L. We therefore have

$$e \sim \frac{\bar{f}}{L} \tag{S33}$$

as the typical crossover at which epistatic interactions dominate the additive fitness.

Figure S6 shows the degree of adaptation as a function of the site selection coefficient f, for both the strictly additive model (blue curve) and the epistatic model at the crossover, i.e. $e = \bar{f}/L$. Clearly, although the epistatic contribution to the fitness function is of the same magnitude as the additive part, the degree of adaptation is very robust under epistasis. Figure S7 shows the total degree of adaptation for varying values of the scaled epistasis parameter $\hat{e} = e(L-1)/\bar{f}$. The predicted crossover is given by $\hat{e} = 1$. In summary, epistasis does not quantitatively change our model and results, as long the epistatic interactions are of the same magnitude as the additive component of the fitness.

7.5 Mutation rate during approach to equilibrium

In the main text we show the degree of adaptation during an approach to equilibrium. Here we show additionally how the mutation rate itself depends on time in such a scenario. Figure S8 shows the distribution of fitness effects of new mutations at three different time points. Initially, the population is poorly adapted, so there are many beneficial mutations available. As the population approaches equilibrium, more mutations become deleterious, as seen in the plot. We use the same simulation protocol as described in the main text: the initial population has been evolved in stationarity with $2N\gamma = 0.1$. To observe the approach to equilibrium, we set $\gamma = 0$.

7.6 Non-exponential distributions of selection coefficients

In addition to the exponential distribution results shown in the main text, we present two additional cases: For $\kappa = 1/2$ the tail of the distribution is a stretched exponential with broadly distributed selection coefficients. For $\kappa = 2$ we recover a gaussian tail and hence more sharply distributed selection coefficients. In Figure S9 we show simulation results with theory predictions for the three distribution shapes $\kappa = 1/2, 1, 2$. While the case $\kappa = 1$ is predicted most accurately by our theory, our prediction slightly deviates from the simulations using the two non-exponential distributions.

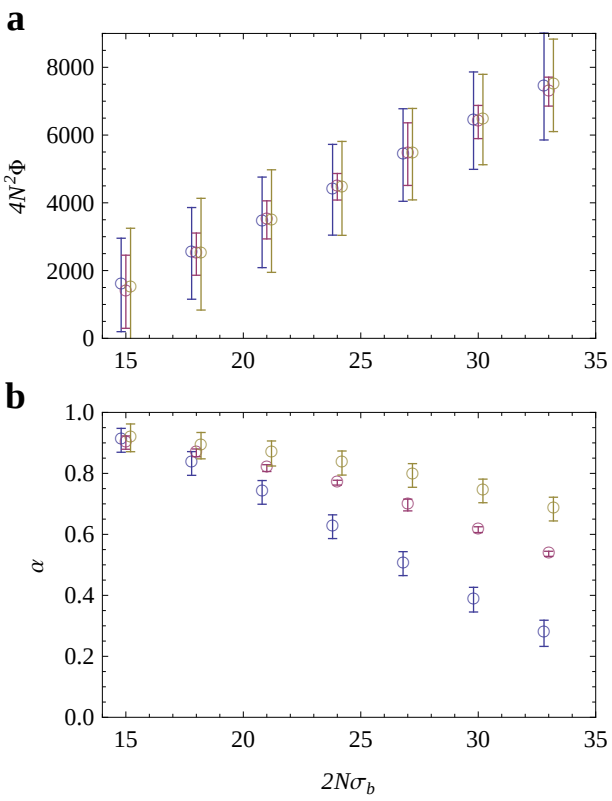

Figure S5: This plot shows a) the scaled fitness flux $4N^2\Phi$ and b) the degree of adaptation α as a function of σ_b at *fixed* $U_\mathrm{b} = 30$. The data points have been obtained by interpolating simulation results with varying γ, L and μ with three different fixed values of $2N\bar{f} = 40, 50, 60$ in blue, red and yellow. Error bars are predicted by the interpolation method (see Numerical Recipes 3rd edition [11] "Kriging"). In a) circles have been shifted horizontally by $\epsilon = \pm 0.2$ to make distinction between the data points possible.

References

[1] Nick H Barton. Linkage and the limits to natural selection. *Genetics*, 140(2):821–41, Jun 1995.

[2] Michael M Desai, Daniel S Fisher, and Andrew W Murray. The speed of evolution and maintenance of variation in asexual populations. *Curr Biol*, 17(5):385–94, Mar 2007.

[3] RA Fisher. The genetical theory of natural selection. *Clarendon Press, Oxford*, 1930.

[4] P J Gerrish and Richard E Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102-103(1-6):127–44, Jan 1998.

[5] Richard A Neher and Boris I Shraiman. Competition between recombination and epistasis can cause a transition from allele to genotype selection. *PNAS*, 106(16):6866–71, Apr 2009.

[6] H Allen Orr. The rate of adaptation in asexuals. *Genetics*, 155(2):961–8, Jun 2000.

[7] S P Otto and M C Whitlock. The probability of fixation in populations of changing size. *Genetics*, 146(2):723–33, Jun 1997.

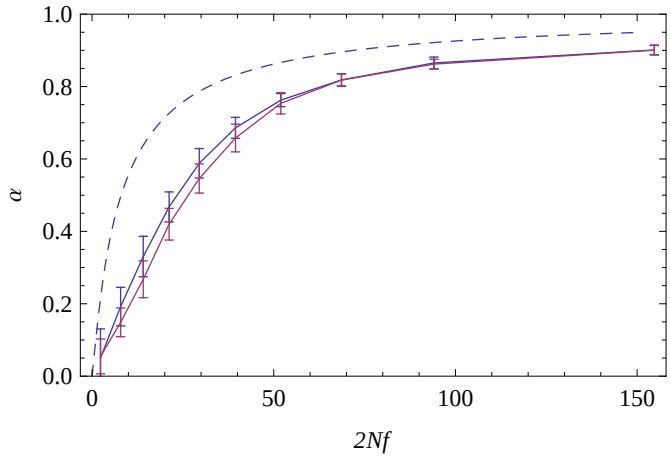

Figure S6: This plot shows the degree of adaptation as a function of the local selection coefficient f_j. The blue curve is the result from a simulation with no epistasis, $e = 0$, while the red curve was simulated with the epistatic fitness component being equal to the additive component: $e = \bar{f}/(L-1)$. Other parameters are $N = 1000$, $L = 1000$, $2N\gamma = 1$, $2N\mu = 0.025$, $2N\bar{f} = 50$.

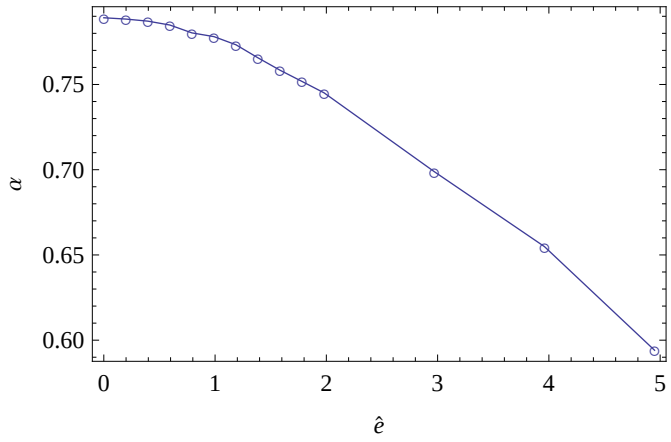

Figure S7: Here we show the total degree of adaptation with the same parameters as in Figure S6, with varying values of scaled epistasis $\hat{e}(L-1)/\bar{f}$. For $\hat{e} > 1$ the epistatic component of the fitness is larger than the additive component.

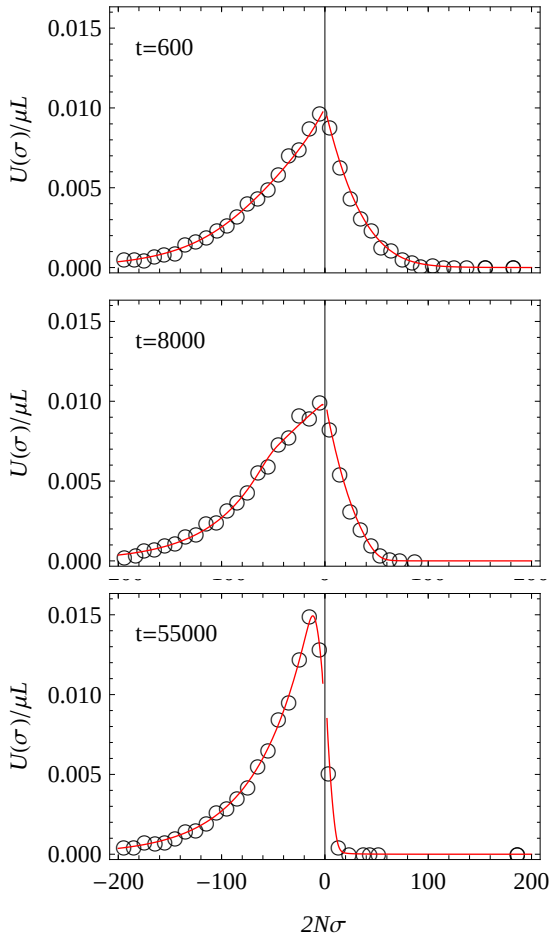

Figure S8: This Figure shows simulation and theory results for an approach to equilibrium, starting from an initially poorly adapted state (see main Text). We plot the distribution of new mutations for three different time points. Simulation results are shown in circles, theory predictions from numerically solving equation 5 in the main text are shown in red. Parameters are $N = 1000$, $L = 1000$, $2N\bar{f} = 50$, $2N\mu = 0.025$. The initial state is a stationary state with $2N\gamma = 0.1$.

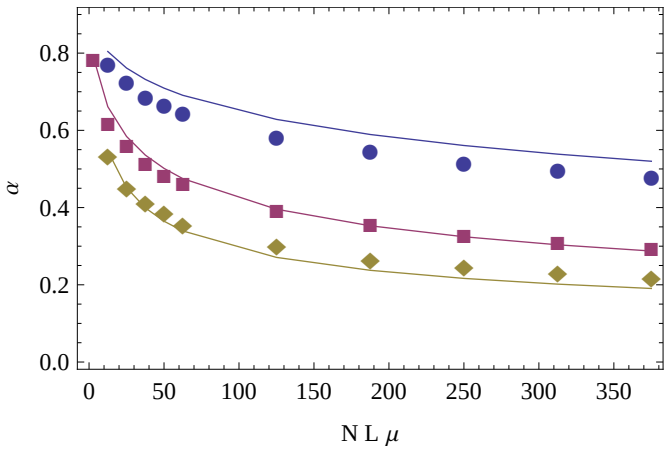

Figure S9: This plot shows simulation results for the degree of adaptation as a function of the total mutation rate. We show results for three values of $\kappa = 0.5$ (Circles), $\kappa = 1$ (Rectangles) and $\kappa = 2$ (Diamonds). Other parameters are $N = 4000$, $2N\bar{f} = 50$, $2N\gamma = 0.1$, $2N\mu = 0.025$.

[8] Su-Chan Park and Joachim Krug. Clonal interference in large populations. *PNAS*, 104(46):18135–40, Nov 2007.

[9] Su-Chan Park, Damien Simon, and Joachim Krug. The speed of evolution in large asexual populations. *J Stat Phys*, 138(1-3):381–410, Jan 2010.

[10] J R Peck. A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics*, 137(2):597–606, Jun 1994.

[11] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. Numerical recipes 3rd edition: The art of scientific computing. *Cambridge University Press*, 2007.

[12] Claus O Wilke. The speed of adaptation in large asexual populations. *Genetics*, 167(4):2045–53, Aug 2004.