

# Evolution of evolvability in rapidly adapting populations

Received: 15 December 2023

Accepted: 29 July 2024

Published online: 11 September 2024

 Check for updates

James T. Ferrare<sup>1</sup> & Benjamin H. Good<sup>1,2,3,4</sup> 

Mutations can alter the short-term fitness of an organism, as well as the rates and benefits of future mutations. While numerous examples of these evolvability modifiers have been observed in rapidly adapting microbial populations, existing theory struggles to predict when they will be favoured by natural selection. Here we develop a mathematical framework for predicting the fates of genetic variants that modify the rates and benefits of future mutations in linked genomic regions. We derive analytical expressions showing how the fixation probabilities of these variants depend on the size of the population and the diversity of competing mutations. We find that competition between linked mutations can dramatically enhance selection for modifiers that increase the benefits of future mutations, even when they impose a strong direct cost on fitness. However, we also find that modest direct benefits can be sufficient to drive evolutionary dead ends to fixation. Our results suggest that subtle differences in evolvability could play an important role in shaping the long-term success of genetic variants in rapidly evolving microbial populations.

The benefits of new mutations can manifest over multiple timescales. Some mutations alter the short-term fitness of an organism, while others can also affect the rates and fitness benefits of future mutations. Examples are common in the microbial world. Mutations in DNA repair genes can generate mutator strains with dramatically elevated mutation rates<sup>1–3</sup>. These variants also alter the molecular spectrum of new mutations, which can shift the relative probabilities of adaptive mutations in addition to their overall rates<sup>4–6</sup>. Other classes of mutations can open or close adaptive pathways through epistatic interactions with other genes<sup>7–12</sup>. Striking examples of these ‘evolvability modifiers’ have been observed in laboratory evolution experiments<sup>8,10,11,13</sup> (Fig. 1a), and they are thought to play a critical role in cancer<sup>14–16</sup> and the evolution of antibiotic resistance<sup>17–19</sup>. But despite their potential importance, it is difficult to predict when these long-term evolutionary benefits should be favoured by natural selection.

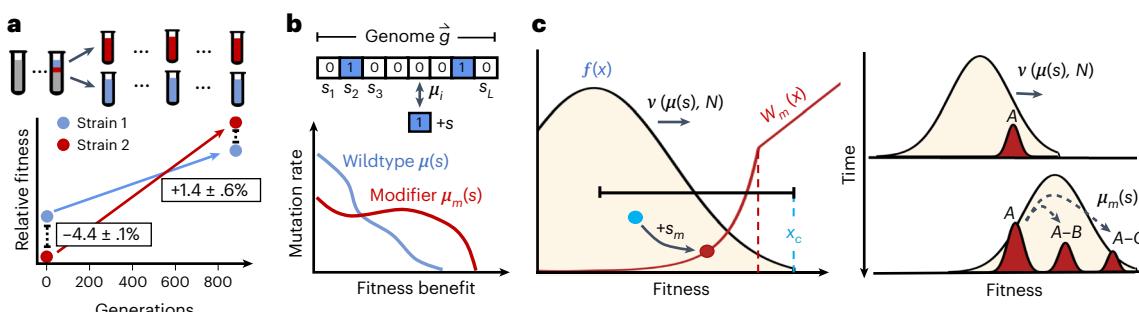
Classical arguments from modifier theory suggest that in a constant environment, asexual populations will select for mutations that maximize their long-term mean fitness<sup>20–22</sup> (Supplementary Section 1). This simple result applies for infinite populations near mutation-selection equilibrium. Both conditions are frequently violated in

adapting populations, as a beneficial variant can fix before its long-term costs or benefits are fully realized. This greediness creates an inherent tension between short-term and long-term fitness gains<sup>23–25</sup>.

While some of these trade-offs can be understood in simple cases where mutations accumulate one by one<sup>22,26–30</sup> (Supplementary Section 2), most microbial populations reside in a qualitatively different regime. In many cases of interest, from laboratory evolution experiments<sup>10,31–33</sup> to natural populations of viruses<sup>34–37</sup>, bacteria<sup>38–40</sup> and certain cancers<sup>14,15</sup>, multiple beneficial mutations can arise and compete within the population at the same time. The competition between these linked variants (‘clonal interference’<sup>41</sup>) ensures that a successful lineage must often generate multiple additional mutations to fix, which amplifies the indirect selection on their mutational neighbourhood. However, while recent work has started to explore these effects for mutation-rate modifiers alone<sup>26</sup>, we currently lack an analytical framework for understanding more general differences in evolvability in the high-diversity regimes most relevant for microbes.

Our limited understanding of these dynamics leaves many basic questions unresolved: How does natural selection balance the short-term costs or benefits of a mutation with its longer-term

<sup>1</sup>Biophysics Program, Stanford University, Stanford, CA, USA. <sup>2</sup>Department of Applied Physics, Stanford University, Stanford, CA, USA. <sup>3</sup>Department of Biology, Stanford University, Stanford, CA, USA. <sup>4</sup>Chan Zuckerberg Biohub – San Francisco, San Francisco, CA, USA.  e-mail: [bhgood@stanford.edu](mailto:bhgood@stanford.edu)



**Fig. 1 | Modelling indirect selection on evolvability in rapidly adapting asexual populations.** **a**, An empirical example of an evolvability modifier from ref. 8. Two *E. coli* strains isolated from a long-term evolution experiment<sup>78</sup> had an initial fitness difference of -4% (Methods). The less-fit strain showed a higher rate of adaptation in replay experiments, which allowed it to consistently overcome its initial disadvantage after ~900 generations of evolution. Fitness differences denote mean  $\pm$  s.e.m. from 20 independent replays. **b**, These long-term benefits can be modelled by the accumulation of beneficial mutations at a large number of linked genomic loci. The fitness benefits of the mutations are summarized by

their distribution  $\mu(s)/ds$ , which denotes the total rate of producing mutations with fitness effects  $s \pm ds/2$ . An evolvability modifier changes this distribution to a new value,  $\mu_m(s)$ . **c**, A modifier mutation (A) with a direct cost or benefit ( $s_m$ ) arises on a genetic background from the wildtype fitness distribution,  $f(x)$ , which has a maximum relative fitness  $x_c$  (left). The modifier lineage competes with the wildtype population as they both acquire further mutations (B and C; right). The outcome of this competition is described by the conditional fixation probability  $w_m(x + s_m)$ , which shows a sharp transition at a critical initial fitness  $x_{cm} \neq x_c$  (left).

impact on the fitness landscape? Which future mutations matter most for determining a lineage's long-term evolutionary fate? And how do the answers to these questions depend on extrinsic factors such as the size of the population or the diversity of competing mutations? In this Article, we address these questions by developing a population genetic theory of indirect (or 'second-order') selection that explicitly accounts for interference among competing beneficial mutations.

## Results

### Modelling indirect selection in rapidly adapting populations

While evolvability can be defined in many ways<sup>23–25</sup>, we focus on a simple model of indirect selection that is motivated by empirical examples such as Fig. 1a (Methods). We consider an asexual population of  $N$  individuals that can acquire beneficial mutations at a large number of linked genetic loci. In a constant environment, the mutations accessible to each genotype  $\vec{g}$  can be summarized by their distribution of fitness effects (DFE), denoted by  $\mu(s|\vec{g})/ds$ , which represents the per generation rate of producing mutations with fitness effects  $s \pm ds/2$  (Fig. 1b and Extended Data Table 1). We will initially assume that the beneficial sites are sufficiently numerous and epistasis sufficiently weak that the DFE remains approximately constant over the relevant timescales (which we determine self-consistently below). This implies that the rate of adaptation of the population will approach a steady-state value  $v(\mu(s), N)$  that depends on the size of the population and the shape of the DFE<sup>42,43</sup>. Given these assumptions, the simplest possible evolvability modifier is a mutant that shifts the DFE to a new shape,  $\mu(s) \rightarrow \mu_m(s)$ , which is maintained for several additional substitutions (Fig. 1b). This minimal model can be viewed as the lowest-order term in a more general expansion in the genotype dependence of  $\mu(s|\vec{g})$  (Extended Data Fig. 1 and Methods). It generalizes the notion of a mutator allele to capture more subtle changes in evolvability while bypassing the enormous complexity of the underlying fitness landscape.

If the modifier takes over the population, the rate of adaptation will shift to a new value  $v_m \equiv v(\mu_m(s), N)$  that reflects its altered supply of mutations (Fig. 1a). Following previous work<sup>8,44,45</sup>, we define an evolvability-enhancing mutation to be one that increases  $v_m$ , while an evolvability-decreasing mutation has the opposite effect. However, natural selection does not act on the long-term rate of adaptation directly: before a modifier can reach these high frequencies, it must initially grow from a single founding individual, where stochastic fluctuations and competition with other lineages both play an important

role. The outcome of this competition can be described by the fixation probability,  $p_{\text{fix}}(\mu(s) \rightarrow \mu_m(s))$ , which provides a quantitative measure of the mutant's long-term reproductive value<sup>46</sup>.

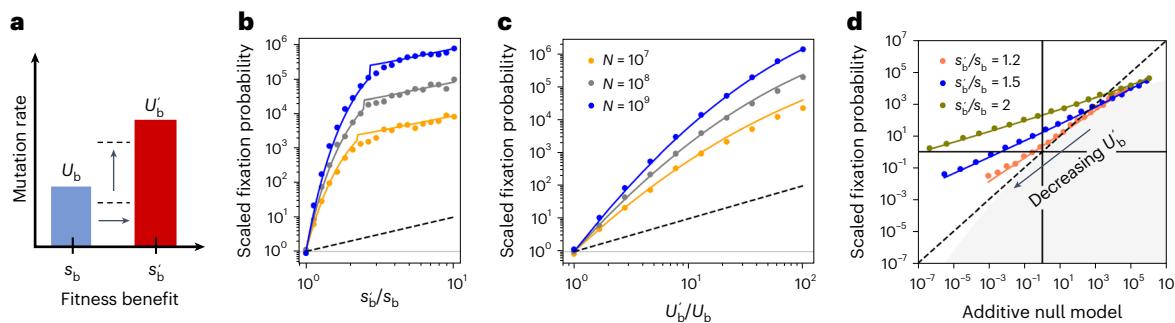
In large populations, the fate of a modifier mutation will strongly depend on its initial genetic background. While this distribution is complicated at the genetic level, previous work has shown that progress can be made by grouping individuals by their relative fitness and modelling the resulting dynamics in fitness space<sup>42,43,47</sup>. The distribution of fitnesses in the background population will approach a steady-state shape  $f(x)$  that increases in fitness at rate  $v \equiv v(\mu(s), N)$  (Fig. 1c). This distribution has a characteristic width,  $x_c$ , which also depends on the DFE and population size, and roughly coincides with the location of the fittest individuals in the population. A new modifier mutation will arise on a genetic background with a relative fitness drawn from  $f(x)$  and will then compete with the wildtype population while acquiring further mutations from the modified DFE  $\mu_m(s)$  (Fig. 1c, right). The outcome of this competition can be summarized by the conditional fixation probability  $w_m(x) \equiv p_{\text{fix}}(\mu(s) \rightarrow \mu_m(s)|x)$ , which depends on the mutant's initial relative fitness  $x$  (Fig. 1c, left). Building on previous work<sup>26</sup>, we show that this conditional fixation probability can often be approximated by the solution to the branching process recursion,

$$0 \approx \underbrace{x \times w_m(x)}_{\text{selection}} + \underbrace{\int \mu_m(s) [w_m(x+s) - w_m(x)] ds}_{\text{further mutations}} - \underbrace{v(\mu(s), N) \times \partial_x w_m(x)}_{\text{adaptation of the wildtype}} - \underbrace{\frac{1}{2} \times w_m(x)^2}_{\text{genetic drift}}, \quad (1)$$

which represents a balance between (1) the growth of the lineage due to selection, (2) the production of further mutations, (3) the adaptation of the wildtype population and (4) the stochastic effects of genetic drift (Methods). The overall fixation probability of the modifier can then be obtained by averaging over the relative fitness of its initial genetic background,

$$p_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m) = \int w_m(x + s_m) \times f(x) dx, \quad (2)$$

where we have also allowed the modifier to have a direct cost or benefit  $s_m$ . Together, equations (1) and (2) provide a quantitative framework for understanding the trade-offs between direct or ('first-order') selection on an immediate fitness change and



**Fig. 2 | Interference between linked mutations enhances indirect selection for evolvability.** **a**, A simplified model for the DFE in Fig. 1b, where all mutations confer the same characteristic fitness benefit. An evolvability modifier will change the overall mutation rate ( $U_b \rightarrow U'_b$ ), the overall selection strength ( $s_b \rightarrow s'_b$ ) or both. **b,c**, The fixation probability (scaled by the neutral expectation,  $1/N$ ) for a selection-strength modifier,  $\tilde{p}_{\text{fix}}(s_b \rightarrow s'_b)$  (b), or a mutation-rate modifier,  $\tilde{p}_{\text{fix}}(U'_b \rightarrow U_b)$  (c). Symbols denote the results of forward-time simulations (Methods) for  $s_b = 10^{-2}$ ,  $U_b = 10^{-5}$  and  $N = 10^7$ – $10^9$ . Solid lines denote our theoretical predictions (Supplementary Section 8), while the dashed lines denote the null expectation in the absence of clonal interference (Supplementary

Section 2). **d**, Fixation probability of a compound modifier,  $\tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b))$ , compared with an additive null model,  $\tilde{p}_{\text{fix}}(U_b \rightarrow U'_b) \times \tilde{p}_{\text{fix}}(s_b \rightarrow s'_b)$ . Symbols denote the results of forward-time simulations for  $s_b = 10^{-2}$ ,  $U_b = 10^{-5}$  and  $N = 10^8$ ; the y coordinates are obtained from simulations, while the x coordinates are obtained from the theoretical predictions in **b** and **c**. Solid lines denote our theoretical predictions, which deviate substantially from the additive expectation (dashed lines). The grey region indicates forbidden combinations that cannot arise in our theory (Supplementary Section 8).

indirect (or ‘second-order’) selection on the rates and benefits of future mutations.

The fixation probability of a neutral variant [ $\mu_m(s) = \mu(s), s_m = 0$ ] is always equal to  $1/N$  (ref. 48; Supplementary Section 3). This provides a natural scale for interpreting the fixation probability in equation (2). Modifiers with  $p_{\text{fix}} \gg 1/N$  are strongly favoured by natural selection, while those with  $p_{\text{fix}} \ll 1/N$  are effectively purged. For this reason, it will be convenient to examine the scaled fixation probability,  $\tilde{p}_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m) \equiv N \times p_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m)$ , so that the sign of  $\log \tilde{p}_{\text{fix}}$  coincides with the net ‘direction’ of natural selection<sup>49</sup>.

### Selection for evolvability in a simple fitness landscape

To understand the indirect selection pressures encoded in equations (1) and (2), we start by considering a simple model for the DFE, where deleterious mutations are neglected and all new mutations confer the same characteristic fitness benefit  $s_b$  (we eventually relax both assumptions below). If  $U_b$  denotes the total rate at which these mutations occur, then an evolvability modifier will change either the selection strength ( $s_b \rightarrow s'_b$ ), the mutation rate ( $U_b \rightarrow U'_b$ ) or some combination of the two parameters (Fig. 2a). This simplified model allows us to obtain an analytical solution for the fixation probability that is valid for empirically relevant scenarios where  $s_b \gg U_b \gg 1/N$  (Methods). It will be convenient to express these results in terms of the key fitness scales  $v$  and  $x_c$  in the wildtype population (Fig. 1c and Methods), which satisfy  $x_c \gg s_b \gg \sqrt{v}$  in the parameter ranges above.

To tease apart the contributions of indirect selection, we begin with the simplest case where there are no direct costs or benefits ( $s_m = 0$ ), and consider mutation-rate and selection-strength changes independently (Fig. 2b,c). For a pure selection-strength modifier ( $s_b \rightarrow s'_b$ ), we find that the fixation probability initially increases sharply with the new selection strength,

$$\log \tilde{p}_{\text{fix}}(s_b \rightarrow s'_b) \approx \left[ 1 - \left( \frac{s_b}{s'_b} \right)^2 \right] \log(\sqrt{(Ns_b)(NU_b)}), \quad (3)$$

before saturating to a linear dependence for larger values of  $s'_b$  (Fig. 2b). (Note: to streamline notation, we have omitted the arguments of the fixation probability that are held constant in equation (3).) The rapid increase in Fig. 2b is qualitatively different than that observed for mutation-rate modifiers<sup>26</sup>,

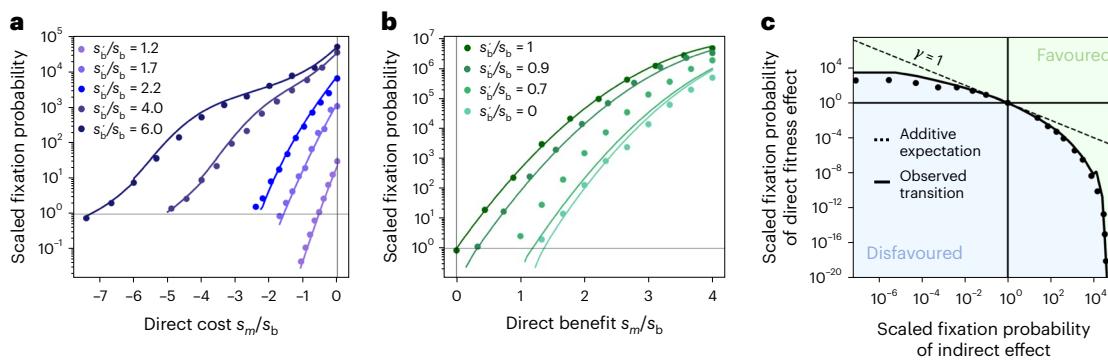
whose indirect benefits increase more slowly with the fold change in the mutation rate (Fig. 2c). In both cases, the fixation probabilities are orders of magnitude larger than the proportional scaling expected when mutations accumulate via discrete selective sweeps (Fig. 2b,c and Supplementary Section 2). This gap grows increasingly large as the supply of beneficial mutations ( $NU_b$ ) increases, showing that the competition between linked mutations can dramatically enhance selection on heritable differences in evolvability.

The origin of this behaviour can be heuristically explained using the key fitness scales in Fig. 1c. For small changes in the selection strength, successful modifiers typically arise in the high-fitness ‘nose’ of the population ( $x \approx x_c$ ) and must acquire  $\approx x_c/s_b$  additional mutations before they outcompete their rivals in the nose. In each step  $j$ , a selection-strength modifier produces  $\approx \exp[\tau \times j(s'_b - s_b)]$  more mutations than a wildtype individual in the time that it takes for the nose to advance by one mutation ( $\tau \approx s_b/v$ ). Multiplying these contributions together leads to the exponential scaling observed in equation (3). The linear saturation at larger values of  $s'_b$  occurs when a single additional mutation is sufficient to ensure fixation. However, unlike in small populations, these successful modifiers still arise on anomalously fit genetic backgrounds ( $x \approx \sqrt{x_c s_b}$ ), so they are able to hitchhike to higher initial frequencies and increase their overall probability of producing an additional mutation. Both examples illustrate that indirect selection acts over a limited horizon, which will be important when considering extensions to more complex fitness landscapes below.

The strength of indirect selection can also be quantified by comparing our expressions with the corresponding fixation probability of a ‘first-order’ mutation,

$$\log \tilde{p}_{\text{fix}}(s) \approx \frac{x_c s}{v}, \quad (5)$$

which has been studied in previous work<sup>42,43,50</sup> (Methods). Comparing this expression with equation (3) shows that even fractional changes in the selection strength ( $\Delta s_b/s_b \approx \sqrt{s_b/x_c} \lesssim 1$ ) can generate fixation probabilities as large as a typical beneficial mutation ( $s \approx s_b$ ). This contrasts with the behaviour observed for mutation-rate modifiers in equation (4), where the mutation rate must increase by several orders



**Fig. 3 | Trade-offs between direct and indirect selection.** **a,b,** Fixation probability of a selection-strength modifier with a direct cost (**a**) or benefit (**b**). Symbols denote the results of forward-time simulations for  $s_b = 10^{-2}$ ,  $U_b = U'_b = 10^{-5}$  and  $N = 10^8$ , while solid lines denote our theoretical predictions (Supplementary Section 8). **c,** Phase diagram illustrating the transition between favoured (green) and unfavoured (blue) modifiers. Symbols denote results of forward time simulations for the selection-strength modifiers in **a** and **b**;

x coordinates show the measured fixation probabilities in the absence of a direct cost or benefit,  $\tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b))$ ; the y coordinates show the predicted fixation probability of a first-order mutation,  $\tilde{p}_{\text{fix}}(s_m)$ , where  $s_m$  is inferred from the x intercept in **a** and **b** ( $\tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b), s_m) = 1$ ). Solid lines show our theoretical predictions (Supplementary Section 8), which exhibit large deviations from the additive expectation ( $\gamma = 1$ , dashed line).

of magnitude [ $\log(\Delta U_b/U_b) \approx s_b^2/v \gg 1$ ] to achieve the same effect. This shows that larger populations can more efficiently select on changes to the fitness benefits of future mutations, compared with the overall rate at which they occur.

The fixation probability of a modifier that changes the mutation rate and selection strength at the same time can be understood using these basic building blocks. We find that the dominant contributions can be expressed as a linear combination of equations (3) and (4),

$$\log \tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b)) \approx \underbrace{\log \tilde{p}_{\text{fix}}(s_b \rightarrow s'_b)}_{\text{change in selection strength}} + \alpha \times \underbrace{\log \tilde{p}_{\text{fix}}(U_b \rightarrow U'_b)}_{\text{change in mutation rate}}, \quad (6a)$$

where the weighting factor  $\alpha$  is given by

$$\alpha \approx \max \left\{ \left( s_b/s'_b \right)^2, s_b/x_c \right\}. \quad (6b)$$

The presence of this additional weighting factor implies that mutation-rate and selection-strength modifiers do not additively combine. Instead, equation (6) shows that increases in the average fitness benefit ( $s'_b > s_b$ ) will temper selection on the mutation rate ( $\alpha < 1$ ), while decreases in  $s'_b$  will amplify it ( $\alpha > 1$ ). This non-additivity arises because larger selection-strength modifiers lower the number of mutations required to fix by a factor of  $(s'_b/s_b)^2$ , which diminishes the compounding effects of the altered mutation rate. These differences can be large and can alter the overall sign of selection on the modifier (Fig. 2d). Moreover, as the individual terms in equation (6) depend on the underlying parameters in different ways, the sign of selection can also vary as a function of the population size and the basal mutation rate<sup>26,28</sup>. Together, these examples illustrate how selection-strength modifiers can lead to qualitatively different behaviour than expected for mutation-rate changes alone and that even modest shifts in  $s_b$  can frequently overpower order-of-magnitude differences in  $U_b$ .

### Trade-offs between direct and indirect selection

We are now in a position to understand how natural selection balances the short-term costs and benefits of a mutation with its longer-term impact on evolvability. Across a broad range of parameters, we find that the fixation probability can be naturally decomposed into contributions from direct and indirect selection,

$$\log \tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b), s_m) \approx \underbrace{\log \tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b))}_{\text{indirect selection}} + \gamma \times \underbrace{\log \tilde{p}_{\text{fix}}(s_m)}_{\text{direct selection}}, \quad (7a)$$

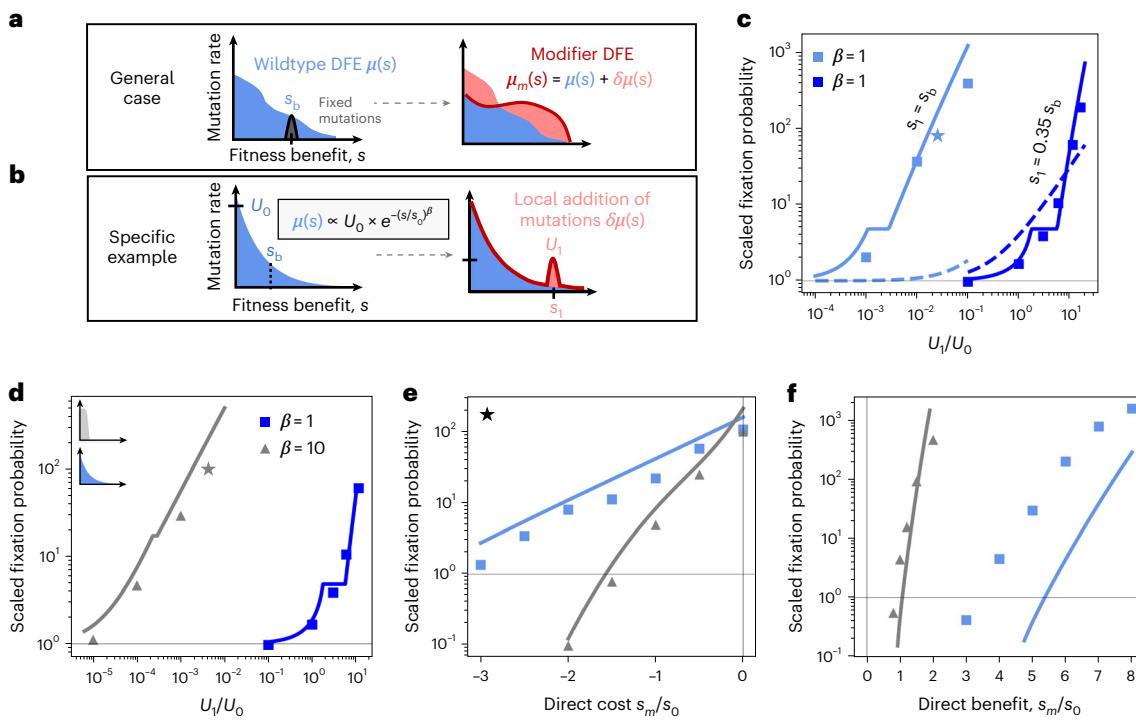
where  $\tilde{p}_{\text{fix}}(s_m)$  and  $\tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b))$  are given by equations (5) and (6) above, and  $\gamma$  is an additional weighting factor satisfying

$$\gamma \approx (s_b/s'_b) \left[ 1 - (x_c s_b/v)^{-1} \log(U'_b/U_b) \right]. \quad (7b)$$

As above, the presence of this additional weighting factor implies that differences in the DFE will generally modulate the effects of first-order selection on fitness. The direction of this effect depends on whether  $\gamma$  is greater or smaller than 1.

We find that modifiers that would be strongly favoured in the absence of a direct cost or benefit [ $\tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b)) \gg 1$ ] lead to a weighting factor  $\gamma \lesssim 1$ , which reduces the relative contribution from  $\tilde{p}_{\text{fix}}(s_m)$ . As a result, these evolvability-enhancing mutations can remain positively selected even when they impose a large direct cost on fitness (for example, larger than the size of a typical driver mutation; Fig. 3a,c). This contrasts with the classical behaviour observed for discrete selective sweeps, where direct costs larger than a typical driver mutation will generally prevent a modifier from fixing (Supplementary Section 2). These results imply that larger populations are better able to sacrifice short-term fitness for longer-term gains in evolvability.

The opposite behaviour occurs when short-term fitness benefits are linked to long-term reductions in evolvability. In this case, modifiers that would be strongly disfavoured on their own [ $\tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b)) \ll 1$ ] will generally amplify the relative contribution of a direct fitness benefit (Fig. 3b,c). A striking example of this effect occurs in the extreme case of an evolutionary ‘dead end’, where further beneficial mutations are not available. Generalizing equation (1) to account for this case, we find that direct benefits as small as  $\approx 0.4x_c$  are sufficient to cause an evolutionary dead end to be positively selected ( $\tilde{p}_{\text{fix}} \gg 1$ ), even though they drive the long-term rate of adaptation to 0 when they fix (Fig. 3b). This critical direct benefit is often larger than a single driver mutation ( $s_b$ ), but it is also smaller than the total fitness variation maintained within the population ( $x_c$ ) and only weakly increases with  $NU_b$ . These examples illustrate how the evolutionary foresight of natural selection can be highly asymmetric: larger populations can still greedily select for



**Fig. 4 | Indirect selection on a continuous distribution of fitness effects.** **a**, An evolvability modifier shifts the DFE to a new shape  $\mu_m(s) \equiv \mu(s) + \delta\mu(s)$  (right). The strength of indirect selection depends on how the perturbation  $\delta\mu(s)$  compares to the size of a typical fixed mutation (left). **b**, Specific example in **c–f**. The wildtype DFE is a stretched exponential with shape parameter  $\beta$ , scale parameter  $s_0$  and overall mutation rate  $U_0$  (left). The modifier adds a localized perturbation,  $\delta\mu(s) \approx U_1 \delta(s - s_i)$  (right). **c**, Scaled fixation probability of the modifier in **b** for an exponential DFE ( $\beta = 1$ ) for two different values of  $s_i$ . Symbols denote the results of forward-time simulations for  $s_0 = 10^{-2}$ ,  $U_0 = 10^{-5}$  and  $N = 10^8$ , where the typical

fixed mutation has  $s_b \approx 8.5 \times s_0$  (Methods). Solid lines denote our theoretical predictions (Supplementary Section 8), while dashed lines indicate the null expectation in the absence of clonal interference. **d**, Fixation probability of the modifier for an exponential DFE ( $\beta = 1$ ) and a distribution with a shorter tail ( $\beta = 10$ ); blue and grey curves in the inset illustrate the differences between these background distributions. In both cases,  $s_i \approx 0.03$ , while other parameters are the same as in **c**. **e**, Fixation probability of the modifier with a direct fitness cost. Base parameters are indicated by the stars in **c** and **d**. **f**, Analogous version of **e** for an evolutionary dead end ( $\mu(s) \rightarrow 0$ ) with a direct fitness benefit.

mutations that lower their rate of adaptation, even while they are better able to endure short-term fitness costs to realize long-term evolutionary benefits.

#### Continuous distributions of fitness effects

We have so far focused on a simplified model of the fitness landscape, where all new mutations confer the same characteristic fitness benefit. However, most organisms produce mutations with a range of different fitness effects. A more realistic evolvability modifier will therefore correspond to a continuous perturbation of the DFE,  $\delta\mu(s) = \mu_m(s) - \mu(s)$ , representing the addition or subtraction of mutations with a range of costs or benefits (Fig. 4a). How does indirect selection on these more general differences in the DFE relate to the idealized selection-strength and mutation-rate axes above?

Focusing first on beneficial mutations, we can extend our solution of equation (1) to a large class of wildtype DFEs that have been studied in previous work. In these settings, the distribution of fixed mutations is strongly peaked around a characteristic fitness benefit  $s_b(\mu(s), N)$  (with a corresponding mutation rate  $U_b(\mu(s), N)$ ), even when the underlying DFE has a broader shape<sup>42,43,51,52</sup> (Fig. 4a and Supplementary Section 5.1). By solving equation (1) in this limit (Methods), we find that indirect selection on a general shift  $\mu(s) \rightarrow \mu(s) + \delta\mu(s)$  strongly depends on how the perturbation  $\delta\mu(s)$  relates to the wildtype values of  $U_b$  and  $s_b$ .

For small changes to the DFE, the fixation probability of the modifier initially grows as

$$\log \tilde{p}_{\text{fix}}(\mu(s) \rightarrow \mu(s) + \delta\mu(s)) \approx \frac{x_c}{s_b} \times \int_0^{x_c} \frac{\delta\mu(s)}{U_b} e^{\frac{x_c(s-s_b)}{v}} ds, \quad (8)$$

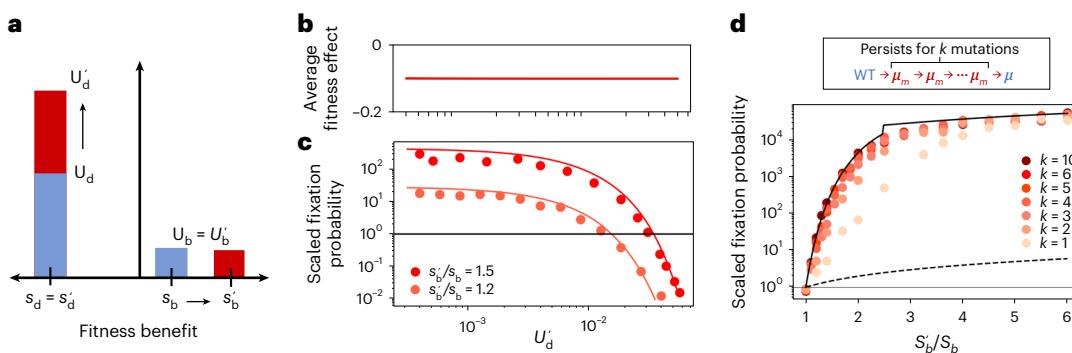
where  $v$  and  $x_c$  again denote the rate of adaptation and nose of the wildtype population (Fig. 1c), which can be calculated from the values of  $N$ ,  $U_b$  and  $s_b$  (Methods). This expression shows how natural selection weighs the addition or subtraction of mutations with different fitness benefits. The critical fitness scale is set by the size of a fixed mutation: when  $s > s_b$ , even small increases in the net mutation rate ( $\int_{s_b}^{\infty} \delta\mu(s) ds \ll U_b$ ) can generate large changes in the fixation probability (Fig. 4c). By contrast, fitness benefits smaller than  $s_b$  require many multiples of  $U_b$  to have the same effect (Fig. 4c). Moreover, as the values of  $s_b$  and  $U_b$  both emerge from the competition between linked mutations, the location of this transition can vary with the size of the population and the shape of the wildtype DFE (Fig. 4d).

For larger changes to the DFE, the mutations that fix in a successful modifier lineage will tend to be concentrated around their own characteristic benefit  $s'_b \neq s_b$ . We find that this new fitness scale is determined by a generalization of the integral in equation (8) and depends on the shape of  $\delta\mu(s)$  as well as the wildtype parameters  $s_b$  and  $U_b$  (Methods). In this case, the fate of the modifier can be predicted from our single-s theory in equation (6),

$$\log \tilde{p}_{\text{fix}}(\mu(s) \rightarrow \mu(s) + \delta\mu(s)) \approx \log \tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b)), \quad (9)$$

where  $U'_b$  denotes the corresponding mutation rate for mutations that are sufficiently close to  $s'_b$  (Fig. 4c,d and Methods). Similar results also apply for modifiers with direct costs or benefits, allowing us to extend our results in Fig. 3 to these more general scenarios as well (Fig. 4e,f).

The equivalence principle in equation (9) shows that indirect selection on an arbitrary shift in the DFE can be understood as a combination



**Fig. 5 | Incorporating deleterious mutations and modifiers with finite mutational horizons.** **a**, Example of an evolvability-enhancing mutation ( $s_b \rightarrow s_b'$ ) that decreases mutational robustness by increasing the deleterious mutation rate ( $U_d \rightarrow U_d'$ ). **b**, As deleterious mutations outnumber beneficial variants by several orders of magnitude ( $U_d \gg U_b$ ), the change in the average fitness effect,  $\int s \times \delta\mu(s) ds$ , is dominated by the deleterious portion of the DFE. **c**, Scaled fixation probability of the modifier as a function of the new deleterious mutation rate  $U_d'$ . Symbols denote the results of forward-time simulations for  $s_b = 4 \times 10^{-2}$ ,  $U_b = 10^{-6}$ ,  $s_d = 10^{-1}$ ,  $U_d = 4 \times 10^{-4}$  and  $N = 10^8$ . Solid lines denote our theoretical predictions, where deleterious mutations behave like an effective direct cost  $s_m^{\text{eff}} \approx U_d - U_d'$ . This example illustrates that changes in the deleterious

portion of the DFE have a minor impact on the fixation probability for many biologically relevant mutation rates ( $U_d' \lesssim 10^{-3}$ ). **d**, Another generalization of Fig. 2b, where modifier individuals revert back to the wildtype DFE after  $K$  mutational steps (Supplementary Section 7.2). Symbols denote the results of forward-time simulations for  $N = 10^8$ ,  $s_b = 10^{-2}$  and  $U_b = 10^{-5}$ , while the solid line denotes our theoretical predictions for the minimal modifier model in Fig. 1b. Consistent with our heuristic analysis, the minimal modifier model ( $K \approx \infty$ ) remains highly accurate even for moderate values of  $K$  (for example, 2–3) and for as little as a single mutation when  $s_b'$  is large. In all cases, the fixation probabilities are much larger than expected in the absence of clonal interference (dashed lines).

of the mutation-rate and selection-strength axes in Fig. 2. However, the relevant parameters in this mapping will not coincide with the nominal mean and height of the DFE. Instead, due to the exponential weighting of mutations with  $s \geq s_b$ , otherwise subtle additions to  $\mu(s)$  can be strongly favoured by natural selection, even when they have a negligible impact on the overall mean or mutation rate (Fig. 4c–e). Conversely, large reductions in these global parameters will be nearly invisible to natural selection unless they also deplete mutations near  $s_b$ . This sensitivity to local changes could help explain previous experimental observations in *Escherichia coli* (Fig. 1a), where the potentiation of just a few beneficial genes was sufficient to overcome a large direct fitness cost<sup>8</sup>.

## Discussion

Our results provide a framework for understanding how natural selection balances the short-term costs and benefits of a new mutation with its longer-term impact on evolvability. We have shown that when beneficial mutations are common, the competition between linked mutations can dramatically enhance selection for subtle differences in the mutational neighbourhood, leading to large deviations from the linear scaling predicted by classical evolutionary models<sup>29</sup>. These results suggest that indirect selection could play a previously unappreciated role in driving the success of genetic variants in large microbial populations, from laboratory evolution experiments<sup>8,10,13,33,53</sup> to natural systems such as cancer<sup>14,15</sup>, influenza<sup>34</sup> or SARS-CoV-2<sup>37,54</sup>. This could have important consequences for evolutionary forecasting<sup>35,55–57</sup>, as it implies that the direct fitness effects of such variants might fail to explain their long-term evolutionary success.

Our theory indicates that it could be difficult to detect these evolvability differences using traditional metrics such as the rate of adaptation<sup>8,44,58</sup> or the substitution rate<sup>53</sup>, as the associated changes in these observables are not always large (Supplementary Section 4.1.7). Previously documented cases such as Fig. 1a might therefore only represent a fraction of the selectable variation in evolvability. Our results suggest that future efforts could instead focus on mapping the aggregate changes to the DFE (Fig. 1b), for example, using barcoded lineage tracking<sup>9,59,60</sup> or mutation trap experiments<sup>61</sup>. However, we have also shown that the important changes in this distribution will often occur in its high-fitness tail and are poorly captured by existing heuristics such as the mean mutational effect (Fig. 5a–c; refs. 29,45). They can also depend on external factors such as the population size and the

overall mutation rate. This suggests that the evolvability benefits of a mutation should not be viewed as an intrinsic property of the genotype but rather a collective effect that can vary across populations or within the same population over time. Our theory provides a framework for predicting where these evolutionarily important differences will occur.

While we have primarily focused on the supply of beneficial mutations, our results can also be extended to account for changes in the supply of strongly deleterious mutations (for example, those that are rapidly purged by selection), which behave like an effective direct cost,  $s_m^{\text{eff}} \approx -\int_{-\infty}^{-v/x_c} \delta\mu(s) ds$  (Fig. 5a–c, Extended Data Fig. 2 and Supplementary Section 6). This mapping to Fig. 3 reveals how adapting populations balance the trade-offs (or synergies) between robustness and evolvability<sup>62–64</sup>. On the one hand, it implies that more rapidly adapting populations will generally be more willing to sacrifice short-term robustness for longer-term gains in evolvability (Fig. 5a–c), particularly for the mutation rates that are common in many bacteria<sup>65</sup>. However, it also suggests that at higher mutation rates, large robustness gains could still be preferred even if they eliminate all opportunities for future adaptation (Fig. 3b). These results shed light on when ‘flatter’ or ‘steeper’ regions of the fitness landscape will be favoured by natural selection<sup>22,30,62,66,67</sup>.

Our minimal model also assumed that the indirect benefits of the modifier remain fixed as it competes for dominance in the population. In reality, epistatic interactions could cause these benefits to attenuate—or even reverse—as the modifier acquires further mutations. Our heuristic analysis suggests that our current results will continue to hold as long as the effective parameters in equation (9) remain constant over a typical fixation time (roughly  $v/x_c$  generations, or  $x_c/s_b$  additional mutations). This timescale is often modest in practice (Methods), allowing our minimal model to capture a broader range of epistatic scenarios than its idealized nature might originally suggest (Fig. 5d, Extended Data Figs. 3 and 4 and Supplementary Section 7). Further extensions of this framework to allow for more rapidly varying distributions of fitness effects ('macroscopic epistasis'<sup>68</sup>) could be useful for understanding how large populations navigate complex fitness landscapes<sup>69</sup>.

Finally, while we have focused on the asexual dynamics common in laboratory experiments<sup>8</sup> and somatic evolution<sup>14,15</sup>, natural microbial populations often show some degree of recombination<sup>70</sup>. Widespread recombination will alter our predictions by decoupling the modifier locus from the future mutations that it produces<sup>24,71</sup>. Previous work

suggests that some of our results may still apply on short genomic distances that remain tightly linked over the characteristic fixation time ( $I \approx \frac{1}{r} \times \frac{v}{x_c}$ ) (refs. 72–74). However, modifiers can also benefit from transient linkage to mutations outside these asexual linkage blocks<sup>75</sup>, and the fitness effects of the mutations could change when decoupled from the background that produced them<sup>69</sup>. Understanding the interplay between these forces—and how they generalize to fluctuating environments<sup>62,76</sup>—will be critical for understanding how indirect selection acts in other contexts.

## Methods

### Model and notation

Our basic model considers an asexual population of  $N \gg 1$  individuals whose genomes are composed of  $L \gg 1$  bi-allelic loci. The genotype of each individual is represented by a binary vector  $\vec{g} \in \{0, 1\}^L$ . We assume a constant environment, where the (log) fitness of each genotype is determined by an arbitrary function  $X(\vec{g})$  (the ‘fitness landscape’). Mutations from genotype  $\vec{g}$  to  $\vec{g}'$  occur at a total rate  $M(\vec{g} \rightarrow \vec{g}')$  per individual per generation (the ‘mutational network’).

Modifier mutations can be expressed in this framework by designating an arbitrary site  $m$  as the modifier locus and recalculating  $X(\vec{g})$  and  $M(\vec{g} \rightarrow \vec{g}')$  for the modifier ( $g_m = 1$ ) and wildtype ( $g_m = 0$ ) alleles separately. This yields an analogous pair of functions  $X_m(\vec{g})$  and  $M_m(\vec{g} \rightarrow \vec{g}')$  that represent the fitness landscape and mutational network of the modifier lineage, as well as a direct cost or benefit  $s_m$  that captures the immediate fitness effect of the modifier allele in the genetic background where it arises. We call such a mutation an ‘evolvability modifier’ if it alters either the fitness landscape [ $X_m(\cdot) \neq X(\cdot) + s_m$ ], the mutational network [ $M_m(\cdot, \cdot) \neq M(\cdot, \cdot)$ ] or some combination of the two. This definition is consistent with the operational notion of evolvability in refs. 8,45 and encompasses classical examples such as mutator alleles<sup>4,26,28</sup> as well as shifts in the fitness landscape due to epistatic interactions with the modifier locus<sup>8,45</sup>. However, we note that it does not capture other notions of evolvability, for example, when the benefits of the modifier are only revealed after a shift in environmental conditions<sup>62</sup> or modifiers of other evolutionary parameters such as the recombination rate<sup>49</sup>, migration rate<sup>20</sup> or pleiotropy<sup>77</sup>. Extensions to such scenarios remain an interesting avenue for future work.

Given the definitions above, the impact of an evolvability modifier can be equivalently described by how it changes the local DFE, which is defined by

$$\mu(s|\vec{g}) = \sum_{\vec{g}'} M(\vec{g} \rightarrow \vec{g}') \times \delta(s + X(\vec{g}) - X(\vec{g}')), \quad (10)$$

where  $\delta(\cdot)$  is the Dirac delta function. This distribution is normalized so that  $\mu(s|\vec{g})ds$  represents the total rate that an individual with genotype  $\vec{g}$  produces mutations with fitness effects in the range  $s \pm ds/2$ . This mutational neighbourhood can in principle vary with the genetic background  $\vec{g}$  due to epistatic interactions in the fitness landscape  $X(\vec{g})$  and/or the mutational network  $M(\vec{g}, \vec{g}')$ . For most of this work, we will assume that this background dependence can be captured by the minimal modifier model,

$$\mu(s|\vec{g}) \approx \begin{cases} \mu(s) & \text{if } g_m = 0, \\ \mu_m(s) & \text{if } g_m = 1, \end{cases} \quad (11)$$

in which the wildtype and modifier DFEs can differ from each other but remain approximately constant over the relevant genetic distance scales. This minimal model can be viewed as the lowest-order term in a more general expansion in the genotype dependence of the DFE:

$$\mu(s|\vec{g}) \approx \mu(s) + \sum_{\ell} \delta\mu_{\ell}(s) \times g_{\ell} + \sum_{\ell < \ell'} \delta\mu_{\ell, \ell'}(s) \times g_{\ell}g_{\ell'} + \dots, \quad (12)$$

with a non-zero  $\delta\mu_{\ell}(s)$  term at the modifier locus ( $\ell = m$ ) and all other  $\delta\mu_{\ell, \ell'}$  terms vanishing (Extended Data Fig. 1 and Supplementary Section 3.1). As above, equation (11) allows us to consider both classical mutator alleles (where  $\mu(s)$  increases by a constant factor) as well as shifts in the underlying fitness landscape due to epistatic interactions with the modifier locus. It can also be viewed as a generalization of the ‘survival of the flattest’ models in refs. 22,30 that incorporates beneficial mutations.

We emphasize that our results do not require equation (11) to hold across the entire fitness landscape but only within a smaller region that is explored before the modifier either fixes or goes extinct. We determine the size of this local neighbourhood in Supplementary Section 7 and find that it is often modest, corresponding to just a handful of mutational steps for many empirically relevant parameter values (Fig. 5b and Extended Data Fig. 3). We also show that the assumption in equation (11) is most sensitive to a narrow range of beneficial fitness effects, so that substantial deviations in other parts of the DFE can still have a negligible impact on the fate of the modifier mutation (Figs. 4 and 5b and Supplementary Section 7). This more general definition of equation (11) applies for a broad range of epistatic fitness landscapes, as well as other relevant scenarios such as aneuploidy that are difficult to express in the traditional landscape picture; we consider several concrete examples in Supplementary Sections 3.1 and 7.

To quantify the net selection pressures on a given modifier, we considered its fixation probability,  $p_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m)$ , when arising in a steady-state population of adapting wildtype individuals (Fig. 1c). We use the term direct (or ‘first-order’) selection to refer to cases where  $\mu_m(s) = \mu(s)$ , while indirect (or ‘second-order’) selection refers to cases where  $\mu_m(s) \neq \mu(s)$ .

### Derivation of the conditional fixation probability

To derive the conditional fixation probability in equation (1), we use a key approximation from previous theoretical models of clonal interference<sup>26,42,43,50</sup> and assume that the fate of a successful modifier is determined while it is still at a low frequency in the population. This implies that (1) the mean fitness of the population will remain close to the wildtype trajectory [ $\partial_t \bar{X}(t) \approx v(\mu(s), N) \equiv v$ ] while the fate of the modifier is being determined and (2) that different individuals in the modifier lineage will either fix or go extinct independently. We discuss the conditions of validity of this approximation in Supplementary Section 3.2 and show that it is satisfied for a broad range of modifier alleles (with the exception of the strongest direct benefits, which we consider separately below).

When this separation of timescales holds, the conditional fixation probability can be calculated by extending the branching process formalism in refs. 26,42,43,50. Briefly, if  $p_e(x)$  denotes the extinction probability of a modifier lineage founded with initial relative fitness  $x$ , one can obtain a recursion relation for  $p_e(x)$  by averaging over the offspring that the founding individual produces in the next generation. The resulting recursion relation is given by

$$p_e(x) = \left\langle p_e(x - v)^{n_c} \prod_s p_e(x - v + s)^{n_m(s)} \right\rangle, \quad (13)$$

where  $n_c \sim \text{Poisson}[(1+x)(1 - \int \mu_m(s)ds)]$  denotes the number of clonal offspring of the founding individual, and  $n_m(s) \sim \text{Poisson}[(1+x)\mu_m(s)ds]$  denotes the number of mutant offspring with fitness effect  $s$ . Equating the conditional fixation probability with the non-extinction probability of the branching process, we obtain equation (1) in the main text after expanding equation (13) to lowest order in  $x$ ,  $\mu_m(s)$ ,  $\sqrt{v}$  and  $w_m(x) \equiv 1 - p_e(x)$ . A more detailed derivation of equation (1) using a Langevin framework is outlined in Supplementary Section 3.2.

When  $\mu_m(\cdot) \neq \mu(\cdot)$ , the conditional fixation probability of an evolvability modifier will differ from that of a lineage under direct (or ‘first-order’) selection because different DFEs contribute to the

mutation and mean fitness terms in equation (1). However, as the wildtype DFE enters only through the rate of adaptation  $\nu(\mu(s), N)$ , the conditional fixation probability of a modifier can still be mapped to a direct selection scenario, in which the population is fixed for the modifier allele but has a different population size  $N^*$  that satisfies

$$\nu(\mu_m(s), N^*) = \nu(\mu(s), N). \quad (14)$$

This formal equivalence hints at a deeper relationship between direct and indirect selection, which we exploit in more detail below.

### Solution for the simplest fitness landscape

To calculate the fixation probability of a modifier in the simplest fitness landscape in Fig. 2a, we sought an approximate analytical solution of equations (1) and (2) that applies for empirically relevant parameter regimes where  $s_b \gg U_b \geq 1/N$ . We outline the key steps below, while a detailed derivation is provided in Supplementary Section 4.

Following previous work<sup>26,42,43,50</sup>, we found that the solution for the conditional fixation probability  $w_m(x)$  can be decomposed into two characteristic regions depending on the size of the initial relative fitness  $x$ . For large values of  $x$ , the mutation term in equation (1) is negligible compared with the other three terms, and the solution is well approximated by

$$w_m(x) \approx \frac{2x_{cm} e^{(x^2 - x_{cm}^2)/2\nu}}{1 + (x_{cm}/x) e^{(x^2 - x_{cm}^2)/2\nu}}, \quad (15)$$

where  $x_{cm}$  is a constant of integration that must be determined self-consistently below. In our parameter regime of interest, this ‘shoulder solution’ shows a sharp transition near  $x_{cm}$ , switching from a linear scaling at high relative fitnesses [ $w_m(x) \approx 2x$ ] to a more rapid decay when  $x < x_{cm}$  (Fig. 1). This implies that  $x_{cm}$  can be viewed as a ‘clonal interference threshold’ for the modifier lineage: modifiers with initial relative fitness greater than  $x_{cm}$  will fix if they survive genetic drift, while modifiers with  $x < x_{cm}$  will be strongly impacted by competition with the wildtype population.

For smaller values of  $x$ , the mutation term is no longer negligible, but the genetic drift term is now sub-dominant, and equation (1) reduces to the linear form,

$$0 \approx \underbrace{x \times w_m(x)}_{\text{selection}} + \underbrace{U'_b [w_m(x + s'_b) - w_m(x)]}_{\text{mutation}} - \underbrace{\nu \times \partial_x w_m(x)}_{\text{competition w/wt}}. \quad (16)$$

To solve this equation, it is useful to re-express it in the integral form,

$$w_m(x) = e^{\frac{(x-U'_b)^2}{2\nu}} \int_{-\infty}^{x+s'_b} \frac{U'_b}{\nu} \times e^{-\frac{(y-U'_b)^2}{2\nu}} w_m(y) dy. \quad (17)$$

In Supplementary Section 3.2, we show that this recursion relation has a natural interpretation as an average over the possible mutant offspring that are produced by the founding modifier clone. In our parameter regime of interest, the right-hand side of equation (17) is usually dominated by relative fitnesses  $y$  that are greater than or equal to those on the left-hand side. This implies that equation (17) is naturally telescoping: we can substitute equation (15) into the right-hand side of equation (17) and recursively extend  $w_m(x)$  to progressively lower fitness values. We carry out this procedure in Supplementary Section 4 to obtain a solution for  $w_m(x)$  that is valid across the full range of relative fitness values.

The constant of integration in equation (15) can be determined by the requirement that equations (15) and (17) should match in the overlap region immediately below  $x_{cm}$  where both approximations are valid. This constraint allows us to solve for  $x_{cm}$  as a function of  $s'_b$ ,  $U'_b$ , and the wildtype’s rate of adaptation  $\nu$ . For modifiers with small or moderate indirect effects, this solution takes on a particularly simple form,

$$x_{cm} \approx \frac{\nu}{s'_b} \log \left( \frac{s'_b}{U'_b} \right) + \frac{s'_b}{2}, \quad (18)$$

which applies when  $x_{cm} \gtrsim s'_b$ .

To calculate the net fixation probability of the modifier in equation (2), we must average the solution for  $w_m(x)$  over the distribution of parental fitnesses  $f(x)$ , which has been characterized in previous work<sup>43,50</sup>. In our parameter regime of interest, this fitness distribution is well approximated by a truncated Gaussian profile,

$$f(x) \approx \begin{cases} \frac{1}{\sqrt{2\pi\nu}} e^{-\frac{x^2}{2\nu}} & x \leq x_c, \\ 0 & x > x_c. \end{cases} \quad (19a)$$

where the maximum fitness  $x_c$  coincides with the interference threshold in equation (18) when  $s'_b = s_b$  and  $U'_b = U_b$ ,

$$x_c \approx \frac{\nu}{s_b} \log \left( \frac{s_b}{U_b} \right) + \frac{s_b}{2}, \quad (19b)$$

and the variance coincides with the rate of adaptation,

$$\nu \approx \frac{2s_b^2 \log(Ns_b)}{\log^2(s_b/U_b)}. \quad (19c)$$

Substituting these expressions into equation (2) and integrating over the parental fitness  $x$ , we obtain an approximate solution for the scaled fixation probability,

$$\log \bar{p}_{fix}(\mu(s) \rightarrow \mu_m(s), s_m) \approx \begin{cases} \frac{x_c^2}{2\nu} - \frac{x_{cm}^2}{2\nu} + \frac{(x_{cm}-s'_b)s_m}{\nu} & s_m < 0, \\ \frac{x_c^2}{2\nu} - \frac{x_{cm}^2}{2\nu} + \frac{x_{cm}s_m}{\nu} & s_m > 0, \end{cases} \quad (20)$$

which includes only the leading order terms from the full expressions in Supplementary Section 4.1. Equations (3)–(7) in the main text are all obtained by considering limiting cases of this basic expression, using equations (18) and (19) to substitute for  $x_{cm}$ ,  $x_c$  and  $\nu$ , and taking the limit that  $x_c \gg s_b \gg \sqrt{\nu}$  (Supplementary Section 4.1). Analogous expressions for more strongly selected modifiers, where equation (18) breaks down, are derived in Supplementary Section 4.2.

### Fixation probability of an evolutionary dead end

Modifiers with sufficiently large indirect costs but strong direct benefits cannot be captured by the branching process approximation in equation (1), as their long-term success requires them to grow to a sufficient size where they start to influence the adaptation of the wildtype. The most extreme example of this behaviour occurs for an evolutionary ‘dead end’ (illustrated by the  $s'_b = 0$  line in Fig. 3b), which is unable to produce further beneficial mutations.

We can extend our solution to this case by explicitly considering the dynamics of the modifier lineage over time (Supplementary Section 4.1.6). At short times, these dynamics will still be well approximated by the branching process model above. A dead-end modifier that arises with an initial relative fitness  $x$  will found a clone that grows as

$$f_m(t) \approx \begin{cases} \frac{e^{\nu t - \frac{xt^2}{2}}}{2Nx} & \text{with probability } 2x, \\ 0 & \text{else,} \end{cases} \quad (21)$$

which accounts for the stochastic effects of genetic drift, as well as the constant adaptation of the background population (Supplementary Section 3.2). If the initial fitness of the modifier is small enough that

$f_m(t)$  remains small at later times, then the background population will always overtake the modifier and eventually drive it to extinction. For larger values of  $x$ , the modifier will eventually grow to macroscopic frequencies (for example, 10%), and equation (21) will start to break down. During this time, the background population will have increased in fitness by  $vt^*$ , where  $t^*$  is defined by  $f_m(t^*) \approx 0.1$ . In Supplementary Section 4.1.6, we show that once the modifier reaches this intermediate size, it will transition to >90% frequency extremely rapidly, while the background fitness distribution is effectively frozen in place. If the nose of the fitness distribution has caught up to the modifier in this time ( $x_c + vt^* > x$ ), then the dead-end lineage is still destined for extinction once these fitter individuals start to expand. However, if  $x \geq x_c + vt^*$ , then the modifier clone will rapidly proceed to fixation before the background population can catch up. This analysis suggests that the conditional fixation probability of a dead-end modifier can be approximated by

$$w_m(x) \approx \begin{cases} 2x & \text{if } x > x_{cm} \\ 0 & \text{if } x < x_{cm}, \end{cases} \quad (22a)$$

with an effective interference threshold,

$$x_{cm} \approx x_c + vt^* \approx \sqrt{2} \times x_c. \quad (22b)$$

A more detailed derivation is presented in Supplementary Section 4.1.6, which allows us to calculate the next-order correction to  $x_{cm}$  that we use for our numerical comparisons in Fig. 3. Substituting this expression into equation (2) yields an analogous formula for the overall fixation probability of a dead-end modifier,

$$\log \tilde{p}_{\text{fix}}(\mu(s) \rightarrow 0, s_m) \approx \log \int_0^{\frac{x_c(s_m - x_{cm} + x_c)}{v}} e^{-\frac{vu^2}{2x_c^2}} du, \quad (23)$$

which shows that direct fitness benefits as small as  $s_m \approx x_{cm} - x_c \approx 0.4x_c$  will be sufficient to make them strongly favoured by natural selection.

To account for these nonlinear feedbacks more generally, we use equation (23) in place of equation (20) whenever the interference threshold in equation (18) is larger than upper bound implied by equation (22).

### Continuous distributions of fitness effects

To extend our results to continuous distributions of fitness effects, we sought an analogous solution of equation (1) for scenarios where the wildtype DFE can be analysed using the approaches developed in refs. 42,43,50 (Supplementary Section 5). We find that the conditional fixation probability can again be approximated by a generalization of equations (15) and (17). The large  $x$  solution is the same as equation (15), but the interference threshold ( $x_{cm}$ ) is now determined by a more complex condition,

$$1 \approx \frac{\sqrt{2\pi v}}{2x_{cm}v e^{-\frac{x_{cm}^2}{2v}}} \int \mu_m(s) \times p_{\text{fix}}(s|\mu_m(s), N^*) ds, \quad (24)$$

where  $p_{\text{fix}}(s|\mu_m(s), N^*)$  is the fixation probability of an ordinary beneficial mutation in a population with a background DFE  $\mu_m(s)$  and a population size  $N^*$  defined by equation (14). Writing  $\mu_m(s) \equiv \mu(s) + \delta\mu(s)$ , we can distinguish between two broad regimes depending on whether  $\mu(s)$  or  $\delta\mu(s)$  provides the dominant contribution to the integral in equation (24).

When the integral in equation (24) is dominated by the contributions from  $\mu(s)$ , then  $x_{cm}$  will remain close to the wildtype interference threshold  $x_c$ . This allows us to calculate  $x_{cm}$  and  $p_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m)$  perturbatively, by treating the  $\delta\mu(s)$  term as a small parameter (Supplementary Section 5.2). Writing  $x_{cm} = x_c + \delta x_c$ , we can perform

a Taylor expansion in  $\delta\mu(s)$  to obtain an approximate analytical expression for  $\delta x_c$ :

$$\delta x_c \approx -\frac{v}{s_b} \int \frac{\delta\mu(s)}{U_b} \frac{\tilde{p}_{\text{fix}}(s|\mu(s), N)}{\tilde{p}_{\text{fix}}(s_b|\mu(s), N)} ds, \quad (25)$$

where  $s_b \equiv s_b(\mu(s), N)$  and  $U_b \equiv U_b(\mu(s), N)$  are the effective parameters for the wildtype DFE, and  $\tilde{p}_{\text{fix}}(s|\mu(s), N)$  is the fixation probability of a first-order mutation from equation (5). We obtain the result in equation (8) by substituting equation (25) into equation (20) and using the expression for  $\tilde{p}_{\text{fix}}(s|\mu(s), N)$  from equation (5).

By contrast, when equation (24) is dominated by the contributions from  $\delta\mu(s)$ , the integral will often be strongly peaked around a characteristic value  $s'_b$ , which is defined by

$$s'_b \equiv \operatorname{argmax}_s \{\delta\mu(s) \times p_{\text{fix}}(s|\mu_m(s), N^*)\} \quad (26)$$

The corresponding mutation rate  $U'_b$  is then defined by taking a Laplace approximation of the integral,

$$\int \delta\mu(s) \times p_{\text{fix}}(s|\mu_m(s), N^*) ds \approx U'_b \times p_{\text{fix}}(s'_b|\mu_m(s), N^*), \quad (27)$$

which yields the single- $s$  mapping in equation (9). A more detailed derivation of these results, as well as explicit calculations of the effective parameters for different choices of  $\mu(s)$  and  $\mu_m(s)$ , can be found in Supplementary Section 5. In Supplementary Section 6, we show how these results can be extended to account for changes in the strongly deleterious portion of the DFE.

### Forward time simulations

We validated our theoretical predictions by comparing them to forward-time, Wright–Fisher-like simulations similar to those used by ref. 26. We begin with a population of  $N$  wildtype individuals at generation 0. In each generation, every individual  $i$  in the population produces

$$n_c(i) \sim \text{Poisson} \left( C(t) \times (1 + X_i - \bar{X}(t)) \times (1 - \int \mu(s) ds) \right) \quad (28a)$$

clonal offspring and

$$n_m(i) \sim \text{Poisson} \left( C(t) \times (1 + X_i - \bar{X}(t)) \times \int \mu(s) ds \right) \quad (28b)$$

mutant offspring, where  $X_i$  is the absolute fitness of the individual,  $\bar{X} = \sum_i X_i / \sum_i 1$  is the mean fitness of the population, and  $C(t) = N / \sum_i 1$  is a normalization constant that ensures that the expected population size in the next generation is equal to  $N$ . Each mutant offspring is assigned a new fitness value  $X_i + s$ , where  $s$  is randomly sampled from the normalized version of  $\mu(s)$ . After initialization, each simulation is allowed to ‘burn in’ for  $\Delta t = 2 \times 10^4$  generations so that it reaches a well-defined steady state.

To measure the fixation probability of a modifier lineage, we continue this basic algorithm while allowing individuals in the wildtype population to produce new modifier offspring at a per capita rate  $U_m$ . These modifier individuals reproduce according to an analogous version of equation (28), with  $\mu(s) \rightarrow \mu_m(s)$ . Reversions from modifier to wildtype genotypes are not allowed. Following the burn-in period, we record the number of generations that elapse until a modifier lineage fixes in the population. In the limit that  $U_m \rightarrow 0$ , this fixation time  $T$  is related to the fixation probability through the relation

$$\tilde{p}_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m) = \frac{1}{U_m \langle T \rangle}. \quad (29)$$

For the estimator in equation (29) to apply,  $U_m$  must be small enough that the total time to fixation is much larger than the predicted sweep

time of a successful modifier ( $\geq x_{cm}/v$ ). To ensure that this is true, we repeat this process of measuring the fixation time for a sequence of  $M = 60$  independent simulations with a sequence of modifier mutation rates  $U_{m,1}, U_{m,2}, \dots, U_{m,M}$ , yielding a sequence of fixation times  $T_1, T_2, \dots, T_m$ ; the mutation rate in simulation  $i$  is chosen based on the previous  $T_{i-1}$ ,

$$U_{m,i} \equiv \min \left\{ \frac{U_{m,i-1} T_{i-1}}{T^*}, 10^{-2} \right\}, \quad (30)$$

so that the predicted value of  $T_i$  is  $T^* \equiv 250 \times x_{cm}/v$  generations. The mutation rate is also capped at  $10^{-2}$  so that successful modifier lineages primarily compete against the background population while small. The sequence is started at  $U_{m,1}=1/T^*$  and allowed to ‘burn in’ for 10 iterations before the  $T_i$  are recorded. The fixation probability is calculated from the maximum likelihood estimator,

$$\bar{p}_{fix}(\mu(s) \rightarrow \mu_m(s), s_m) \approx \frac{1}{\sum_{i=10}^{60} U_{m,i} T_i}. \quad (31)$$

Numerical procedures used for calculating the theoretical curves in each of the figures are described in Supplementary Section 8.

### Empirical example from Fig. 1

The relative fitness estimates for the modifier example in Fig. 1a were obtained from ref. 8. This study examined two strains of *E. coli* that were isolated from generation 500 of Lenski’s long-term evolution experiment<sup>78</sup>. Relative fitnesses of the two strains at the first time point in Fig. 1a were obtained from head-to-head competitions under the same conditions as the original experiment. We converted the  $W$  values reported by ref. 8 into relative (log) fitness estimates using the procedure described in Supplementary Section 8. This yielded a log fitness difference of  $\Delta X = -4.4\%$  ( $-4.17\%$ ,  $-4.66\%$ ) (mean and 95% confidence intervals from replicate competition assays). The fitness measurements at the second time point were obtained after evolving each isolate for an additional 883 generations in 20 independent replay experiments and performing pooled fitness measurements of the evolved populations. Numerical conversion of the corresponding  $W$  values yielded a log fitness difference of  $\Delta X = 1.4\%$  ( $0.2\%$ ,  $2.6\%$ ).

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

Fitness measurements and confidence intervals in Fig. 1a were obtained from the Supplementary Information of ref. 8. Simulation results in the remaining figures are available in the accompanying source data files. Source data are provided with this paper.

### Code availability

Source code for forward-time simulations, numerical calculations and figure generation are available via Github ([https://github.com/bgoodlab/evolution\\_of\\_evolvability](https://github.com/bgoodlab/evolution_of_evolvability)).

### References

- Matic, I. et al. Highly variable mutation rates in commensal and pathogenic *Escherichia coli*. *Science* **277**, 1833–1834 (1997).
- LeClerc, J. E., Li, B., Payne, W. L. & Cebula, T. A. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274**, 1208–1211 (1996).
- Loeb, L. A. Human cancers express mutator phenotypes: origin, consequences and targeting. *Nat. Rev. Cancer* **11**, 450–457 (2011).
- Sane, M., Diwan, G. D., Bhat, B. A., Wahl, L. M. & Agashe, D. Shifts in mutation spectra enhance access to beneficial mutations. *Proc. Natl Acad. Sci. USA* **120**, e2207355120 (2023).
- Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**, 703–705 (1997).
- Couce, A., Guelfo, J. R. & Blázquez, J. Mutational spectrum drives the rise of mutator bacteria. *PLoS Genet.* **9**, e1003167 (2013).
- Freddolino, P. L., Goodarzi, H. & Tavazoie, S. Fitness landscape transformation through a single amino acid change in the rho terminator. *PLoS Genet.* **8**, e1002744 (2012).
- Woods, R. J. et al. Second-order selection for evolvability in a large *Escherichia coli* population. *Science* **331**, 1433–1436 (2011).
- Aggeli, D., Li, Y. & Sherlock, G. Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. *Nat. Commun.* **12**, 5193 (2021).
- Johnson, M. S. et al. Phenotypic and molecular evolution across 10,000 generations in laboratory budding yeast populations. *Elife* **10**, e63910 (2021).
- Blount, Z. D., Borland, C. Z. & Lenski, R. E. Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **105**, 7899–7906 (2008).
- Palmer, M. E., Moudgil, A. & Feldman, M. W. Long-term evolution is surprisingly predictable in lattice proteins. *J. R. Soc. Interface* **10**, 20130026 (2013).
- Barrick, J. E., Kauth, M. R., Strelloff, C. C. & Lenski, R. E. *Escherichia coli* rpoB mutants have increased evolvability in proportion to their fitness defects. *Mol. Biol. Evol.* **27**, 1338–1347 (2010).
- Karlsson, K. et al. Deterministic evolution and stringent selection during preneoplasia. *Nature* **618**, 383–393 (2023).
- Colom, B. et al. Mutant clones in normal epithelium outcompete and eliminate emerging tumours. *Nature* **598**, 510–514 (2021).
- Fabre, M. A. et al. The longitudinal dynamics and natural history of clonal haematopoiesis. *Nature* **606**, 335–342 (2022).
- Gifford, D. R. et al. Identifying and exploiting genes that potentiate the evolution of antibiotic resistance. *Nat. Ecol. Evol.* **2**, 1033–1039 (2018).
- Vogwill, T., Kojadinovic, M. & MacLean, R. C. Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of pseudomonas. *Proc. R. Soc. B* **283**, 20160151 (2016).
- Cummins, E. A., Snaith, A. E., McNally, A. & Hall, R. J. The role of potentiating mutations in the evolution of pandemic *Escherichia coli* clones. *Eur. J. Clin. Microbiol. Infect. Dis.* <https://doi.org/10.1007/s10096-021-04359-3> (2021).
- Altenberg, L., Liberman, U. & Feldman, M. W. Unified reduction principle for the evolution of mutation, migration, and recombination. *Proc. Natl Acad. Sci. USA* **114**, E2392–E2400 (2017).
- Wagner, G. P. Feedback selection and the evolution of modifiers. *Acta Biotheor.* **30**, 79–102 (1981).
- Wilke, C., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* **412**, 331–333 (2001).
- Payne, J. L. & Wagner, A. The causes of evolvability and their evolution. *Nat. Rev. Genet.* **20**, 24–38 (2019).
- Pigliucci, M. Is evolvability evolvable? *Nat. Rev. Genet.* **9**, 75–82 (2008).
- Hansen, T. F., Houle, D., Pavličev, M. & Pélabon, C. (eds) *Evolvability: A Unifying Concept in Evolutionary Biology?* (MIT Press, 2023); <https://doi.org/10.7551/mitpress/14126.003.0022>
- Good, B. H. & Desai, M. M. Evolution of mutation rates in rapidly adapting asexual populations. *Genetics* **204**, 1249–1266 (2016).
- van Gestel, J. et al. Short-range quorum sensing controls horizontal gene transfer at micron scale in bacterial communities. *Nat. Commun.* **12**, 2324 (2021).

28. Raynes, Y., Wylie, C. S., Sniegowski, P. D. & Weinreich, D. M. Sign of selection on mutation rate modifiers depends on population size. *Proc. Natl Acad. Sci. USA* **115**, 3422–3427 (2018).
29. Tuffaha, M. Z., Varakunan, S., Castellano, D., Gutenkunst, R. N. & Wahl, L. M. Shifts in mutation bias promote mutators by altering the distribution of fitness effects. *Am. Nat.* **202**, 503–518 (2023).
30. van Nimwegen, E., Crutchfield, J. P. & Huynen, M. Neutral evolution of mutational robustness. *Proc. Natl Acad. Sci. USA* **96**, 9716–9720 (2001).
31. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
32. Levy, S. F. et al. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).
33. Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
34. Strelkowa, N. & Lässig, M. Clonal interference in the evolution of influenza. *Genetics* **192**, 671–682 (2012).
35. Zanini, F. et al. Population genomics of intrapatient HIV-1 evolution. *Elife* **4**, e11282 (2015).
36. Rochman, N. D. et al. Ongoing global and regional adaptive evolution of SARS-CoV-2. *Proc. Natl Acad. Sci. USA* **118**, e2104241118 (2021).
37. Kemp, S. A. et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**, 277–282 (2021).
38. Barroso-Batista, J. et al. The first steps of adaptation of *Escherichia coli* to the gut are dominated by soft sweeps. *PLoS Genet.* **10**, e1004182 (2014).
39. Zhao, S. et al. Adaptive evolution within gut microbiomes of healthy people. *Cell Host Microbe* **25**, 656–667.e8 (2019).
40. Lieberman, T. D. et al. Genetic variation of a bacterial pathogen within individuals with cystic fibrosis provides a record of selective pressures. *Nat. Genet.* **46**, 82–87 (2014).
41. Neher, R. A. Genetic draft, selective interference, and population genetics of rapid adaptation. *Annu. Rev. Ecol. Evol. Syst.* **44**, 195–215 (2013).
42. Good, B. H., Rouzine, I. M., Balick, D. J., Hallatschek, O. & Desai, M. M. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. *Proc. Natl Acad. Sci. USA* **109**, 4950–4955 (2012).
43. Fisher, D. S. Asexual evolution waves: fluctuations and universality. *J. Stat. Mech. Theory Exp.* **2013**, P01011 (2013).
44. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014).
45. Wagner, A. Evolvability-enhancing mutations in the fitness landscapes of an RNA and a protein. *Nat. Commun.* **14**, 3624 (2023).
46. Barton, N. H. & Etheridge, A. M. The relation between reproductive value and genetic contribution. *Genetics* **188**, 953–973 (2011).
47. Neher, R. A., Shraiman, B. I. & Fisher, D. S. Rate of adaptation in large sexual populations. *Genetics* **184**, 467–481 (2010).
48. Birk, C. W. & Walsh, J. B. Effects of linkage on rates of molecular evolution. *Proc. Natl Acad. Sci. USA* **85**, 6414–6418 (1988).
49. Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
50. Good, B. H. & Desai, M. M. Deleterious passengers in adapting populations. *Genetics* **198**, 1183–1208 (2014).
51. Desai, M. M. & Fisher, D. S. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798 (2007).
52. Hegreness, M., Shores, N., Hartl, D. & Kishony, R. An equivalence principle for the incorporation of favorable mutations in asexual populations. *Science* **311**, 1615–1617 (2006).
53. Nguyen Ba, A. N. et al. High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**, 494–499 (2019).
54. Karthikeyan, S. et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* **609**, 101–108 (2022).
55. Łukasz, M. & Lässig, M. A predictive fitness model for influenza. *Nature* **507**, 57–61 (2014).
56. Lee, J. M. et al. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl Acad. Sci. USA* **115**, E8276–E8285 (2018).
57. Łukasz, M. et al. Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389–395 (2022).
58. Wiser, M. J., Ribeck, N. & Lenski, R. E. Long-term dynamics of adaptation in asexual populations. *Science* **342**, 1364–1367 (2013).
59. Ascensao, J. A., Wetmore, K. M., Good, B. H., Arkin, A. P. & Hallatschek, O. Quantifying the local adaptive landscape of a nascent bacterial community. *Nat. Commun.* **14**, 248 (2023).
60. Couce, A. et al. Changing fitness effects of mutations through long-term bacterial evolution. *Science* <https://doi.org/10.1126/science.add1417> (2024).
61. Frenkel, E. M., Good, B. H. & Desai, M. M. The fates of mutant lineages and the distribution of fitness effects of beneficial mutations in laboratory budding yeast populations. *Genetics* **196**, 1217–1226 (2014).
62. Draghi, J. A., Parsons, T. L., Wagner, G. P. & Plotkin, J. B. Mutational robustness can facilitate adaptation. *Nature* **463**, 353–355 (2010).
63. Johnson, M. S., Martsul, A., Kryazhimskiy, S. & Desai, M. M. Higher-fitness yeast genotypes are less robust to deleterious mutations. *Science* **366**, 490–493 (2019).
64. Zheng, J., Guo, N. & Wagner, A. Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* **370**, eabb5962 (2020).
65. Sung, W., Ackerman, M. S., Miller, S. F., Doak, T. G. & Lynch, M. Drift-barrier hypothesis and mutation-rate evolution. *Proc. Natl Acad. Sci. USA* **109**, 18 488–18 492 (2012).
66. Wylie, C. S. & Shakhnovich, E. I. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc. Natl Acad. Sci. USA* **108**, 9916–9921 (2011).
67. Archetti, M. Survival of the steepest: hypersensitivity to mutations as an adaptation to soft selection. *J. Evol. Biol.* **22**, 740–750 (2009).
68. Good, B. H. & Desai, M. M. The impact of macroscopic epistasis on long-term evolutionary dynamics. *Genetics* **199**, 177–190 (2015).
69. Agarwala, A. & Fisher, D. S. Adaptive walks on high-dimensional fitness landscapes and seascapes with distance-dependent statistics. *Theor. Popul. Biol.* **130**, 13–49 (2019).
70. Hanage, W. P. Not so simple after all: bacteria, their population genetics, and recombination. *Cold Spring Harb. Perspect. Biol.* **8**, a018069 (2016).
71. Gardner, A. & Kalinka, A. T. Recombination and the evolution of mutational robustness. *J. Theor. Biol.* **241**, 707–715 (2006).
72. Weissman, D. B. & Hallatschek, O. The rate of adaptation in large sexual populations with linear chromosomes. *Genetics* **196**, 1167–1183 (2014).
73. Neher, R. A., Kessinger, T. A. & Shraiman, B. I. Coalescence and genetic diversity in sexual populations under selection. *Proc. Natl Acad. Sci. USA* **110**, 15 836–15 841 (2013).

74. Good, B. H., Walczak, A. M., Neher, R. A. & Desai, M. M. Genetic diversity in the interference selection limit. *PLoS Genet.* **10**, e1004222 (2014).
75. Dawson, K. J. Evolutionarily stable mutation rates. *J. Theor. Biol.* **194**, 143–157 (1998).
76. Barnett, M., Zellar, L. & Rainey, P. B. Experimental evolution of evolvability. Preprint at *bioRxiv* <https://www.biorxiv.org/content/early/2024/05/03/2024.05.01.592015> (2024).
77. Guillaume, F. & Otto, S. P. Gene functional trade-offs and the evolution of pleiotropy. *Genetics* **192**, 1389–1409 (2012).
78. Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991).
79. Kondrashov, F. A. & Kondrashov, A. S. Multidimensional epistasis and the disadvantage of sex. *Proc. Natl Acad. Sci. USA* **98**, 12 089–12 092 (2001).
80. Leigh, E. G. The evolution of mutation rates. *Genetics* **73**, Suppl 73:1–18 (1973).

## Acknowledgements

We thank D. Wong for useful discussions and S. Walton, O. Ghosh and Z. Liu for comments and feedback on the manuscript. This work was supported in part by the Alfred P. Sloan Foundation (FG-2021-15708) and a Terman Fellowship from Stanford University. B.H.G. is a Chan Zuckerberg Biohub San Francisco Investigator.

## Author contributions

Conceptualization: J.T.F. and B.H.G.; theory and methods development: J.T.F. and B.H.G.; analysis: J.T.F. and B.H.G.; writing: J.T.F. and B.H.G.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41559-024-02527-0>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41559-024-02527-0>.

**Correspondence and requests for materials** should be addressed to Benjamin H. Good.

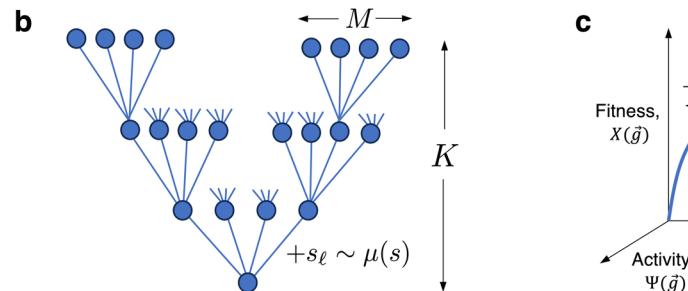
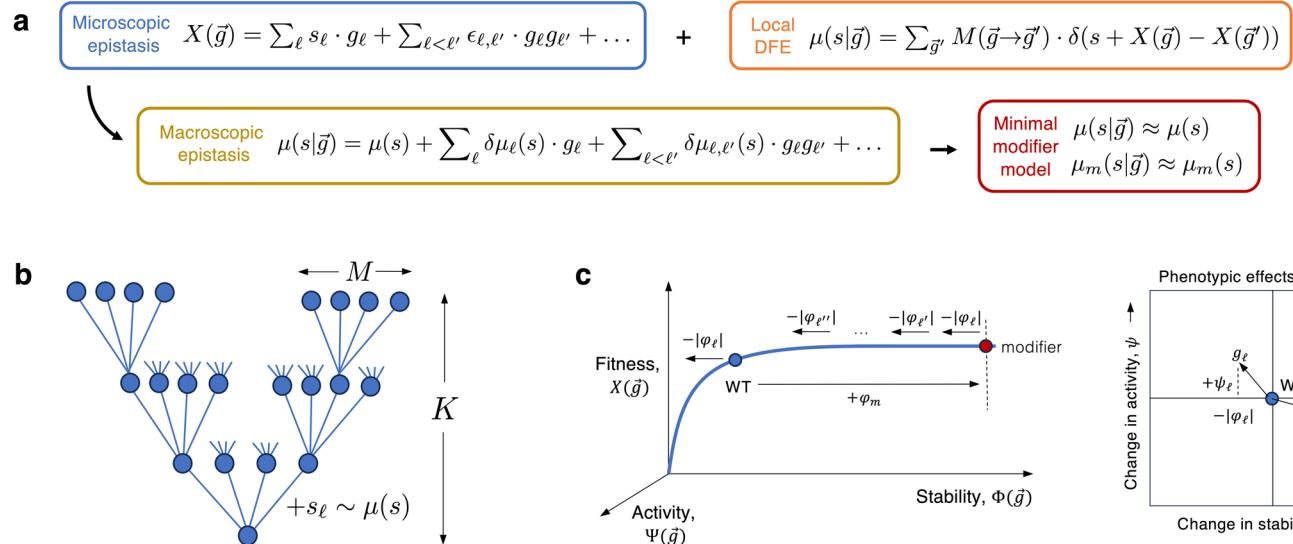
**Peer review information** *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

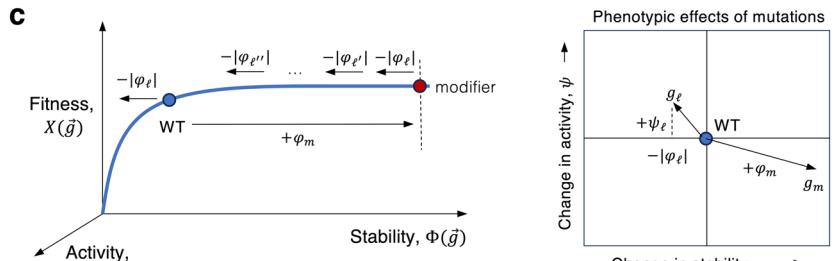
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

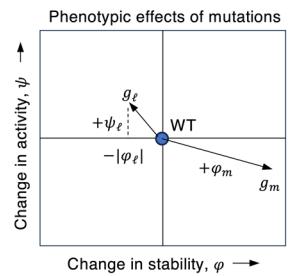
© The Author(s), under exclusive licence to Springer Nature Limited 2024

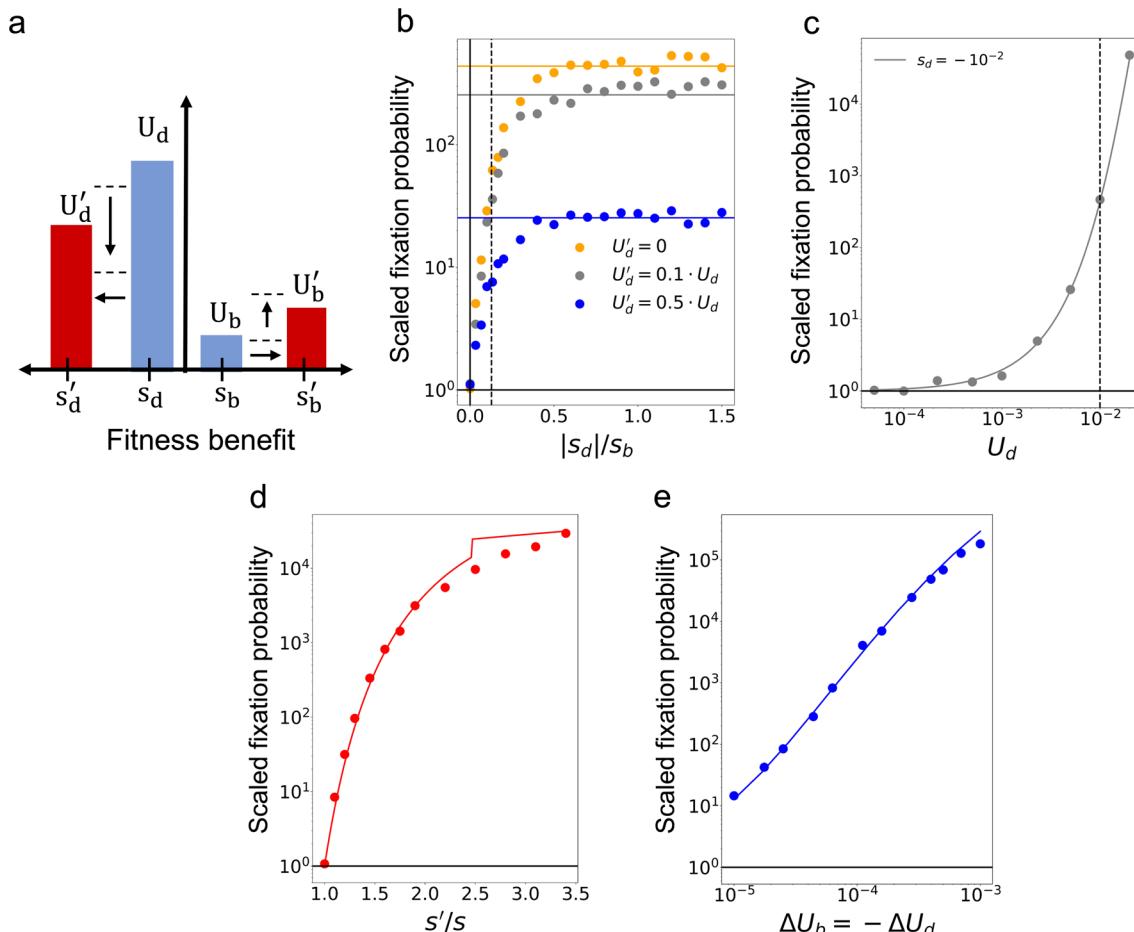


**Extended Data Fig. 1 | Examples of epistatic fitness landscapes that satisfy the minimal model in Fig. 1.** (a) The evolvability modifier in Fig. 1b can be viewed as the lowest order term in a general macroscopic epistasis expansion (left, SI Section 3.1). Different fitness landscapes can produce the same macroscopic behavior. (b,c) Examples of highly epistatic fitness landscapes that satisfy the simple model above. (b) A ‘maximally epistatic’ landscape of branching uphill paths, which generalizes the model in refs. 79,80. Each step  $k=1, \dots, K$  of a given path can access  $M \ll L$  beneficial mutations; all other genotypes have fitness zero.



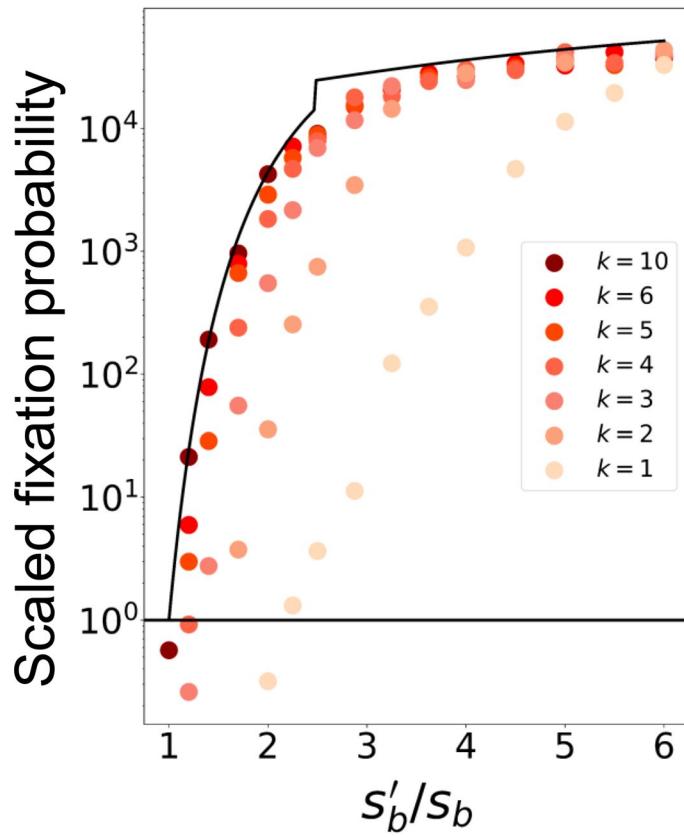
(c) A fitness landscape formed by a non-linear combination of two global phenotypic effects of mutations, for example, stability,  $\Phi(\vec{g})$ , and activity,  $\Psi(\vec{g})$ . Individual mutations can affect both traits simultaneously (right). Stabilizing mutations can act like modifier alleles by potentiating the fitness benefits of mutations that would destabilize the protein on their own (left). In particular, a strongly stabilizing mutation can allow  $K \approx \varphi_m / |\varphi_{\ell}|$  new mutations to accumulate before their effects on stability become important. See SI Section 3.1 for more details.





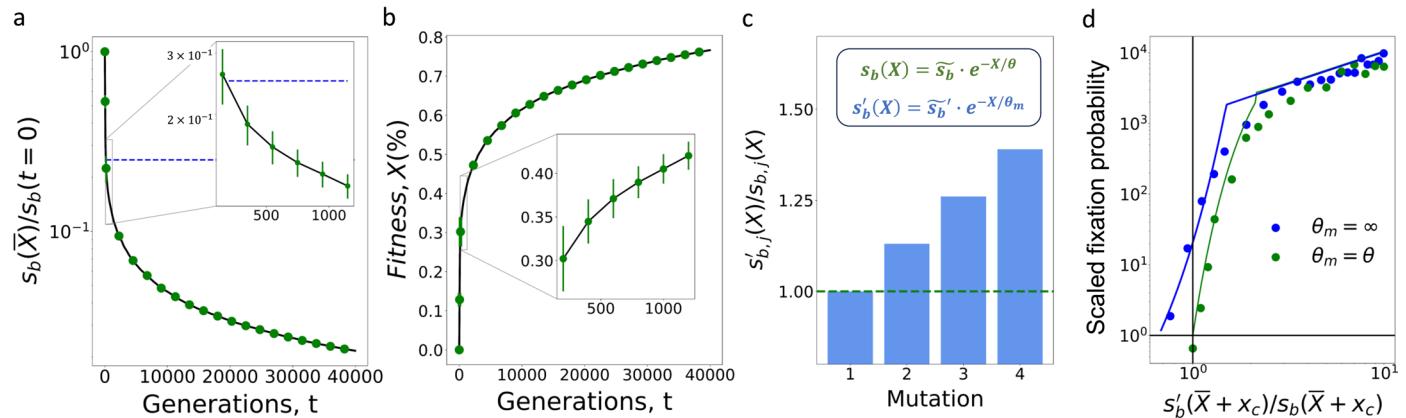
**Extended Data Fig. 2 | Deleterious mutations and indirect selection for robustness.** **(a)** A generalization of the simplified model in Fig. 5a, where the modifier can also shift the typical fitness cost  $s_d$ . **(b, c)** Fixation probability of a robustness-enhancing modifier with  $U'_d < U_d$  (and all other parameters are held fixed). Symbols denote the results of forward time simulations for  $N=10^8$ ,  $s_b=10^{-2}$ , and  $U_b=10^{-5}$ , while solid lines denote our theoretical predictions in SI Section 6. Panel (b) shows that the purging mutations approximation holds across a broad range of fitness costs, with the dashed line marking the predicted transition to quasi-neutrality ( $|s_d| \approx v/x_c$ ). Panel (c) shows that selection for increased robustness is relatively weak unless  $U_d \gtrsim s_b$  (dashed line). **(d)** Fixation probability of a modifier that imposes a tradeoff between robustness and evolvability by increasing the strength of selection on beneficial and deleterious

mutations simultaneously. Results are shown for  $s_b = |s_d| = s$  and  $U_d = 10^{-2}$ , with the remaining parameters the same as panel b. Since  $U_d \gg U_b$ , this example shows that strong selection for evolvability can occur for modifiers reduce the average fitness effect of mutations ( $\Delta\bar{s} \propto U_b\Delta s - U_d\Delta s < 0$ ). **(e)** Fixation probability of modifier that enhances robustness and evolvability at the same, by shifting mutations from deleterious to beneficial ( $U_d - U'_d = U'_b - U_b$ ). Symbols denote results of forward-time simulations with  $s_d = -10^{-2}$  and  $U_d = 10^{-2}$ , with the remaining parameters the same as panel (b). Lines denote our theoretical predictions in the absence of deleterious mutations ( $U_d = U'_d = 0$ ). This example shows that enhancements in evolvability are weighted more strongly than comparable increases in robustness, even when nearly all new mutations are deleterious ( $U_b \ll U_d$ ).



**Extended Data Fig. 3 | Relaxing the assumption that modifiers permanently change the mutation spectrum.** An alternative version of the model in Fig. 5c, where the modifier reverts to an evolutionary dead-end ( $\mu_m(s)=0$ ) after  $K$  mutations. Symbols denote the results of forward-time simulations for  $N=10^8$ ,  $s_b=10^{-2}$ , and  $U_b=10^{-5}$ , while the line denotes our theoretical predictions for the

minimal modifier model in Fig. 1b (that is  $K=\infty$ ). Even in this extreme case, our minimal modifier model (solid line) remains highly accurate for moderate values of  $K$ , and as little as  $K=1$  in the quasi-sweeps regime. This demonstrates that large populations can only 'see' across the fitness landscape for  $\approx x_{cm}/s'_b$  additional mutations (SI Section 7.2).



**Extended Data Fig. 4 | Selection for evolvability in the presence of diminishing returns epistasis.** **(a, b)** A simple model of global diminishing returns epistasis motivated by the empirical example in ref. 58 (SI Section 7.3). The fitness effects of new mutations shrink as the population adapts (panel a), leading to a decelerating rate of adaptation over time (panel b). Points denote the results of forward-time simulations for the distribution of fitness effects  $\mu(s\vec{g}) = U_b \cdot \delta(s - \bar{s}_b \cdot e^{-X(\vec{g})/\theta})$ , with  $U_b = 10^{-5}$ ,  $\bar{s}_b = 10^{-1}$ ,  $\theta = 0.2$ , and  $N = 10^7$ ; points are connected by solid lines to aid visualization. **(c, d)** The fixation probability of an evolvability modifier that arises at the beginning of the inset in panel a, where the fitness trajectory is still decelerating. Green symbols in (c) show a selection-strength modifier with the same diminishing returns schedule

as the background population ( $\theta_m \approx \theta$ ), while the blue symbols show an alternate example where the modifier avoids future diminishing returns once it arises ( $\theta_m \approx \infty$ ). The green line illustrates the predictions from the ‘adiabatic’ approximation in SI Section 7.3, demonstrating that the permanent modifier model [ $\mu_m(s\vec{g}) \approx \mu_m(s)$ ] provides a good approximation when the local selection strengths are properly renormalized. The blue line shows the predictions from our heuristic analysis in SI Section 7.3, which accounts for the additional benefits that accrue for the modifier lineage when  $\theta_m \gg \theta$  (panel d). This example illustrates that the evolvability advantages that accrue from large differences in diminishing returns epistasis can drive modest deviations from our existing theory when  $\theta$  grows close to  $x_c$ .

**Extended Data Table 1 | Table of mathematical symbols. Definitions of mathematical symbols used in the main text, along with locations where they are used**

Symbol	Definition	Figure/Equation
$N$	Population size	-
$\mu(s \vec{g})$	Distribution of fitness effects (DFE) of new mutations produced by genotype $\vec{g}$	Eq. (10)
$\mu(s)$	DFE in the wildtype background	Fig. 1B; Eq. (11)
$v$	Rate of adaptation of the wildtype population	Fig. 1A,C; Eq. (19)
$\mu_m(s)$	New DFE in the modifier background	Fig. 1B
$s_m$	Direct cost or benefit of the modifier allele	Fig. 1C; Eq. (2)
$v_m$	Rate of adaptation in a population fixed for the modifier allele	Fig. 1A
$p_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m)$	Fixation probability of a modifier mutation arising in an adapting wildtype population	Eq. (2)
$\tilde{p}_{\text{fix}}(\mu(s) \rightarrow \mu_m(s), s_m)$	Fixation probability scaled by the neutral expectation ( $1/N$ )	-
$w_m(x)$	Conditional fixation probability of a modifier lineage with an initial relative fitness $x$	Fig. 1C; Eq. (1)
$f(x)$	Relative fitness distribution of individuals in the wild-type population	Fig. 1C; Eq. (19)
$x_c$	Maximum relative fitness of individuals in the wildtype population	Fig. 1C; Eq. (19)
$s_b$	Fitness benefit of a typical fixed mutation in the wildtype population	Figs. 2A and 4A
$U_b$	Total mutation rate for mutations close to $s_b$	Figs. 2A and 4A
$s'_b, U'_b$	New values of $s_b, U_b$ for the modifier lineage	Fig. 2A; Eq. (26)
$\tilde{p}_{\text{fix}}(s_b \rightarrow s'_b)$	Fixation probability of a pure selection-strength modifier ( $s_m = 0, U'_b = U_b$ )	Fig. 2B; Eq. (3)
$\tilde{p}_{\text{fix}}(U_b \rightarrow U'_b)$	Fixation probability of a pure mutation-rate modifier ( $s_m = 0, s'_b = s_b$ )	Fig. 2C; Eq. (4)
$\tilde{p}_{\text{fix}}((U_b, s_b) \rightarrow (U'_b, s'_b))$	Fixation probability of compound modifier with $s_m=0$	Fig. 2C; Eq. (6)
$\alpha$	Weighting factor capturing non-additive interactions between selection-strength and mutation-rate changes	Fig. 2C; Eq. (6)
$\tilde{p}_{\text{fix}}(s)$	Fixation probability of an ordinary mutation [ $\mu_m(s) = \mu(s)$ ] with direct cost or benefit $s$	Fig. 3C; Eq. (5)
$\gamma$	Weighting factor capturing non-additive interactions between indirect and direct selection	Fig. 3C; Eq. (7)
$\delta\mu(s) \equiv \mu_m(s) - \mu(s)$	Continuous perturbation to the DFE	Fig. 4A
$U_0, s_0, \beta$	Total mutation rate, scale parameter, and shape parameter for the wildtype DFE in Fig. 4	Fig. 4A, bottom
$U_1, s_1$	Mutation rate and fitness effect of the local DFE perturbation in Fig. 4	Fig. 4A, bottom
$U_d, s_d$	Mutation rate and fitness cost for the deleterious mutations in Fig. 5A	Fig. 5A
$U'_d, s'_d$	New values of $U_d, s_d$ for the modifier lineage	Fig. 5A

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection	No software was used for data collection.
Data analysis	Information on code availability is provided under the "code availability" entitled section of the manuscript. The polyfit function in Numpy version 1.24 and fsolve and quad function in Scipy 1.10 were used for analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Source code for forward-time simulations and numerical calculations are available at Github: [https://github.com/bgoodlab/evolution\\_of\\_evolvability](https://github.com/bgoodlab/evolution_of_evolvability)

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	N/A
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	N/A
Recruitment	N/A
Ethics oversight	N/A

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	N/A
Research sample	N/A
Sampling strategy	N/A
Data collection	N/A
Timing and spatial scale	N/A
Data exclusions	N/A
Reproducibility	N/A
Randomization	N/A
Blinding	N/A

Did the study involve field work?       Yes       No

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

**Materials & experimental systems**

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	Plants

**Methods**

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input checked="" type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

**Plants**

Seed stocks

N/A

Novel plant genotypes

N/A

Authentication

N/A