Global epistasis emerges from a generic model of a complex trait

Gautam Reddy*
NSF-Simons Center for Mathematical & Statistical Analysis of Biology, Harvard University, Cambridge, MA 02138

Michael M. Desai†
NSF-Simons Center for Mathematical and Statistical Analysis of Biology, Harvard University, Cambridge, MA 02138
Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
Quantitative Biology Initiative, Harvard University, Cambridge, MA 02138, USA and Department of Physics, Harvard University, Cambridge, MA 02138, USA.

(Dated: June 15, 2021)

Abstract

Epistasis between mutations can make adaptation contingent on evolutionary history. Yet despite widespread "microscopic" epistasis between the mutations involved, microbial evolution experiments show consistent patterns of fitness increase between replicate lines. Recent work shows that this consistency is driven in part by global patterns of diminishing-returns and increasing-costs epistasis, which make mutations systematically less beneficial (or more deleterious) on fitter genetic backgrounds. However, the origin of this "global" epistasis remains unknown. Here we show that diminishing-returns and increasing-costs epistasis emerge generically as a consequence of pervasive microscopic epistasis. Our model predicts a specific quantitative relationship between the magnitude of global epistasis and the stochastic effects of microscopic epistasis, which we confirm by re-analyzing existing data. We further show that the distribution of fitness effects takes on a universal form when epistasis is widespread, and introduce a novel fitness landscape model to show how phenotypic evolution can be repeatable despite sequence-level stochasticity.

Keywords: evolution, epistasis, complex trait, predictability

* gautam_nallamala@fas.harvard.edu
† mdesai@oeb.harvard.edu

I. INTRODUCTION

Despite the idiosyncrasies of epistasis, a number of laboratory microbial evolution experiments show systematic patterns of convergent phenotypic evolution and declining adaptability. A striking example is provided by the *E.coli* long-term evolution experiment (LTEE) (Figure 1a): 12 replicate populations that adapt in parallel show remarkably similar trajectories of fitness increase over time [1, 2], despite stochasticity in the identity of fixed mutations and the underlying dynamics of molecular evolution [3, 4]. Similar consistent patterns of fitness evolution characterized by declining adaptability over time have also been observed in parallel yeast populations evolved from different genetic backgrounds and initial fitnesses [5] (Figure 1b) and in other organisms [6–12]. Declining adaptability is thought to arise from diminishing-returns epistasis [5, 13, 14], where a global coupling induced by epistatic interactions systematically reduces the effect size of individual beneficial mutations on fitter backgrounds. Diminishing-returns manifests as a striking linear dependence of the fitness effect of a mutation on background fitness (Figure 1c). While diminishing-returns can be rationalized as the saturation of a trait close to a fitness peak, recent work shows a similar dependence on background fitness even for deleterious mutations, which become more costly on higher fitness backgrounds [15]. This suggests that fitter backgrounds are also less robust to deleterious effects (Figure 1d), a phenomenon that has been termed increasing-costs epistasis. The origin of the global coupling that results in these effects is unknown.

Put together, these empirical observations suggest that the contributions to the fitness effect, s_i, of a mutation at a locus i in a given genetic background can be written as

$$s_i = s_{\text{additive},i} + s_{\text{genotype},i} - c_i y, \tag{1}$$

where $s_{\text{additive},i}$ is the additive effect of the mutation, $s_{\text{genotype},i}$ is its genotype-dependent epistatic contribution independent of the background fitness y (i.e., idiosyncratic epistasis), and c_i quantifies the magnitude of global epistasis for locus i. Eq. (1) reflects the observation that the strength of global epistasis depends on the specific mutation and applies independently of whether its additive effect is deleterious (increasing-costs) or beneficial (diminishing-returns). Over the course of adaptation in a fixed environment, global epistatic feedback on mutational effects can lead to a long-term decrease in adaptability. If this feedback dominates, Eq. (1) suggests that the dependence of the fitness effect on evolutionary history is summarized entirely by the current fitness, and therefore results in predictable fitness evolution.

Here, we show that diminishing-returns and increasing-costs epistasis are a simple consequence of widespread epistasis. This is consistent with recent work [16] that proposes a similar argument to explain these phenomena. However, while the core idea is similar, we present here an alternative framework based on the Fourier analysis of fitness landscapes, which leads to new insights and quantitative predictions. In particular, our framework leads to novel predictions for the relationship between the magnitude of global epistasis and the stochastic effects of microscopic epistasis, which we confirm by re-analyzing existing data. Extending this framework, we further quantify how the distribution of fitness effects shifts as the organism adapts and how the fitness effect of a mutation depends on the sequence of mutations that have fixed over the course of adaptation (i.e., historical contingency). While specific historical relationships depend on the genetic architecture, we introduce a novel fitness landscape model with an intuitive architecture for which the entire history is summarized by the current fitness. Using this fitness landscape model, we investigate the long-term dynamics of adaptation and elucidate the architectural features that lead to predictable fitness evolution.

RESULTS

Diminishing-returns and increasing-costs epistasis

We begin by examining the most general way to express the relationship between genotype and fitness (i.e., to describe the fitness landscape). A map between a quantitative trait (such as fitness), y, and the underlying genotype can be expressed as a sum of combinations of ℓ biallelic loci $x_1, x_2, \ldots, x_\ell$ that take on values $x_i = \pm 1$ [17–21]:

$$y = \bar{y} + \sum_i f_i x_i + \sum_{i>j} f_{ij} x_i x_j + \sum_{i>j>k} f_{ijk} x_i x_j x_k + \ldots, \tag{2}$$

where $\bar{y}$ is a constant that sets the overall scale of fitness. The symmetric convention $x_i = \pm 1$ for the two allelic variants is less often used than $x_i = 0, 1$, but it is an equivalent formulation, which we employ here because it will prove more convenient for our purposes (see [22] for a discussion). The coefficients of terms linear in x_i represent the additive contribution of each locus to the fitness (i.e. its fitness effect averaged across genotypes at all other loci), the higher-order terms quantify epistatic interactions of all orders, and $\bar{y}$ is the average fitness across all possible genotypes.

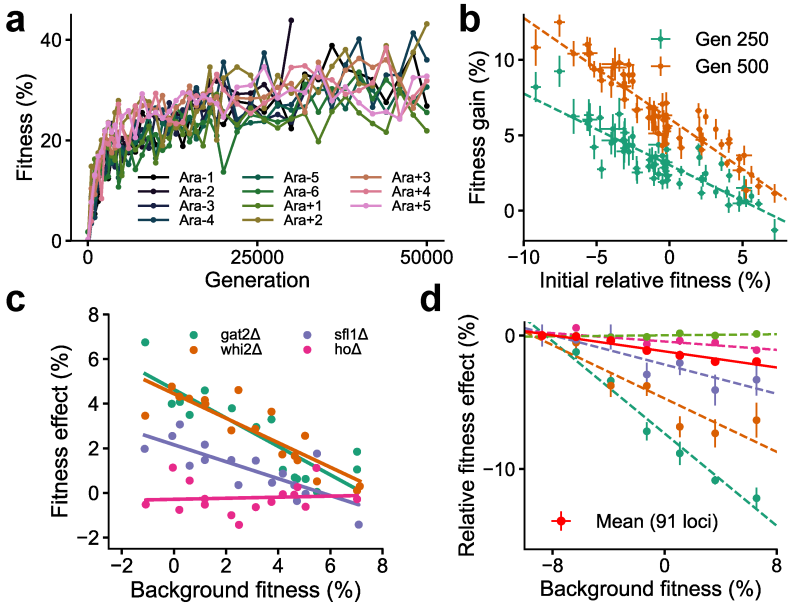

Figure 1. Declining adaptability and global epistasis in microbial evolution experiments. (a) Convergent phenotypic evolution in the *E.coli* long-term evolution experiment: the fitness relative to the common ancestor of 11 independently adapting populations over 50,000 generations is shown (data from [1]). The 12th population, Ara+6, has limited data and is not shown. (b) Yeast strains with lower initial fitness adapt faster (data from [5]). The fitness gain after 250 (green) and 500 (orange) generations of 640 independently adapting populations with 64 different founders and 10 replicates of each founder. Mean and SE are computed over replicates. (c) Diminishing returns of specific beneficial mutations on fitter backgrounds for three knocked out genes (green, orange and purple) (data from [5]). Control in pink. (d) Increasing costs of specific deleterious mutations on fitter backgrounds (data from [15]). The fitness effect relative to the least fit background for the mean over 91 mutations (in red) and five of the 91 mutations are shown. Linear fits for the five specific mutations and the mean using dashed and solid lines respectively are shown.

Importantly, Eq. (2) makes apparent the idiosyncrasies induced by epistasis: a mutation at a locus with ℓ interacting partners has an effect composed of $2^{\ell-1}$ contributions.

To explicitly compute the fitness effect of a mutation at locus i on a particular genetic background, we simply flip the sign of x_i, keeping all other x_j constant, and write down the difference in fitness that results. This fitness effect will generally involve a sum over a large number of terms involving the f's in Eq. (2). While this may suggest that an analysis of fitness effects via

Eq. (2) is intractable, the analysis in fact simplifies considerably if the locus has a significant number of independent interactions that contribute to the fitness (i.e., provided that the number of independent, nonzero epistatic terms associated to the locus is large). In this case, we show that the fitness effects of individual mutations decrease linearly with background fitness and the fluctuations around this linear trend are normally-distributed. In other words, widespread independent idiosyncratic epistatic interactions lead to the observed patterns of diminishing-returns and increasing-costs epistasis.

We present a derivation of this result in the SI. Here we explain the key intuition using a heuristic argument. The argument is based on a simple idea: for a well-adapted organism ($y > \bar{y}$) with complex epistatic interactions, a mutation is more likely to disrupt rather than enhance fitness. To be quantitative, consider a highly simplified scenario where some number N of the f's in Eq. (2) are ± 1 at random and the others are 0. In this case, the fitness of a given genotype is a sum of N_+ and N_- interactions that contribute positively and negatively to the trait respectively, each with unit magnitude, so that $y = \bar{y} + N_+ - N_-$. When positive and negative interactions balance, the organism is in a "neutrally-adapted" state ($y \approx \bar{y}$). By selecting for positive interactions, adaptation generates a bias so that $N_+ > N_-$ and $y > \bar{y}$. If locus i involved in a fraction v_i of all of $N = N_+ + N_-$ interactions is mutated, the effect of the mutation, on average, is to flip the sign of $N_+ v_i$ positive interactions and $N_- v_i$ negative interactions. The new fitness is then $y_i = y - 2N_+ v_i + 2N_- v_i = \bar{y} + (1-2v_i)(y-\bar{y})$ and thus $s_i = y_i - y = -2v_i(y-\bar{y})$. The negative linear relation between the background fitness, y, and the fitness effect of the mutation, s_i, is immediately apparent and emerges as a systematic trend simply due to a sampling bias towards positive interactions. Of course, while this relation is true on average, it is possible that locus i affects more or less positive interactions due to sampling fluctuations. Provided only that N is large and the interactions are independent, these fluctuations are approximately Gaussian with magnitude $\sqrt{N v_i (1 - v_i)}$.

This basic argument holds beyond the simple model with unit interactions. In the more general case, if the mutation is directed from $x_i = -1 \to +1$, we show in the SI that its fitness effect, s_i, on a background of fitness y can be written as

$$s_i = \underbrace{2f_i(1-\tilde{v}_i)}_{\text{additive}} - \underbrace{2\tilde{v}_i(y-\bar{y})}_{\text{global epistasis}} + \underbrace{\tilde{\epsilon}_i}_{\text{genotype}}, \tag{3}$$

where[1]

$$\tilde{v}_i \equiv \frac{\left(\sum_{j\neq i} f_{ij}^2 + \sum_{j>k\neq i} f_{ijk}^2 + \dots\right) - \left(\sum_{j\neq i} f_j f_{ij} + \sum_{j>k\neq i} f_{jk} f_{ijk} + \dots\right)}{\sum_{j\neq i}(f_j - f_{ij})^2 + \sum_{j>k\neq i}(f_{jk} - f_{ijk})^2 + \dots}, \tag{4}$$

and $\tilde{\epsilon}_i$ is a genotype and locus-dependent term which is distributed across genotypes with mean zero and variance expressed in terms of the f's from Eq. (2) (see SI for details). The numerator of $\tilde{v}_i$ in Eq. (4) is proportional to the covariance of fitness effects and background fitness and the denominator is the variance of background fitness across genotypes. A similar equation for the case $x_i = +1 \to -1$ can be derived. The choice of $+1 \to -1$ or $-1 \to +1$ is simply a matter of convention. If the convention is reversed, the coefficients of odd-order in Eq. (2), i.e., $f_i, f_{ijk}, \dots$, should also switch signs. It can be easily checked that reversing the signs of these quantities in the expression for $\tilde{v}_i$ above leads to the expression for $\tilde{v}_i$ when $x_i = +1 \to -1$.

Note that in general $\tilde{v}_i$ is not guaranteed to be positive and $\tilde{\epsilon}_i$ is arbitrary and determined by the genotype-fitness map. However, consistent patterns emerge when locus i has a large number of independent, nonzero epistatic terms and the additive effects $f_1, f_2, \dots$ of its interacting partners are not much larger than the epistatic terms (defined further below), which we call the widespread-epistasis (WE) limit. In the WE limit, $\tilde{\epsilon}_i$ is normally-distributed across genotypes with variance proportional to $\tilde{v}_i(1 - \tilde{v}_i)$. This follows from the same reasoning as in our heuristic argument with unit interactions above (see SI for details). In addition, $\tilde{v}_i$ is typically positive, giving rise to a negative linear trend (i.e. diminishing-returns and increasing-costs). We can see this by taking the third and higher-order terms in Eq. (4) to be zero, in which case $\tilde{v}_i$ is positive if $\sum_{j\neq i} f_{ij}^2 > \sum_{j\neq i} f_j f_{ij}$. This will typically be true in the WE limit because we expect $\sum_{j\neq i} f_{ij}^2$ to scale with the number of interacting partners ℓ, while each term in $\sum_{j\neq i} f_j f_{ij}$ can be positive or negative and thus the sum scales as $\sqrt{\ell}$ if the terms are independent. Thus when locus i has a large number of interacting partners, $\tilde{v}_i$ is typically positive unless the magnitude of the additive terms (a) is much larger than the magnitude of the epistatic terms (e), $a \gg e\sqrt{\ell}$. This argument is easily extended to the case when the third and higher-order terms are non-zero (see SI); the upshot is that the bias towards $\tilde{v}_i$ positive gets stronger with increasing epistasis.

The conditions for the WE limit are more likely to hold when the number of loci, ℓ, that affect the trait is large. Therefore, we expect to generically observe patterns of diminishing-returns and

[1] In the following equation and similar ones henceforth, a summation such as $\sum_{j>k\neq i} f_{ijk}^2$ is meant to denote a sum over pairs j, k, where each pair appears only once and no pair which includes index i appears. Symmetry of the f's w.r.t interchanged indices is also assumed (e.g., $f_{ijk} = f_{jik}$).

increasing-costs epistasis for a complex trait involving many loci. Importantly, whether we observe a negative linear trend does not depend on the magnitude of a locus' epistatic interactions relative to its own additive effect, but rather relative to the additive effects of its interacting partners. If we are not in the WE limit, and instead the additive effects dominate (i.e., $a \gg e\sqrt{l}$), then Eq. (4) suggests that the slope of the linear trend can be either positive or negative. We will show further below that recent experimental data demonstrates that both scenarios can be relevant: some loci have $a \ll e\sqrt{l}$ while others have $a \gg e\sqrt{l}$, with the former creating a bias towards the observed negative linear trends that characterize diminishing-returns and increasing-costs epistasis.

We note that Eq. (3) immediately leads to testable quantitative predictions: in the WE limit, the distribution of the residuals, $\tilde{\epsilon}_i$, obtained from regressing s_i and y is entirely determined by the slope of the regression, $-2\tilde{v}_i$. Specifically, we predict that these residuals (the deviations of individual genotype fitnesses from the overall diminishing-returns or increasing-costs trend) should be normally distributed with a variance proportional to $\tilde{v}_i(1-\tilde{v}_i)$. However, this condition only applies if diminishing-returns arises from the WE limit. It does not hold if epistasis is negligible, if locus i interacts significantly with only a few other dominant loci, or if the epistatic terms are interrelated (e.g., when global epistasis arises from a nonlinearity applied to an unobserved additive trait [23–25]). The latter case may still lead to a negative linear trend, but the statistics of the residuals will differ from Eq. (3) (see SI for a discussion).

It is convenient to subsequently work with the symmetric version of Eq. (3), where the fitness effects of both $x_i = -1 \to +1$ and its reversion $x_i = +1 \to -1$ (whose fitness effect is negative of the former) are included in the regression against their respective background fitness. In this case, the additive term is averaged out, and we show (SI) that in the WE limit,

$$s_i = -2v_i(y - \bar{y}) + 2\sqrt{v_i(1-v_i)}\eta_i, \tag{5}$$

where η_i depends on the genetic background and the locus, and is normally-distributed with zero mean and variance V, and

$$v_i \equiv \frac{V_i}{V} = \frac{f_i^2 + \sum_{j\neq i} f_{ij}^2 + \ldots}{\sum_k f_k^2 + \sum_{k>l} f_{kl}^2 + \ldots}. \tag{6}$$

Here V is the total genetic variance due to all loci (i.e., the variance in fitness across all possible genotypes) while V_i is the contribution to the total variance by the f's involving locus i. We therefore refer to v_i as the *variance fraction* of locus i. We show further below that for certain fitness landscapes, v_i can also be interpreted as the fraction of pathways affected by a locus. For

these reasons, we focus on v_i, which is half of the negative slope, rather than the slope. Note that the v_i's do not sum to one unless there is no epistasis (with epistasis, $\sum_i v_i > 1$, reflecting the fact that the variance contributed by different loci overlap). While the directed mutation case discussed previously is the relevant one when presenting experimental data (for e.g., Figure 1c,d), it is conceptually simpler to work with the symmetric case. These two cases coincide and $v_i \approx \tilde{v}_i$ in the WE limit if the additive effect of a locus is small (i.e., $f_i^2 \ll \sum_{j\neq i} f_{ij}^2 + \sum_{j>k\neq i} f_{ijk}^2 + \dots$).

Our results show that the variance fraction v_i plays an important role. It determines the slope of the negative relationship between the fitness effect and background fitness. At the same time, it determines the magnitude of the idiosyncratic fluctuations away from this trend. We also note that this slope can be used to experimentally probe the contribution of a locus to the trait (i.e., its variance fraction) taking into account *all* orders of epistasis, which circumvents the estimation of the individual f's in Eq. (2). The theory additionally predicts that the slope obtained by regressing the sum of fitness effects of two mutations at loci i, j against background fitness is proportional to $v_{ij} = v_i + v_j - 2e_{ij}$, where e_{ij} quantifies the magnitude of epistatic interactions of all orders between i and j (SI).

Importantly, while the fitness effects of individual mutations (and hence the distribution of fitness effects) may change over the course of evolution due to epistasis, the distribution of variance fractions (DVF) across loci, $P(v)$, is an invariant measure of the range of effect-sizes available to the organism during adaptation. As we will see, this means that the DVF plays an important role in determining long-term adaptability.

Numerical results and experimental tests

To illustrate our analytical results, we first demonstrate that the effects described above are reproduced in numerical simulations. To do so, we numerically generated a genotype-phenotype map of the form in Eq. (2), with $\ell = 400$ loci and an exponential DVF, $P(v) = \bar{v}^{-1} e^{-v/\bar{v}}$, where $\bar{v} = 0.02$ (Methods). This DVF is shown in Figure 2a. Note that $\bar{v}\ell \gg 1$ corresponds to an epistatic landscape; $\bar{v}\ell = 8$ chosen here thus corresponds to a model within the WE limit (note that $\tilde{v}_i \approx v_i$ in this parameter range). Using this numerical landscape, we measured the fitness effect of mutations at 30 loci across 640 background genotypes with a range of fitnesses (Figure 2b). Our results recapitulate the predicted linear dependence on background fitness (Figure 1c,d),

with a negative slope equal to twice the variance fraction predicted from Eq. (5). We further simulated the evolution of randomly generated genotypes similar to the experimental procedure used in Kryazhimskiy *et al.* [5] (Figure 2c), finding that our results reproduce the patterns of declining adaptability observed in experiments (Figure 1b). Note that ∼10 mutations are fixed during this simulated evolution; declining adaptability here is not due to a finite-sites effect.

As described previously, Eq. (5) implies a proportional relationship between the magnitude of global epistasis (quantified by the slope of the relationship between the fitness effect of a mutation and the background fitness) and the magnitude of microscopic epistasis (quantified by the residual variance around this linear trend); see also Figure 3a. We verify this relationship in simulations (Figure 2d). We predict that the slope obtained by regressing the sum of fitness effects of two mutations at loci i, j against background fitness is proportional to $v_{ij} = v_i + v_j - 2e_{ij}$. We further assume that $e_{ij} = O(\bar{v}^2)$ (specifically, $e_{ij} = v_i v_j$ for the genotype-phenotype map used for numerics). Since v_i and v_j are typically small for a complex trait, we expect near-additivity $v_{ij} \approx v_i + v_j$ and that any deviations are sub-additive, which is confirmed in simulations (Figure 2e,f).

While testing the latter prediction on double mutants requires further experiments, we can immediately test the relationship between the slope and the distribution of residuals from existing experimental data. To do so, we re-analyzed the data from Johnson *et al.* [15], which measured the fitness effect of 91 insertion mutants on about 145 backgrounds. These background strains were obtained by crossing two yeast strains that differed by $\approx 40,000$ SNPs. Of these 40,000 loci, $\ell \approx 40$ have been identified as causal loci with currently available mapping resolution [26]. In Figure 3, we show the estimated $\tilde{v}_i$ (negative one-half of the slope of the best-fit line) and the variance fraction v_i for each of the 91 mutations. These mutations were selected after screening for nonzero effect, and thus the DVF is biased upwards. The mean variance fraction is $\bar{v} \approx 0.06$. The wide range of v_i observed in the data implies that the epistatic influence of loci varies greatly across loci and we will show further below that this is crucial for maintaining a supply of beneficial mutations even when the organism is well-adapted to the environment.

Our theoretical results imply that we expect the linear relationship between background fitness and fitness effect to be negative if the additive effects of a locus' interacting partners are not much larger than the epistatic terms. Specifically, we define the additivity of interacting loci (AoIL) for

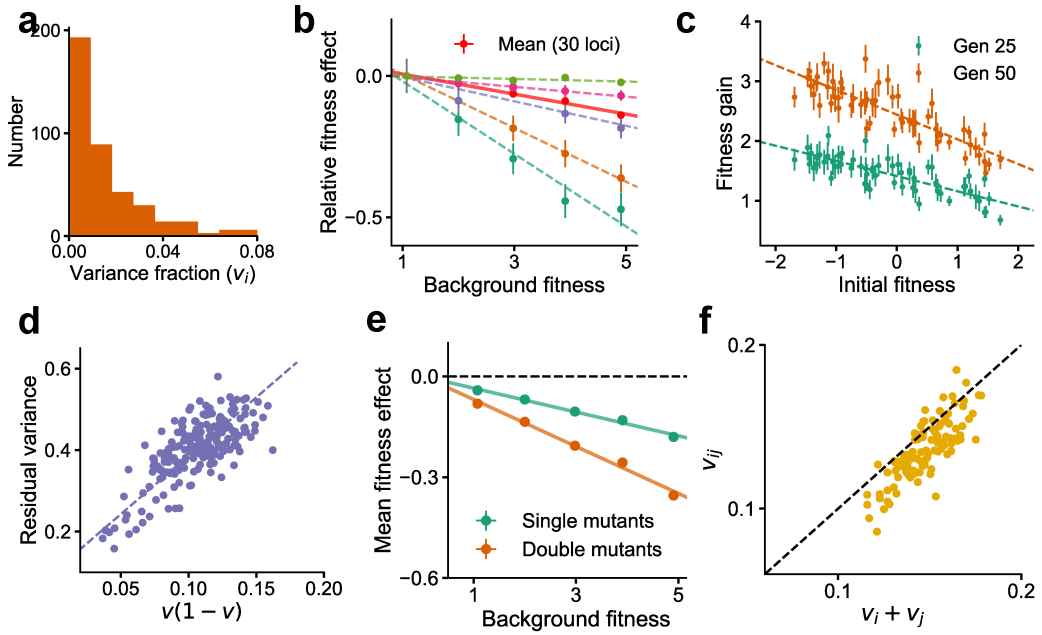

Figure 2. Global epistasis is recapitulated in a generic model of a complex trait and leads to testable predictions. (a) The distribution of variance fractions (DVF) over 400 loci for the simulated genotype-phenotype map. (b) The predicted linear relationship between fitness effect (relative to the fitness effect on the least fit background) and background fitness for the mean over 30 randomly chosen loci (red, solid line) and five loci (dashed lines in colors) is recapitulated. The slope of the linear fit for each locus is proportional to its variance fraction, v (slope $= -2v$). Mean and SE are over backgrounds of approximately equal fitness. See Methods for more details. (c) The mean fitness gain after 25 (green) and 50 (orange) generations of simulated evolution of 768 independently adapting populations with 64 unique founders and 12 replicates each. Means and SEs are computed over the 12 replicates. Error bars are s.e.m. (d) The relationship predicted from theory between the residual variance from the linear fit for each locus and its slope is confirmed in simulations. (e) The mean fitness effect for single mutants at 30 loci and double mutants from all possible pairs of the 30 loci. The slope for the double mutants is predicted to be roughly twice that of single mutants. (f) The estimated variance fraction of a double mutant with mutations at two loci is predicted from theory and confirmed in simulations to be approximately the sum of the variance fractions for single mutations at the two loci. Sub-additivity is due to epistasis between the two loci. See Methods for more details.

locus i as

$$\text{AoIL}(i) \equiv \frac{|\sum_{j\neq i} f_j f_{ij} + \sum_{j>k\neq i} f_{jk} f_{ijk} + \ldots|}{\left(\sum_{j\neq i} f_{ij}^2 + \sum_{j>k\neq i} f_{ijk}^2 + \ldots\right) + |\sum_{j\neq i} f_j f_{ij} + \sum_{j>k\neq i} f_{jk} f_{ijk} + \ldots|}, \tag{7}$$

which we show can be estimated from data (Methods and SI). If the AoIL is less than half, Eq. (4) implies that the linear trend is guaranteed to be negative. If instead the AoIL is greater than 0.5, the trend can be either positive or negative. The data shows a range of AoIL between 0 and 1 across loci. As predicted by our theory, we find that the loci with AoIL < 0.5 always show negative trends and the ones with AoIL > 0.5 show both negative and positive trends (Figure 3c). Importantly, the sign of the trend is determined by the AoIL and not by the additivity of the mutated locus, which we define as

$$\text{Additivity}(i) \equiv \frac{f_i^2}{f_i^2 + \sum_{j\neq i} f_{ij}^2 + \sum_{j>k\neq i} f_{ijk}^2 + \ldots}. \tag{8}$$

The additivity across loci also has a wide range. However, small additivity does not necessarily imply a negative trend (Figure 3d).

We next used the data from Johnson *et al.* [15] to analyze the relationship between the slope of the linear trend and the residual variance around this trend. We find that the experimental data confirms our theoretical prediction that the residual variance is proportional to $\tilde{v}_i(1-\tilde{v}_i)$ if the AoIL is small (Figure 3e, $R^2 = 0.5$ for loci with AoIL < 0.5 and $R^2 = 0.42$ for all loci). The Gaussian-distributed term in Eq. (3) also predicts the shape of the distribution of the residuals given the variance fractions, which aligns well with the empirical distribution of the residuals (Figure 3f).

Together, these theoretical results and our reanalysis of experimental data show that linear patterns of global diminishing-returns and increasing-costs epistasis are a simple consequence of widespread epistatic interactions. The distribution of variance fractions observed in data (Figure 3b) further implies that the epistatic influence of different loci on fitness can vary across a wide range. In what follows, we show that these two observations can be put together to make general predictions about the distribution of fitness effects, and consequently the long-term dynamics of adaptation. The key ingredient that enables this analysis (including Eq. (5)) is that in the WE limit, fitness and fitness effects are jointly normal (with respect to a uniform distribution over all possible genotypes), which allows us to quantify complex dependencies between these variables in terms of pairwise covariances.

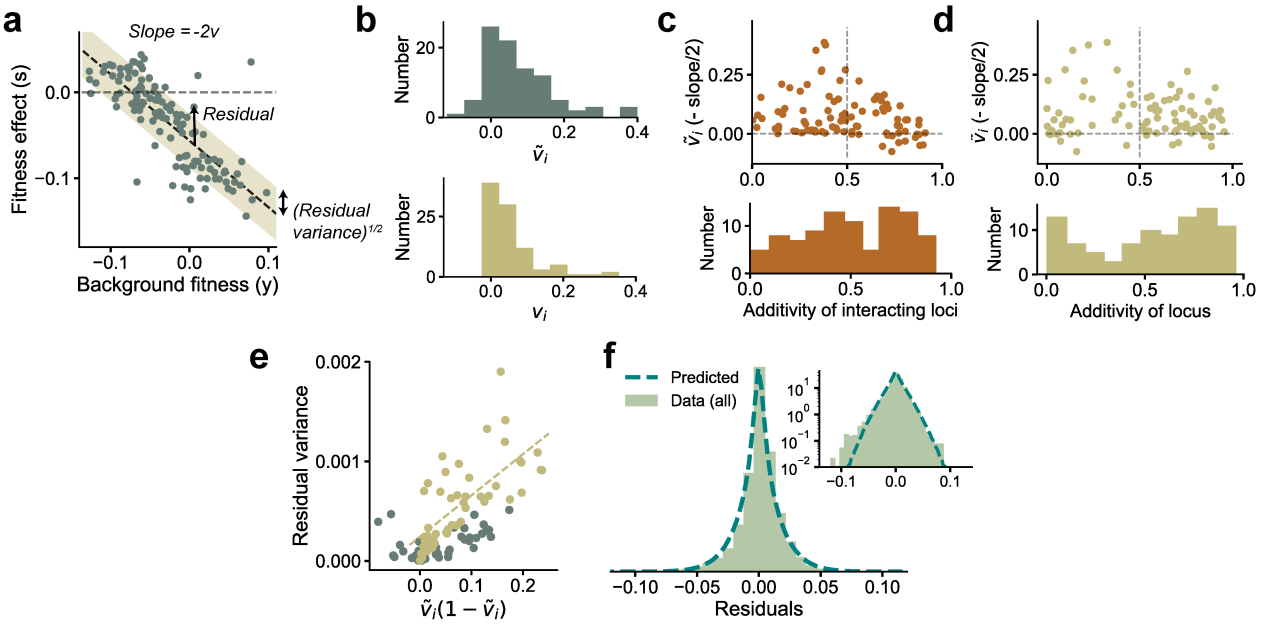

Figure 3. Experimental observations from Johnson *et al.* [15] are consistent with theoretical predictions. (a) The fitness effect of one of the 91 mutations from [15] plotted against background fitness. (b) The distribution of the measured $\tilde{v}_i$ (negative one-half of the slope from (a)) and variance fractions v_i for the 91 insertion mutations. (c,d) $\tilde{v}_i$ plotted against the additivity of interacting loci and the additivity of the mutated locus (see main text for definitions). The histograms are shown below the plots. The sign of the trend depends on the additivity of interacting loci rather than the additivity of the mutated locus. (e) The measured variance of the residuals against the prediction $\tilde{v}_i(1-\tilde{v}_i)$, shown here for the 91 mutations. The yellow circles correspond to the loci with AoIL < 0.5. The best-fit line (yellow dashed line) to these loci has $R^2 = 0.50$ ($R^2 = 0.42$ for all points). (f) The shape of the distribution of residuals pooled from all 91 mutations aligns well with the prediction from Eq. (3). The variances of the two distributions are matched. Inset: same plot in log-linear scale. See Methods for more details.

The distribution of fitness effects

Long-term adaptation is determined by the distribution of fitness effects (DFE) of possible mutations and the stochastic dynamical processes that lead to fixation. While Eq. (5) represents the distribution of the fitness effects of a specific mutation at locus i over *all* genotypes in the population that have fitness y, we are instead interested in the DFE, where fitness effects are measured for all the mutations arising in the background of a *particular* genotype that has fitness

y. For now we ignore the influence of evolutionary history on the DFE; we expand on that complication in the following Section.

Examining the DFE over ℓ loci for a randomly chosen genotype of fitness y can be thought of as sampling the fitness effects $s_1, s_2, \ldots, s_\ell$ from the conditional joint distribution $P(s_1, s_2, \ldots, s_\ell|y)$, which generally depends on epistasis. If the number of independent, nonzero epistatic terms is large, then $P(s_1, s_2, \ldots, s_\ell|y)$ is a multivariate normal distribution defined by the means and covariances of the $\ell + 1$ variables $y, s_1, s_2, \ldots, s_\ell$, which in turn can be computed in terms of the f's from Eq. (2). In particular, the conditional means and covariances are $\mathrm{Mean}_y(s_i) = -2v_i(y - \bar{y}), \mathrm{Cov}_y(s_i, s_j) = 4V(e_{ij} - v_i v_j)$, where e_{ij} is the epistatic variance fraction between loci i and j and $e_{ii} = v_i$. This implies that the conditional correlation between fitness effects is $(e_{ij} - v_i v_j)/\sqrt{v_i v_j (1 - v_i)(1 - v_j)}$.

The DFE simplifies considerably if we make certain additional assumptions on the magnitude of epistatic interactions. If we assume the typical variance fraction $\bar{v}$ is small (i.e., $\bar{v} \ll 1$) and also that e_{ij} is $O(\bar{v}^2)$, then correlations are $O(\bar{v})$ and thus negligible. Then, in a particular sample $s_1, s_2, \ldots, s_\ell$, we can think of each s_i as being drawn independently with mean $-2v_i(y - \bar{y})$ and variance $4v_i V$. To compute the DFE, $\rho(s|y)$, we first sample the variance fraction from the DVF, $P(v)$, and then sample a Gaussian random variable with the aforementioned mean and variance. This leads to the DFE

$$\rho(s|y) = \int_0^1 dv (2\sqrt{vV})^{-1} P(v) \varphi\left(\frac{s + 2v(y - \bar{y})}{2\sqrt{vV}}\right), \tag{9}$$

where φ is the standard normal pdf. Curiously, the correlations between s_i's vanish when $e_{ij} = v_i v_j$, in which case the above equation is exact and the DFE is determined entirely by the DVF. Further below, we introduce a specific fitness landscape model for which this relation does hold. Diminishing-returns is naturally incorporated in Eq. (9): the mean of s is $-2\bar{v}(y - \bar{y})$, i.e., the DFE shifts progressively towards deleterious values with increasing fitness.

Historical contingency in adaptive trajectories

A key unresolved question is the extent to which evolutionary history influences the DFE and the dynamics of adaptation [27]. That is, what does our theory say about historical contingency?

Suppose a clonal population of fitness y_0 accumulates k successive mutations resulting in

fitnesses $y_1, y_2, \ldots, y_k$. By virtue of arising on the same ancestral background, the fitness gain of a new mutation, s_{k+1}, is in general correlated with the full sequence of past fitnesses and the identity of the k mutations through its epistatic interactions with them. Based on these correlations, we use well-known properties of conditional normal distributions [28] to write

$$s_{k+1} = \sum_{i=0}^{k} w_{k+1,i} y_i + \epsilon, \tag{10}$$

where the weights $w_{k+1,i}$ depend on the variance fraction (v_{k+1}) of the new mutation and its epistatic interactions with past mutations. Here ϵ is the normally-distributed residual that depends on the initial genotype and the weights (SI). Eq. (10) is a generalization to a sequence of mutations of Eq. (5), which we can think of as the special case where $k = 0$.

To gain intuition, it is useful to first analyze Eq. (10) when $k = 1$ (i.e., to compute the effect of a second mutation conditional on the first). In this case, we show in the SI that

$$s_2 \simeq -2v_2(y_1 - \bar{y}) + \frac{v_1 v_2 - e_{12}}{v_1} s_1 + \epsilon, \tag{11}$$

where $s_1 = y_1 - y_0$ is the fitness effect due to mutation 1. The first term on the right hand side is the dependence on the fitness of the immediate ancestor, similar to the corresponding term in Eq. (5). The second term quantifies the influence of epistasis between loci 1 and 2 on s_2. When $e_{12} = v_1 v_2$, dependence on s_1 vanishes entirely and s_2 depends only on y_1. In contrast, if loci 1 and 2 do not interact, $e_{12} = 0$, and s_2 is, on average, larger *if* the mutation at 1 is beneficial compared to when it is deleterious. This has an intuitive interpretation: diminishing-returns applies to the overall fitness and the mechanism through which it acts is epistasis. However, if mutations 1 and 2 do not interact, then the increase in fitness corresponding to mutation 1 does not actually reduce the effect of mutation 2 (as expected by diminishing-returns) so the expected effect of mutation 2 is larger. This analysis suggests that during adaptation, since selection favors mutations with stronger fitness effects on the current background, a mutation that interacts less with previous mutations is more likely to be selected.

To identify the conditions under which history plays a minimal role, we would like to examine when s_{k+1} depends only on the current fitness, y_k, and is independent of both the past fitnesses and idiosyncratic epistasis. If this were true, then Eq. (5) would apply for new mutations that arise through the course of a single evolutionary path (i.e., the fitness effect of a new mutation is "memoryless" and depends only on its variance fraction and the current fitness). Surprisingly, such

a condition does exist. We show that this occurs when the magnitude of epistatic interactions between the new mutation and the k previous mutations, $e_{k+1,1:k}$, satisfies a specific relation: $e_{k+1,1:k} = v_{k+1}v_{1:k}$, where $v_{1:k}$ is the combined variance fraction of the k previous mutations (SI). In general, this condition is not satisfied, implying that there will be historical contingency which can be analyzed using the framework above. Remarkably, it turns out that a fitness landscape model for which the condition is satisfied does exist and arises from certain intuitive assumptions on the organization of biological pathways and cellular processes. This fitness landscape model additionally serves as an example of a landscape where global epistasis can vary substantially across loci. We describe this model below.

The connectedness model

We introduce the "connectedness" model (CN model, for short). In this model, each locus i is involved in a fraction μ_i of independent "pathways", where each pathway has epistatic interactions between all loci involved in that pathway (Figure 4a). The probability of an epistatic interaction between three loci (i, j, k) is then proportional to $\mu_i\mu_j\mu_k$, since this is the probability that these loci are involved in the same pathway. When the number of loci ℓ is large, we show that in this model, $v_i = \mu_i/(1+\mu_i)$, and when ℓ is small, $v_i = \mu_i/\bar{\mu}\ell$, where $\bar{\mu}$ is the average over all loci (SI). The CN model therefore has a specific interpretation: the outsized contribution to the fitness from certain loci (large v_i) is due to their involvement in many different complex pathways (large μ_i) and not from an unusually large perturbative effect on a few pathways. The distribution, $P(\mu)$, across loci determines the DVF.

Statistical fitness landscapes such as the NK model and the Rough Mt. Fuji model [27, 29–33] are related to the CN model. Specifically, the CN model is a sub-class of the broader class of generalized NK models (see [34] for a review). However, often-studied fitness landscape models have one important difference that distinguishes them and gives qualitatively different dynamics of adaptation (shown further below): in contrast to the CN model, classical fitness landscapes are typically 'regular.' That is, the variance fraction of every locus is assumed to be the same (except the star neighborhood model which has a bimodal DVF [34]).

The CN model is equivalent to a Gaussian fitness landscape with exponentially-decaying correlations (SI). The CN model has tunable ruggedness, where the landscape transitions from

additivity to maximal epistasis with increasing $\bar{\mu}$. Maximal epistasis corresponds to $\mu_i = 1$ (and hence $v_i = 1/2$) for all i. From Eq. (5), this implies that the new fitness after a mutation occurs is independent of the previous fitness, consistent with the expectation from a House-of-Cards model [35] (where genotypes have uncorrelated fitness). Regular fitness landscape models with exponentially-decaying correlations have memoryless fitness effects under the restrictive assumption that every locus is equivalent [27]. We show that the dynamics of adaptation of the more general CN model are also memoryless, i.e., the condition detailed in the previous section holds true (SI). Yet, as we show below, the predicted dynamics for the CN model are very different to those from a regular fitness landscape model.

We emphasize that the well-connectedness assumed for the CN model is not a requirement for Eq. (5) to hold. However, how diminishing-returns influences the long-term dynamics of adaptation depends on the specific genetic architecture and the corresponding fitness landscape. Consider for example an alternative model of genetic networks organized in a modular structure (Figure 4b). In this model, each locus is part of a single module, and interacts epistatically with other loci in that module to determine the fitness of that module; overall fitness is then determined as a function of the module fitnesses. In this case, the variance contributed by a locus is due to its additive contribution and from epistasis between loci restricted to its module. While the argument for diminishing-returns still applies to the fitness as a whole, it follows from the same argument that diminishing-returns should also apply to each module separately. Consequently, the dynamics of adaptation for the modular model are different from the CN model. For simplicity, we analyze the dynamics of adaptation for the CN model and postpone a discussion of how the dynamics differ for different models to subsequent work.

The dynamics of adaptation

We now examine the DFE that follows from Eq. (5) and what that implies for long-term adaptation under the conditions for memoryless fitness effects. We henceforth assume a large number of loci with sparse epistasis (though the total number of nonzero epistatic terms is still large). This implies that $\ell \gg 1$, $v_i \ll 1$ and $\bar{v}\ell \gg 1$; for simplicity, we also assume strong-selection-weak-mutation (SSWM) selection dynamics and $s \ll 1, Ns \gg 1$, where s are fitness effects and N is the population size. Under these conditions, a mutation sweeps and fixes in a

population before another one arises. The probability of fixation of a beneficial mutation, $p_{\rm fix}$, is then proportional to its fitness effect [36].

It is convenient to rescale fitnesses based on the total variance in fitness across all possible genotypes by defining $z = V^{-1/2}(y - \bar{y}), \sigma = V^{-1/2}s, \nu = V^{-1/2}\eta$. Note that ν is normally-distributed with zero mean and unit variance. Here z has an intuitive interpretation as the "adaptedness" of the organism. When the organism is neutrally-adapted ($|z| \ll 1$), positive and negative epistatic contributions to the fitness are balanced and diminishing-returns is negligible. Diminishing-returns is relevant when the organism is well-adapted ($z \gg 1$). Below, we give the intuition behind our analysis, which is presented in full detail in the SI.

In the neutrally-adapted regime, the linear negative feedback in Eq. (5) is negligible and the DFE is determined by the distribution of $\simeq v^{1/2}\nu$. Loci with large v can lead to a DFE with a long tail. If $\bar{v}$ is the typical variance fraction of a locus, the fitness increases as $z \sim n_s \bar{v}^{1/2}$, where n_s is the number of substitutions. Since $\bar{v}$ is a measure of overall epistasis, this implies that epistasis speeds adaptation in the neutrally-adapted regime by allowing access to more influential beneficial mutations.

Fitness increases until the effect of the negative feedback cannot be neglected. From Eq. (5), this happens when $\bar{v}z \sim \bar{v}^{1/2}\nu$ (i.e., when $z^2 \sim \bar{v}^{-1}$). Intuitively, fitness begins to plateau when its accumulated benefit from substitutions is comparable to the scale of the total genetic variance ($n_s\bar{v} \sim 1$) and further improvements are due to rare positive fluctuations. In this well-adapted regime, diminishing-returns and increasing-costs epistasis strongly constrain the availability of beneficial mutations, whose effects can be quantified in this model: for a mutation to have a fitness effect σ, we require from Eq. (5) that $\nu \simeq \sigma/2v^{1/2} + v^{1/2}z$, which has probability $\sim e^{-\nu^2/2}$. Beneficial effects of large σ arise when ν has a large positive deviation. The most likely v that leads to a particular σ is when ν is smallest (i.e., at $v^* \simeq \sigma/2z$), in which case $\nu \simeq \sqrt{2\sigma z}$, yielding a tail probability $\sim e^{-\sigma z}$. Remarkably, the beneficial DFE in the well-adapted regime is quite generally an exponential distribution independent of the precise form of the DVF (unless it is singular). In particular, we show in the SI that for the DFE, $\rho(\sigma|z)$,

$$\frac{\rho(\sigma|z)}{\rho(-\sigma|z)} = e^{-\sigma z}, \tag{12}$$

which depends solely on the adaptedness of the organism. The exponential form arises because of the Gaussianity of ν, but the argument can be easily extended to ν with non-Gaussian tails.

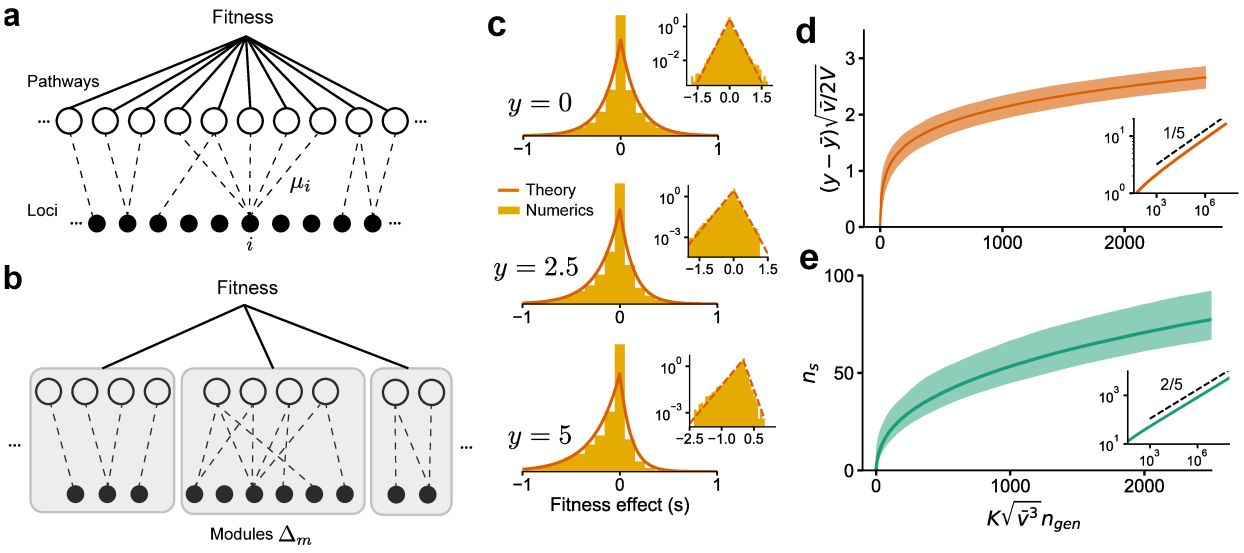

Figure 4. The DFE and long-term adaptation dynamics predicted for the connectedness model. (a) Schematic of the connectedness (CN) model, where each locus is associated with a fraction μ of pathways that contribute to the organism's fitness. (b) An alternative model with modular organization, where sets of loci interact only within the pathways specific to a single module. (c) The DFE predicted from Eq. (14) matches those obtained from simulated evolution of genotypes from the CN model. 128 randomly drawn genotypes (400 loci) with initial fitness y close to zero are evolved to $y = 2.5$ and $y = 5$ and the DFE is measured across loci and genotypes. We chose $\bar{y} = 0$ and $V = 1$ so that y represents adaptedness. Insets: same plots in log-linear scale. Note that the number of beneficial mutations acquired during the simulated evolution (∼10-20) is much less than the total number of loci (400). (d) For a neutrally-adapted organism, the theory predicts quick adaptation to a well-adapted state beyond which the adaptation dynamics are independent of the specific details of the genotype-fitness map. Shown here is the mean adaptation curve predicted under strong-selection-weak-mutation (SSWM) assumptions, which leads to a power-law growth of fitness with exponent 1/5 in the well-adapted regime (inset). (e) The number of fixed beneficial mutations under SSWM, which grows as a power-law with exponent 2/5 in the well-adapted regime (inset). The shaded region is the 95% confidence interval around the mean for (c) and (d). See Methods and SI for more details.

An exponential beneficial DFE has been previously proposed by Orr [37] but arises here due to a qualitatively different argument. Orr's result instead follows from extreme value theory: Suppose the fitness effects of ℓ loci ($\ell \gg 1$) are sampled from a DFE $\rho(\sigma)$ and $F(\sigma) \equiv \int_{-\infty}^{\sigma} \rho(\sigma')d\sigma'$. Then,

the probability that a beneficial mutation has at least a certain effect size σ is $P(\sigma_b \geq \sigma) = \frac{1-F(\sigma)}{1-F(0)} \approx \frac{\ln F(\sigma)}{\ln F(0)}$, where the latter approximation holds when beneficial mutations are rare (i.e., $1-F(0)$ is small). A well-known result from extreme value theory (Gumbel's law [38, 39]) implies that for a large family of distributions $\rho(\sigma)$ and for $\ell \gg 1$, we have $-\ell \ln F(\sigma) \propto e^{-k\sigma}$ (for some constant k) and therefore $P(\sigma_b \geq \sigma) = e^{-k\sigma}$. This argument is consistent with our results, but does not yield the dependence of k on adaptedness and the rate of beneficial mutations without additional information about $\rho(\sigma)$.

Under SSWM assumptions, from Eq. (12), the typical effect size of a fixed mutation is $\sigma_{\rm fix} \sim z^{-1}$, which typically has a variance fraction,

$$v^*_{\rm fix} \simeq \sigma_{\rm fix}/2z \sim 1/2z^2. \tag{13}$$

The above relation makes precise the effects of increasing-costs epistasis on adaptation. As adaptation proceeds, the delicate balance of high fitness configurations constrains fixed beneficial mutations to have *moderate* variance fractions. A mutation of small variance fraction is likely to confer small benefit and is lost to genetic drift, while one with a large variance fraction is more likely to disrupt an established high fitness configuration.

This intuition is not captured in regular fitness landscape models, which assume statistically equivalent loci, i.e., $v_i = \bar{v}$ for all i and $P(v) = \delta(v - \bar{v})$ is singular. From Eq. (9), we see that this leads to a Gaussian DFE whose mean decreases linearly with increasing fitness, in contrast to the exponential DFE in our theory. The key difference is the lack of loci with intermediate effect, which drive adaptation in the well-adapted regime. As a consequence, the rate of beneficial mutations declines exponentially ($U_b \sim e^{-\bar{v}z^2/2}$) and the fitness thus sharply plateaus at $z \sim \bar{v}^{-1/2}$. In contrast, our theory predicts a much slower depletion of beneficial mutations, $U_b \sim z^{-2}$ (SI). The rate of adaptation is $dz/dt \sim U_b p_{\rm fix}\sigma_{\rm fix} \sim z^{-4}$ (since $p_{\rm fix} \sim \sigma_{\rm fix}$), which leads to a slow but steady power-law gain in fitness, $z \sim t^{1/5}$. The rate of fixation of beneficial mutations is $dn_s/dt \sim U_b p_{\rm fix} \sim z^{-3} \sim t^{-3/5}$, which gives $n_s \sim t^{2/5}$.

We verify our analytical results using numerics. As before, we generated a genotype-phenotype map using the CN model with an exponential DVF, $P(v) = \bar{v}^{-1}e^{-v/\bar{v}}$ and $\ell = 400$ loci. The DFE can be calculated exactly by plugging in this $P(v)$ in Eq. (9):

$$\rho(\sigma|z) = \frac{\bar{v}^{-1}}{2\sqrt{2\bar{v}^{-1} + z^2}} e^{-\sigma z/2 - |\sigma|\sqrt{2\bar{v}^{-1}+z^2}/2}. \tag{14}$$

We simulated the evolution of randomly generated genotypes from $z = 0$ to $z = 2.5$ and $z = 5$ and the DFE across all loci was measured (we chose $\bar{y} = 0, V = 1$ so that $y = z, s = \sigma$). The theoretical prediction for the DFE, Eq. (14), closely aligns with the numerical results (Figure 4c).

Due to computational constraints, it is difficult to simulate evolution deep into the well-adapted regime. To compute the shape of adaptive trajectories and their variability, we instead simulated SSWM dynamics using the DFE directly from Eq. (14), beginning from a neutrally-adapted fitness ($z = 0$). Typical trajectories (Figure 4d) show rapid adaptation to the well-adapted regime beyond which the fitness grows slowly as $t^{1/5}$, as predicted from theory. The predictions for the number of fixed beneficial mutation are also re-capitulated (Figure 4e).

DISCUSSION

Recent empirical studies have observed consistent patterns of diminishing-returns and increasing-costs epistasis. Our model gives a simple explanation for these observations. In particular, we showed that these patterns are generic consequences of widespread microscopic epistatic interactions. The intuition underlying this result is that a random mutation typically has a larger disruptive effect on the delicate balance of microscopic epistasis that underpins a fitter background. Our model predicts a quantitative relationship between the magnitudes of global epistasis (i.e., the negative slope of diminishing-returns and increasing-costs epistasis) and microscopic epistasis, which we confirmed using existing data (Figure 3).

A similar explanation for diminishing-returns and increasing-costs epistasis has been recently proposed by Lyons *et al.* [16]. While our core argument for diminishing-returns and increasing-costs epistasis is the same as in that work, our Fourier analysis framework dissects the features of the fitness landscape necessary to observe these phenomena in terms of experimentally measurable average effects (i.e., the f's in Eq. (2)). In particular, we show that the additivity of a locus' interacting partners critically determines whether the trend is negative or unbiased. In addition, the Fourier analysis framework yields predictions for the distribution of fitness effects, the historical influence of past mutations on the fitness effect of a newly mutated site and motivates the proposed 'connectedness' fitness landscape model. The analysis of experimental data presented in Lyons *et al.* complements the experimental data considered here, lending further empirical

support for the prevalence of epistasis and its importance in determining long-term adaptability.

Our model leads to other experimentally testable predictions. The most direct and accessible test of the theory is to measure the fitness for all possible combinations of mutations at ∼10-15 significant loci and compare (using Eq. (6)) the magnitude of global epistasis to the measured fitness coefficients (the f's). Additionally, we predict that the magnitude of global epistasis of a double mutant should be nearly the sum of magnitudes of the corresponding single mutants, and any deviations should be biased towards sub-additivity. Since the predictions involve measuring residual variance, experimental noise can be an important confounding factor.

The observation that diminishing-returns occurs as a "regression to the mean" effect on certain fitness landscapes has been noted previously [40, 41]. The theory developed here quantifies precisely when we should expect to observe these patterns. We emphasize that our key result, Eq. (5), is a general statistical relation that holds if epistasis is widespread, irrespective of the specific genetic architecture and the corresponding fitness landscape. Weak epistasis with many loci is sufficient to observe noticeable patterns of global epistasis. However, the argument fails if the contribution of a locus is purely additive or when epistasis is limited to one or a handful of other loci. In the latter case, we expect the fitness effect of a mutation to be dominated by the allelic states of its partner loci, and thus take on a few discrete values. A few examples from Johnson *et al.* [15] indeed exhibit this pattern, (e.g. cases where the fitness effect of a specific mutation depends primarily on the allelic state at a single other locus).

We highlight a distinction between global epistasis discussed in this work and another form of global epistasis (also known as "nonspecific" epistasis) typically used in protein evolution to describe nonspecific epistatic interactions due to a nearly additive trait transformed by a nonlinear function [23–25, 42, 43]. This nonlinear function creates systematic relationships between epistasis terms and breaks the condition of independent epistatic terms required for our arguments to apply. Specific nonlinearities such as an exponential function may indeed lead to a negative linear trend on average, but the structure of the residuals differs from the one in Eq. (5) and observed in data.

A surprising empirical observation is that the negative linear relationship between fitness effect and ancestral fitness characteristic of global epistasis has different slopes for different loci. Our model identifies the negative slope as twice the fraction of variance contributed by a locus to the trait. To explain the wide range of variance fractions (VF) observed in data, we developed the connectedness (CN) model, a framework to think about the organization of cellular processes

that can lead to loci of widely varying VFs. In the CN model, loci have a large VF due to their involvement in many different pathways rather than due to a large effect on a single pathway. The CN model can be viewed as a statistical fitness landscape where loci can have a range of VFs, specified by the distribution of variance fractions (DVF). In the special case of every locus having the same VF, the CN model corresponds to a fitness landscape with tunable ruggedness and exponentially-decaying correlations.

Extending our framework to incorporate adaptation, we showed that the distribution of fitness effects (DFE) depends only on the current fitness, rather than the entire evolutionary history, under the intuitive assumptions behind the CN model. The theory therefore gives a simple explanation for why phenotypic evolution can be predictable, even while the specific mutations that underlie this evolution are highly stochastic.

Our framework has an implicit notion of 'adaptedness' without referencing a Gaussian-shaped phenotypic optimum, often assumed in models of adaptation (e.g. Fisher's geometric model) [44–46]. Over the course of adaptation, the DFE shifts towards deleterious values, reflecting diminishing-returns, which naturally arises from our basic arguments. For a well-adapted organism, we show that the DFE for beneficial mutations takes on an exponential form, and leads to universal adaptive dynamics. While an exponential DFE for beneficial mutations has been proposed previously based on extreme value theory [37], our result arises due to an entirely different argument: the tail of the beneficial DFE is determined by loci of intermediate size whose disruptive effect due to increasing-costs is small, yet whose effect size is large enough not to be lost due to genetic drift.

Our theory further predicts declining adaptability, with rapid adaptation in a neutrally-adapted regime followed by much slower increases in fitness, resulting in power-law adaptive trajectories when the organism is well-adapted. This is consistent with observations from the *E.coli* LTEE [1, 2]. Our model predicts a quicker decline in the number of substitutions ($n_s \sim t^{2/5}$) compared to the near linear trend observed in the LTEE data [4]. However, the dynamics of fixation in the LTEE deviate strongly from SSWM assumptions. This may explain the discrepancy, although we note that existing theory has only analyzed the effects of clonal interference and other breakdowns in SSWM assumptions for a constant DFE and weak epistasis [47, 48]. Further work will be required to understand how these effects interact with global epistasis. For example, we may expect that the effect of a highly beneficial mutation at a segregating locus is more likely to be

attenuated due to interference from subsequent deleterious mutations, while a less-fit lineage has a larger pool of beneficial mutations and is thus more likely to 'leapfrog' over more-fit lineages.

METHODS AND MATERIALS

The code and data to generate the figures are available at [49].

Simulations

We use a fitness landscape model with ℓ loci to generate the genotype-fitness map. Each locus is assigned a sparsity μ from $P(\mu)$, which is an exponential distribution with mean $\bar{\mu}$. Each of M independent pathways sample loci with each locus i having probability μ_i of being selected to a pathway. We choose $\ell = 400, \bar{\mu} = 0.02, M = 500$ so that $\bar{\mu}\ell = 8$ ensures signficant epistasis. All loci in a pathway interact with each other, where additive and higher-order coefficient terms of all orders were drawn independently from a standard normal distribution. The total fitness is the sum of contributions from the M pathways. We normalize the coefficients so that the sum of squares of all coefficients is 1, i.e., the total variance across genotypes is 1. The mean, $\bar{y}$ is close to zero from our sampling procedure. The above procedure is a simple and efficient way to generate epistatic terms to order $\sim$20, beyond which the computational requirements are limited by the exponentially increasing demand. Note that the effects described in the paper were also observed with only pairwise and cubic epistatic terms.

The variance fractions shown in Figure 2a can be calculated numerically from the definition. From the theory, given our choice of $P(\mu)$, these should follow an exponential distribution with mean $\bar{v} \approx \bar{\mu}/(1+\bar{\mu})$. There may be deviations since M is finite whereas the calculations assume $M \to \infty$. To generate Figure 2b, in order to get a range of background fitnesses, we first sample 128 random genotypes. These have fitnesses close to zero; in order to obtain a range of fitness values, we simulated the evolution of these 128 genotypes up to $y = 1, 2, 3, 4, 5$ under strong-selection-weak-mutation (SSWM) assumptions to get $128 \times 5 = 640$ genotypes at roughly five fitness values. The fitness effect of applying a mutation (i.e., flipping its sign) is measured for 30 randomly chosen loci (which are kept fixed) over each of the 640 genotypes. This is shown for five of the 30 and for the mean over the 30 loci in Figure 2b.

To generate Figure 2c, we sampled 64 random genotypes and 12 replicates of each. The evolution of these 768 genotypes was simulated for a total of 50 generations with a mutation rate of 1 per generation. The mean fitness gain over the 12 replicates is plotted for each of the 64 founders against their initial fitness.

To generate Figure 2d, the residuals are measured using the same procedure as for the experimental data analysis described below for the initial 128 genotypes at $y \approx 0$ and the 30 loci with the largest variance fraction.

Double mutants were created by mutating all pairs of the 30 randomly chosen loci on the 640 evolved genotypes. Their mean fitness effect was computed and plotted along with the mean fitness effect for single mutants, shown in Figure 2e. The variance fraction of the pair of loci for the double mutant was estimated as before and compared to the sum of the estimated variance fractions of the corresponding single mutants. This is shown in Figure 2f.

To generate the plots in Figure 4c, we simulated the evolution of 128 randomly sampled genotypes to $y = 2.5$ and $y = 5$. The fitness effect of 200 randomly sampled loci was measured and the distribution is plotted.

Analysis of the data from Johnson *et al.*

The data from Johnson *et al.* [15] consists of the fitness after the addition of 91 insertion mutations on each of 145 background genotypes. The fitness of a particular mutation at locus i can be modeled as

$$y_i = -c_i y + b_i + \text{Residual}_i(g), \tag{15}$$

where y_i, y are the mutant and background fitnesses respectively, c_i, b_i are constants for each locus and the residual $\text{Residual}_i(g)$ depends on the background genotype g.

We estimate the variance fraction $v_i = (1-\hat{\rho}_i)/2$, where the Pearson correlation $\hat{\rho}_i = \text{Corr}(y_i \oplus y, y \oplus y_i)$, where the symbol $\oplus$ denotes that the mutant and background fitness datasets are concatenated. $\tilde{v}_i$ is estimated as the negative one-half of the slope of the best linear fit of $s_i = y_i - y$ and y. The residuals for each of the 145 genotypes for each of the 91 mutations is simply

$$\text{Residual}_i(g) = (y_i + c_i y) - \overline{(y_i + c_i y)}, \tag{16}$$

where the overline represents an average over the 145 genotypes, which is used as an estimate of

the constant term and $c_i = 2\tilde{v}_i - 1$. In Figure 3b, we plot the distribution of estimated v_i and $\tilde{v}_i$. In Figure 3c, we compute the AoIL for each locus using Eq. (7), which we show in the SI to be $|\mathrm{Cov}(s_i, y_i + y)|/(|\mathrm{Cov}(s_i, y_i + y)| + \mathrm{Var}(s_i))$. In Figure 3d, we compute the additivity using Eq. (8). The additive effect is $f_i = \overline{(y_i - y)}/2$ and $\mathrm{Var}(s_i)/4$ gives the sum of squares of the epistatic terms (SI). In Figure 3e, we compute the variance of the residuals across the 145 genotypes for each locus and plot it against the locus' estimated $\tilde{v}_i(1-\tilde{v}_i)$. In Figure 3f, we plot the distribution of residuals over all genotypes and loci. The prediction is that in the WE limit the distribution of residuals is determined by $2\tilde{v}_i(1 - \tilde{v}_i)\eta$, where η is a Gaussian random variable. We multiply $\sqrt{\tilde{v}_i(1 - \tilde{v}_i)}$ for each locus with 10,000 i.i.d standard normal RVs, pool the resulting numbers for all loci and plot the predicted distribution in Figure 3f. The distributions are variance-matched. While Figure 3e shows that the variance of the residuals aligns with the theoretical prediction of being proportional to slope, Figure 3f shows that the data is also consistent with the predicted Gaussianity of the background-genotype-dependent contribution.

ACKNOWLEDGMENTS

We thank Sergey Kryazhimskiy, Andrew Murray, Milo Johnson, and members of the Desai lab for comments on the manuscript. G.R was supported by the NSF-Simons Center for Mathematical and Statistical Analysis of Biology at Harvard (award number #1764269) and the Harvard FAS Quantitative Biology Initiative. M.M.D. acknowledges support from the Simons Foundation (Grant 376196), NSF Grant PHY-1914916, and NIH grant R01GM104239.

[1] Wiser MJ, Ribeck N, Lenski RE. Long-term dynamics of adaptation in asexual populations. Science. 2013;342(6164):1364–1367.

[2] Lenski RE, Wiser MJ, Ribeck N, Blount ZD, Nahum JR, Morris JJ, et al. Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli. Proceedings of the Royal Society B: Biological Sciences. 2015;282(1821):20152292.

[3] Tenaillon O, Barrick JE, Ribeck N, Deatherage DE, Blanchard JL, Dasgupta A, et al. Tempo and mode of genome evolution in a 50,000-generation experiment. Nature. 2016;536(7615):165–170.

[4] Good BH, McDonald MJ, Barrick JE, Lenski RE, Desai MM. The dynamics of molecular evolution over 60,000 generations. Nature. 2017;551(7678):45–50.

[5] Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science. 2014;344(6191):1519–1522.

[6] Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nature Reviews Genetics. 2003;4(6):457–469.

[7] Perfeito L, Sousa A, Bataillon T, Gordo I. Rates of fitness decline and rebound suggest pervasive epistasis. Evolution. 2014;68(1):150–162.

[8] Wünsche A, Dinh DM, Satterwhite RS, Arenas CD, Stoebel DM, Cooper TF. Diminishing-returns epistasis decreases adaptability along an evolutionary trajectory. Nature Ecology & Evolution. 2017;1(4):1–6.

[9] Sanjuán R, Cuevas JM, Moya A, Elena SF. Epistasis and the adaptability of an RNA virus. Genetics. 2005;170(3):1001–1008.

[10] Couce A, Tenaillon OA. The rule of declining adaptability in microbial evolution experiments. Frontiers in genetics. 2015;6:99.

[11] Jerison ER, Kryazhimskiy S, Mitchell JK, Bloom JS, Kruglyak L, Desai MM. Genetic variation in adaptability and pleiotropy in budding yeast. Elife. 2017;6:e27167.

[12] Schenk MF, Szendro IG, Salverda ML, Krug J, De Visser JAG. Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. Molecular biology and evolution. 2013;30(8):1779–1787.

[13] Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. Negative epistasis between beneficial mutations in an evolving bacterial population. Science. 2011;332(6034):1193–1196.

[14] Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science. 2011;332(6034):1190–1192.

[15] Johnson MS, Martsul A, Kryazhimskiy S, Desai MM. Higher-fitness yeast genotypes are less robust to deleterious mutations. Science. 2019;366(6464):490–493.

[16] Lyons DM, Zou Z, Xu H, Zhang J. Idiosyncratic epistasis creates universals in mutational effects and evolutionary trajectories. Nature Ecology & Evolution. 2020;p. 1–9.

[17] Hordijk W, Stadler PF. Amplitude spectra of fitness landscapes. Advances in Complex Systems. 1998;1(01):39–66.

[18] Neher RA, Shraiman BI. Statistical genetics and evolution of quantitative traits. Reviews of Modern Physics. 2011;83(4):1283.

[19] Weinberger ED. Fourier and Taylor series on fitness landscapes. Biological cybernetics. 1991;65(5):321–330.

[20] Szendro IG, Schenk MF, Franke J, Krug J, De Visser JAG. Quantitative analyses of empirical fitness landscapes. Journal of Statistical Mechanics: Theory and Experiment. 2013;2013(01):P01005.

[21] Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-order epistasis? Current opinion in genetics & development. 2013;23(6):700–707.

[22] Poelwijk FJ, Krishna V, Ranganathan R. The context-dependence of mutations: a linkage of formalisms. PLoS computational biology. 2016;12(6):e1004771.

[23] Starr TN, Thornton JW. Epistasis in protein evolution. Protein Science. 2016;25(7):1204–1218.

[24] Sailer ZR, Harms MJ. Detecting high-order epistasis in nonlinear genotype-phenotype maps. Genetics. 2017;205(3):1079–1088.

[25] Otwinowski J, McCandlish DM, Plotkin JB. Inferring the shape of global epistasis. Proceedings of the National Academy of Sciences. 2018;115(32):E7550–E7558.

[26] Bloom JS, Kotenko I, Sadhu MJ, Treusch S, Albert FW, Kruglyak L. Genetic interactions contribute less than additive effects to quantitative trait variation in yeast. Nature communications. 2015;6(1):1–6.

[27] Agarwala A, Fisher DS. Adaptive walks on high-dimensional fitness landscapes and seascapes with distance-dependent statistics. Theoretical population biology. 2019;130:13–49.

[28] Eaton ML. Multivariate statistics: a vector space approach. JOHN WILEY & SONS, INC, 605 THIRD AVE, NEW YORK, NY 10158, USA, 1983, 512. 1983;.

[29] Neidhart J, Szendro IG, Krug J. Exact results for amplitude spectra of fitness landscapes. Journal of theoretical biology. 2013;332:218–227.

[30] Kauffman SA, Weinberger ED. The NK model of rugged fitness landscapes and its application to maturation of the immune response. Journal of theoretical biology. 1989;141(2):211–245.

[31] Stadler PF, Happel R. Random field models for fitness landscapes. Journal of Mathematical Biology. 1999;38(5):435–478.

[32] Aita T, Uchiyama H, Inaoka T, Nakajima M, Kokubo T, Husimi Y. Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: Application to prolyl endopeptidase

and thermolysin. Biopolymers: Original Research on Biomolecules. 2000;54(1):64–79.

[33] Altenberg L. Back T, Foger DB, Michalewicz Z, editors. NK Fitness Landscapes, In "The Handbook of Evolutionary Computation". Oxford Univ. Press; 1997.

[34] Hwang S, Schmiegelt B, Ferretti L, Krug J. Universality classes of interaction structures for NK fitness landscapes. Journal of Statistical Physics. 2018;172(1):226–278.

[35] Kauffman S, Levin S. Towards a general theory of adaptive walks on rugged landscapes. Journal of theoretical Biology. 1987;128(1):11–45.

[36] Haldane JBS. A mathematical theory of natural and artificial selection, part V: selection and mutation. In: Mathematical Proceedings of the Cambridge Philosophical Society. vol. 23. Cambridge University Press; 1927. p. 838–844.

[37] Orr HA. The distribution of fitness effects among beneficial mutations. Genetics. 2003;163(4):1519–1526.

[38] Gumbel EJ. Statistics of extremes. Courier Corporation; 2004.

[39] Majumdar SN, Pal A, Schehr G. Extreme value statistics of correlated random variables: a pedagogical review. Physics Reports. 2020;840:1–32.

[40] Draghi JA, Plotkin JB. Selection biases the prevalence and type of epistasis along adaptive trajectories. Evolution. 2013;67(11):3120–3131.

[41] Greene D, Crona K. The changing geometry of a fitness landscape along an adaptive walk. PLOS Comput Biol. 2014;10(5):e1003520.

[42] Otwinowski J. Biophysical inference of epistasis and the effects of mutations on protein stability and function. Molecular biology and evolution. 2018;35(10):2345–2354.

[43] Husain K, Murugan A. Physical Constraints on Epistasis. Molecular Biology and Evolution. 2020 05;Cvaa128.

[44] Fisher R. The genetical theory of natural selection. Clarendon Press; 1930.

[45] Orr HA. The genetic theory of adaptation: a brief history. Nature Reviews Genetics. 2005;6(2):119–127.

[46] Martin G, Lenormand T. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. Evolution. 2006;60(5):893–907.

[47] Good BH, Rouzine IM, Balick DJ, Hallatschek O, Desai MM. Distribution of fixed beneficial mutations and the rate of adaptation in asexual populations. Proceedings of the National Academy of

Sciences. 2012;109(13):4950–4955.

[48] Schiffels S, Szöllősi GJ, Mustonen V, Lässig M. Emergent neutrality in adaptive asexual evolution. Genetics. 2011;189(4):1361–1375.

[49] Reddy G. github:greddy992/global_epistasis; 2020. Available from: `https://github.com/greddy992/global_epistasis`.