



# A null model for the distribution of fitness effects of mutations

Olivier Cotto<sup>a,b,c</sup> and Troy Day<sup>a,b,1</sup>

Edited by Marcus Feldman, Stanford University, Stanford, CA; received October 25, 2022; accepted April 28, 2023

The distribution of fitness effects (DFE) of new mutations is key to our understanding of many evolutionary processes. Theoreticians have developed several models to help understand the patterns seen in empirical DFEs. Many such models reproduce the broad patterns seen in empirical DFEs but these models often rely on structural assumptions that cannot be tested empirically. Here, we investigate how much of the underlying “microscopic” biological processes involved in the mapping of new mutations to fitness can be inferred from “macroscopic” observations of the DFE. We develop a null model by generating random genotype-to-fitness maps and show that the null DFE is that with the largest possible information entropy. We further show that, subject to one simple constraint, this null DFE is a Gompertz distribution. Finally, we illustrate how the predictions of this null DFE match empirically measured DFEs from several datasets, as well as DFEs simulated from Fisher’s geometric model. This suggests that a match between models and empirical data is often not a very strong indication of the mechanisms underlying the mapping of mutation to fitness.

evolution | adaptation | theory | mutation | fitness landscape

Mutations are a key source of heritable variation and therefore they play a fundamental role in how evolution by natural selection occurs. One of the most important properties governing the fate of a new mutation is its effect on the fitness of the organism in which it appears. Even though mutations are often classified as deleterious, neutral, or beneficial, their fitness effects can be viewed as falling along a continuum. The relative frequency of the different fitness effects of mutations is referred to as the distribution of fitness effects (DFE). Describing and understanding the DFE is important for many reasons, including understanding the likely pathways of adaptation (1), the evolutionary potential for responding to environmental change (2, 3), the driftload (4), and the evolution of sex (5).

Empirical estimates of the DFE are difficult to obtain. The most direct way to estimate the DFE is to induce mutations on identified target genes and then compare the growth rate of each mutant genotype with that of a reference genotype (6). For logistical reasons, these methods have primarily been used in microorganisms (7–9), and in such studies, the DFE is generally found to be bimodal (10). One peak of the DFE corresponds to lethal mutations while the other encompasses mutations that range from mildly deleterious to beneficial. When excluding lethal mutations, and measuring selection as  $s = \exp(r_x - r_y)$  where  $r_x$  and  $r_y$  are the Malthusian growth rates of the focal and reference strains (11), the DFE often has a shape similar to a Gamma distribution (6). For multicellular organisms, a targeted mutation approach is often not feasible, but analogous studies can be done using mutation-accumulation experiments (10, 12). In either case, the goal is to estimate the DFE arising from de novo mutations. This contrasts with other approaches that attempt to infer the DFE from standing genetic variation using sequence data (see ref. 13).

Several theoretical frameworks have been developed to provide an understanding of the DFE and for making predictions about the form of the distribution that we expect to observe. The general approach is to construct a fitness landscape that maps genotypes to fitness (14). For example, the house of cards model (15), the NK model (16), and the rough Mount Fuji model (17), all assign a fitness value directly to each genotype in a way that depends on the mutational “distance” between them. In a second type of model, mutations are assigned a fitness value indirectly, by first considering their phenotypic effect and then mapping this phenotype to fitness. Within this approach, Fisher’s geometric model has been a common choice to analyze various evolutionary questions related to the DFE (e.g., refs. 18–23, reviewed in ref. 24). Fisher’s geometric model assumes that fitness is maximized at a particular value of some quantitative trait and that fitness declines smoothly as the distance of the trait from the optimum increases (24). The popularity of this model stems from its mathematical tractability, its intuitive appeal, and its emergence from primary biological principles (24, 25).

## Significance

The distribution of fitness effects (DFE) of new mutations plays a fundamental role in how evolution by natural selection occurs. A key research goal is therefore to infer properties of the underlying genotype-fitness map from empirically observed DFEs. Here, we show that such an inference is extremely difficult because many different genotype-fitness maps produce the very same DFE. Indeed, we demonstrate that if a genotype-fitness map is chosen at random, then it will almost certainly result in a DFE that has the largest possible information entropy. Subject to certain constraints this “null” DFE is shown to be a Gompertz distribution. We also demonstrate that this null DFE matches empirically measured DFEs from several datasets very well.

Author affiliations: <sup>a</sup>Department of Mathematics and Statistics, Queens University, Kingston, ON, K7L 3N6, Canada; <sup>b</sup>Department of Biology, Queens University, Kingston, ON, K7L 3N6, Canada; and <sup>c</sup>Plant Health Institute Montpellier, Université Montpellier, Institut National de Recherche pour l’Agriculture, l’alimentation et l’Environnement, Centre de coopération Internationale en Recherche Agronomique pour le Développement, Institut de Recherche pour le Développement, Institut Agro, Montpellier, F-34398, France

Author contributions: O.C. and T.D. designed research; performed research; contributed new reagents/analytic tools; analyzed data; and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup>To whom correspondence may be addressed. Email: day@queensu.ca.

This article contains supporting information online at [http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218200120/-/DCSupplemental](https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2218200120/-/DCSupplemental).

Published May 30, 2023.

Some of the above models make predictions that match the properties of empirical DFEs remarkably well (see ref. 6). For example, Fisher's geometric model predicts that the DFE should follow a modified Gamma distribution (20) and also generates relatively accurate predictions about other quantities such as epistasis and dominance (21). Recent studies, however, have shown that several fitness landscape models can produce predictions close to empirical observations (23, 24, 26). Indeed, it might well be very difficult to distinguish between theoretical models based on available data because empirical estimates of the DFE are often inferred from a limited number of observations. Furthermore, the underlying assumptions of some fitness landscape models—even though classically used in theoretical genetics—are very difficult to verify empirically. For example, the predictions of Fisher's geometric model depend on unverifiable assumptions about the complexity of the organism's phenotype, which is usually quantified by some notion of phenotypic dimension (see below for a more precise explanation).

What does a match between the predictions of fitness landscape models and empirical DFEs tell us about the underlying biology of mutation and fitness? Put another way, how much about the underlying "microscopic" biological processes involved in the mapping of genotypes to fitness can be inferred from "macroscopic" observations of the DFE? If many different underlying processes lead to the same predicted DFE then little would be learned about these underlying biological processes from empirical patterns of DFEs alone (27). One way to address this question is to determine the statistical properties of the DFE that are expected based on some kind of null or random underlying model. Any deviation from the null model could then be taken as a signal of the underlying biological processes involved. The goal of this paper is to develop one possible such null model for the DFE. We will show that this null model produces several predictions that match empirical data very well. We conclude by discussing how our results can be used, together with existing fitness landscape models, to better infer the mechanisms underlying empirical measurements of the DFE.

## 1. Modeling the DFE

In this section, we first cast previous theory for the DFE into a more abstract setting in order to better see the relationship between this theory and the null model. Then, we derive the null model.

**1.1 A General Setting for Landscape Models of the DFE.** Landscape models of the DFE can be viewed as having three main ingredients. The first is the set,  $G$ , of all genotypes of the organism. For the moment we make no assumption about how the genotypes in this set are related to one another, nor do we impose any structure on the set. Thus,  $G$  is simply a set containing all distinguishable genotypes in terms of their sequence identity, sequence length, karyotype, etc. Second, for models of the de novo DFE, an assumption is made about how to select a subset  $G_m \subseteq G$  of genotypes through mutation from a wild-type genotype  $g_{WT} \in G$ . Third, some method of assigning a fitness value to each genotype is assumed. With these ingredients, one then asks questions about the resulting distribution of fitness values of genotypes in the subset  $G_m$ .

Rather than working directly with the set of genotypes  $G$ , Fisher's geometric model restricts things by assuming that each genotype in  $G$  can be assigned a real-valued  $n$ -dimensional number  $x \in \mathbb{R}^n$ . In this way, a structure is imposed on the

set  $G$  by assuming that the elements of  $G$  can be embedded in what is, mathematically, called a metric space  $(\mathbb{R}^n, d)$  where  $d$  is the usual (possibly weighted) Euclidean "distance." In particular, under this embedding it then becomes meaningful to talk about the distance between different genotypes. Biologically,  $x$  is usually viewed as an  $n$ -dimensional phenotype corresponding to that genotype, but a precise definition of what constitutes a dimension or a phenotype is not typically given. Furthermore, since not all properties of an organism's phenotype can be described mathematically by a value in  $\mathbb{R}^n$ ,  $x$  does not have a clear biological interpretation. Notice that this also introduces the implicit assumption that there exists a continuum of possible genotypes rather than a discrete set since each genotype's identity can take any value in  $\mathbb{R}^n$ . For the second ingredient, we begin by choosing a particular genotype to label as the wild type. Then, because we have a way to quantify the distance between genotypes, the set of mutant genotypes  $X_m$  (which is the set in  $\mathbb{R}^n$  that corresponds to  $G_m$  in  $G$ ) is then generated by a form of nearest neighbor mutation from the wild type. Specifically, it is usually assumed that the set  $X_m$  is obtained by drawing values of  $x$  from an  $n$ -dimensional Gaussian density centered at  $x_{WT}$ . Finally, for the third ingredient, it is assumed that the fitness of a genotype  $x \in \mathbb{R}^n$  is a Gaussian function of  $x$  centered at some optimum. With these three restrictions, Fisher's geometric model thus essentially maps a particular chosen form for the distribution of mutation effects in phenotype space, through a particular chosen form of fitness function, to the DFE.

The Mt. Fuji and NK models are constructed in a similar way. Again, rather than working directly with the set of genotypes  $G$ , these models restrict things by characterizing each genotype as a string of letters of length  $n$ , where each letter is one of a set of possible choices (e.g., A, T, C, G). Often this is simplified further by allowing only two choices, and therefore a genotype is characterized as a binary string,  $x$ , of zeroes and ones. In this way, a structure is again imposed on the set  $G$ , but now by assuming that the elements of  $G$  can be embedded in a metric space  $(\{0, 1\}^n, d)$  where  $d$  is typically taken to be the hamming distance (i.e., the number of differences between two strings across all of the  $n$  sites). This thereby again allows us to talk meaningfully about the distance between genotypes. For the second ingredient, we again choose a particular genotype to label as the wild type, and the set of mutations is again generated by a form of nearest neighbor mutation from this wild type, but now using the hamming distance as the metric. Finally, for the third ingredient, many such models again simply assume a particular fitness mapping (e.g., fitnesses are assigned randomly from a uniform distribution). This description provides a different way to characterize genotypes than Fisher's geometric model, with the benefit that it has a clear biological interpretation. One shortcoming, however, is that it excludes the possibility of genotypes of different lengths or karyotypes. Nevertheless, as with Fisher's geometric model, these three restrictions mean that again the particular chosen form for the distribution of mutation effects in sequence space is then mapped, through a particular chosen fitness function, to the DFE.

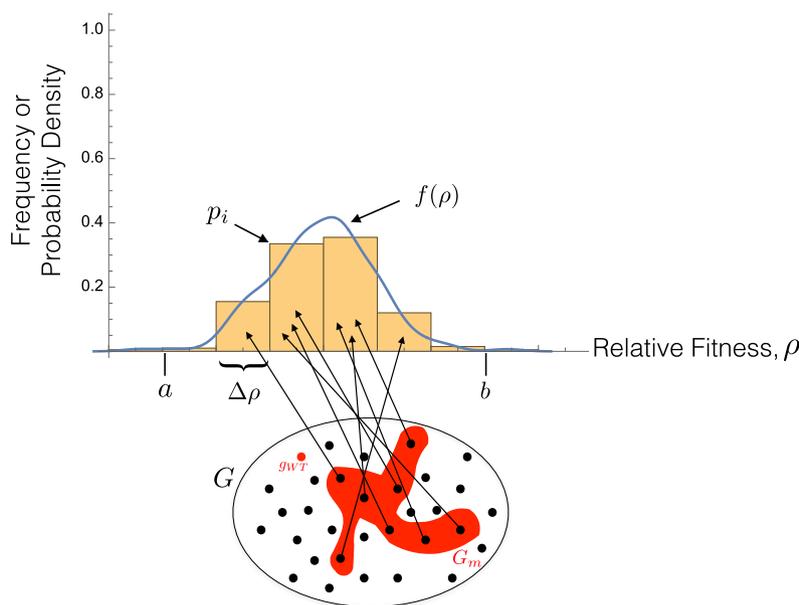
**1.2 Null Model for the DFE.** The null model developed here also falls within this general three-ingredient framework. We focus on asexually replicating organisms whose replication rate is measured while in the exponential growth phase because this is how most of the existing data were collected (7). In this context, empirical studies typically measure the fitness of a mutant genotype by estimating its population size at two (or more) time

points. The ratio of these estimates then gives a multiplicative measure of absolute fitness, which we denote by  $\lambda$ ; i.e.,  $\lambda$  is the factor by which a typical mutant individual multiplies over the time period in question. Sometimes this growth factor is log-transformed to give an additive measure of absolute fitness (i.e., the Malthusian growth rate), but either way the raw data are usually simply estimates of population size at different time points. Since reproduction is inherently a multiplicative process, we therefore take  $\lambda$  to be the fundamental measurement for which we seek a null model. Once a null model is obtained for this multiplicative measure of fitness, we can then derive the corresponding null model for other fitness measures from it (*SI Appendix, Appendix 1*). We wish to emphasize, however, that deriving a null model by starting with a fitness measure other than  $\lambda$  can produce different results, and this should be a focus of future research.

Most studies of the DFE focus on relative fitness rather than absolute fitness. They do so by normalizing the absolute fitness of each mutant genotype with the fitness of a reference genotype. Therefore, defining  $\lambda^*$  as the fitness of a reference genotype, we seek a null model for the values of  $\rho = \lambda/\lambda^*$ . The reference genotype used for  $\lambda^*$  could be the wild type from which the set of mutations is generated but it need not be. As with many previous studies, we focus on genotypes that allow population growth, meaning that  $\lambda \geq 1$ . Notice that this still allows for both beneficial and deleterious genotypes in terms of relative fitness,  $\rho$ . For the moment, we impose an upper bound  $b$  on possible values of relative fitness and we discretize the fitness axis into  $m$  fitness bins, each of width  $\Delta\rho$  (meaning that  $m\Delta\rho = b - \frac{1}{\lambda^*}$ ; Fig. 1). However, these are temporary restrictions and are used only to clarify the presentation and to ensure that the limiting results for arbitrarily large upper bounds  $b$ , and arbitrarily narrow fitness bins  $\Delta\rho$ , are well-justified mathematically. Thus, using  $f(\rho)$  to be an approximating, continuous, DFE, each bar  $i$  of the  $m$  fitness bins in the discretized version will correspond to a probability  $p_i = f(\hat{\rho}_i)\Delta\rho$  for some  $\hat{\rho}_i$  in the  $i$ th interval (Fig. 1).

A null model is a “statistical descriptor of (the) expected pattern (in data) in the absence of a particular mechanism” (28). There is a subtle but important distinction between null models and models of neutrality (e.g., the neutral theory of evolution). A null model is a statistical description of a null hypothesis, whereas a neutral model can also serve as an alternative, mechanistic, or process-based, hypothesis for observed patterns (28). Thus, a null model is not meant to be taken as a theory of how the natural world works. Instead, the chief purpose of a null model is to determine whether the observed data are unusual, or unexpected, if focal mechanism  $X$  were not occurring. By “not occurring,” one typically means that the data are instead generated in a completely unbiased fashion *with respect to the focal mechanism X*.

Landscape models of the DFE are based on specific mechanistic assumptions about how mutations get mapped, through development, to fitness. For example, Fisher’s geometric model assumes that, when mutation occurs, it has a small effect on a set of quantitative traits, that an intermediate value of each trait has the highest fitness according to a Gaussian function, and that there are no other fitness consequences of mutation. The Mt. Fuji model assumes that, when mutation occurs, it has a small effect on genome sequence identity, and that sequence identity is mapped to fitness via a particular function. Therefore, a reasonable null hypothesis for landscape models is that the genotype–fitness map is instead chosen randomly and in an unbiased way from the set of all possible maps. The rationale is that, if observed DFEs do not differ appreciably from this null hypothesis, then even if a landscape model’s predictions also matched the observed DFE very well, this alone would not be sufficient evidence to conclude that the landscape model has explanatory power. Of course, this does not mean that real genotype–fitness maps are actually randomly chosen (failure to reject the null hypothesis does not mean acceptance of the null hypothesis), but simply that the data contained in the DFE do not provide any evidence that the mapping of mutation to fitness is anything other than random.



**Fig. 1.** An example of a genotype–fitness map and the resulting DFE. Region  $G$  denotes the set of all genotypes (indicated by black dots),  $g_{WT}$  the wild type (indicated by a red dot), and the subset  $G_m$  of genotypes generated from  $g_{WT}$  by mutation (indicated as the red region). The set of arrows represent a randomly chosen genotype–fitness map, where  $\rho$  denotes relative fitness. The resulting DFE is shown (orange).  $p_i$  denotes the frequency of bin  $i$  and  $f(\rho)$  denotes the probability density of relative fitness value  $\rho$ .

To construct such a null model, we follow the same three-ingredient framework used by landscape models. For the first ingredient of the null model, we work directly with the set of genotypes  $G$  and do not impose any structure on this set. For the second ingredient, we again choose a wild-type genotype  $g_{WT} \in G$  but, because there is no structure imposed on the set  $G$ , we cannot generate the set of mutations  $G_m$  using any notion of distance from the wild type. In other words, there is no way to distinguish between any *particular* genotypes in  $G$  because there is no structure imposed on the set  $G$ . Thus, all we can do is specify the statistical properties of a set of mutant genotypes  $G_m$  and how this relates to  $g_{WT}$ . We will explain how this is done shortly. To specify the third ingredient, we begin by considering all possible genotype-fitness maps for the set  $G_m$ . Any such map will assign each one of the genotypes in  $G_m$  to some fitness value and so it will produce a DFE. We take the null DFE to be the fitness distribution that arises from a randomly chosen genotype-fitness map. To make this idea concrete, Fig. 1 displays a total of eight genotypes within the set  $G_m$  and six relative fitness bins. Any given genotype-fitness map assigns each of the eight genotypes to one of the six relative fitness bins, and thus, there are a total of  $6^8 = 1,679,616$  genotype-fitness maps. One of these maps is shown in Fig. 1.

A key observation is that some genotype-fitness maps give rise to the same DFE. To appreciate the consequences of this, suppose that there are  $n$  genotypes and  $m$  fitness bins, and let us first consider how many DFEs are possible. A crude upper bound can be obtained by noting that each bin can have anywhere from zero to  $n$  genotypes assigned to it. Thus, bin 1 can have at most  $n + 1$  different possibilities. The same is true for bin 2 etc. but, because the number in each of the bins is interdependent, there must certainly be less than  $(n + 1)^m$  different possible fitness distributions. On the other hand, the number of genotype-fitness maps grows exponentially with  $n$ . In particular, any genotype-fitness map places each genotype into one of the  $m$  fitness bins. Since there are  $m$  possibilities for each of the  $n$  genotypes, this gives a total of  $m^n$  different genotype-fitness maps ( $m = 6$  and  $n = 8$  in Fig. 1). Thus, the number of possible DFEs grows at most polynomially in the number of genotypes, whereas the number of genotype-fitness maps grows exponentially in the number of genotypes. This suggests that, for large genomes (i.e., large  $n$ ), there are many genotype-fitness maps that give rise to some common DFE(s). This already hints at the idea that inferring properties of the genotype-fitness map from the DFE is not likely to be easy.

The above qualitative ideas can be formalized by using a simplified version of Sanov's Theorem (29) from probability theory (see also ref. 30). When the number of genotypes  $n$  is large, the number of genotype-fitness maps that give rise to a particular DFE  $p_i$  is given by

$$e^{nH(p_i) + O(\ln n)}, \quad [1]$$

where  $H(p_i)$  is the information entropy of the DFE (*SI Appendix, Appendix 2*). From expression [1], we can then infer that the DFE with the largest information entropy will be generated by exponentially more genotype-fitness maps than any other distribution. To put this another way, if the number of genotypes  $n$  is large and we were to choose a genotype-fitness map at random from the set of  $m^n$  possible maps, then the DFE that results would almost certainly be one that has the highest possible information entropy (the probability of choosing a genotype-fitness map that gives rise to any other distribution is exponentially smaller; *SI Appendix, Appendix 2*). This result shows that a suitable null

model of the DFE of the set  $G_m$  is the distribution of fitness values with the largest possible information entropy.

Now that we have shown that the null DFE maximizes the information entropy of the fitness distribution on the set  $G_m$ , the final thing we need to specify to completely characterize the null distribution is some information about the subset  $G_m$  that is generated by mutation. Recall that there is no structure imposed on the set  $G$  (e.g., no notion of distance between genotypes) and therefore there is no way to distinguish between the identity of those particular genotypes that are likely to be generated by mutation from  $g_{WT}$  versus the identity of those genotypes that are not (if there was such a way to classify the identity of individual genotypes, then this classification could also be used as a measure of distance between genotypes). Therefore, we can only specify statistical properties of the entire collection of mutations  $G_m$  as a whole.

The simplest approach is to assume nothing about the statistical properties of  $G_m$ . It is then not difficult to show that the resulting DFE is a fixed uniform distribution. However, this is not a particularly useful null model because, even if mutations were randomly mapped to fitness, these fitness values would still depend on the environment in which they were assessed as well as the reference genotype used for computing relative fitness (and a fixed uniform distribution cannot account for these features). Therefore, in keeping with the general approach for constructing a null model (28), we seek to constrain the model in a way that accounts for these generic statistical features of the data. There are many ways to do so but all require that we tie at least some of the statistical properties of  $G_m$  to the environment of the assay and the reference genotype used.

The simplest specification is to assume that the null distribution of relative fitness values of  $G_m$  is "centered" at a specific location on the fitness axis and that this center depends on the environment and the fitness of the reference genotype,  $\lambda^*$ . There are several ways to characterize the center of a distribution, and the most convenient is the mean. Thus, we will take the null DFE to be the fitness distribution that results from choosing a random genotype-fitness map (i.e., the distribution with the largest possible information entropy) subject to the constraint that the DFE of  $G_m$  has a given mean determined by the environment and reference genotype used in the assay. It is well-known in the statistical literature that, if the upper bound  $b$  is large and the bin width  $\Delta\lambda$  is small, this maximum entropy distribution is approximated by the shifted exponential (*SI Appendix, Appendix 3*)

$$f(\rho) = \frac{e^{-\frac{\rho-a}{\rho-a}}}{\rho - a}, \quad [2]$$

where  $a = 1/\lambda^*$  is the lower bound of possible values of relative fitness, and is determined by the reciprocal of the multiplicative fitness  $\lambda^*$  of the reference genotype. Finally, most microbial experiments quantify the DFE using a transformation of multiplicative fitness that corresponds to the ratio of Malthusian growth rates (i.e., the distribution of  $z = r/r^*$ , see ref. 7). Using the relationship  $\rho = e^{(r-r^*)\tau}$  we can write  $z = 1 + \frac{\ln \rho}{\ln \lambda^*}$  and so obtain the null distribution of  $z$ , which is (*SI Appendix, Appendix 3*)

$$g(z) = \alpha\beta e^\alpha e^{\beta z - \alpha e^{\beta z}}, \quad [3]$$

where  $\alpha = \frac{1}{\rho\lambda^* - 1}$  and  $\beta = \ln \lambda^*$ . This is a Gompertz distribution. Both null DFEs [2] and [3] implicitly assume that the wild type is relatively poorly adapted since they assume no

upper bound on the allowable fitness effects of mutations. More general results can be found in *SI Appendix, Appendix 3*.

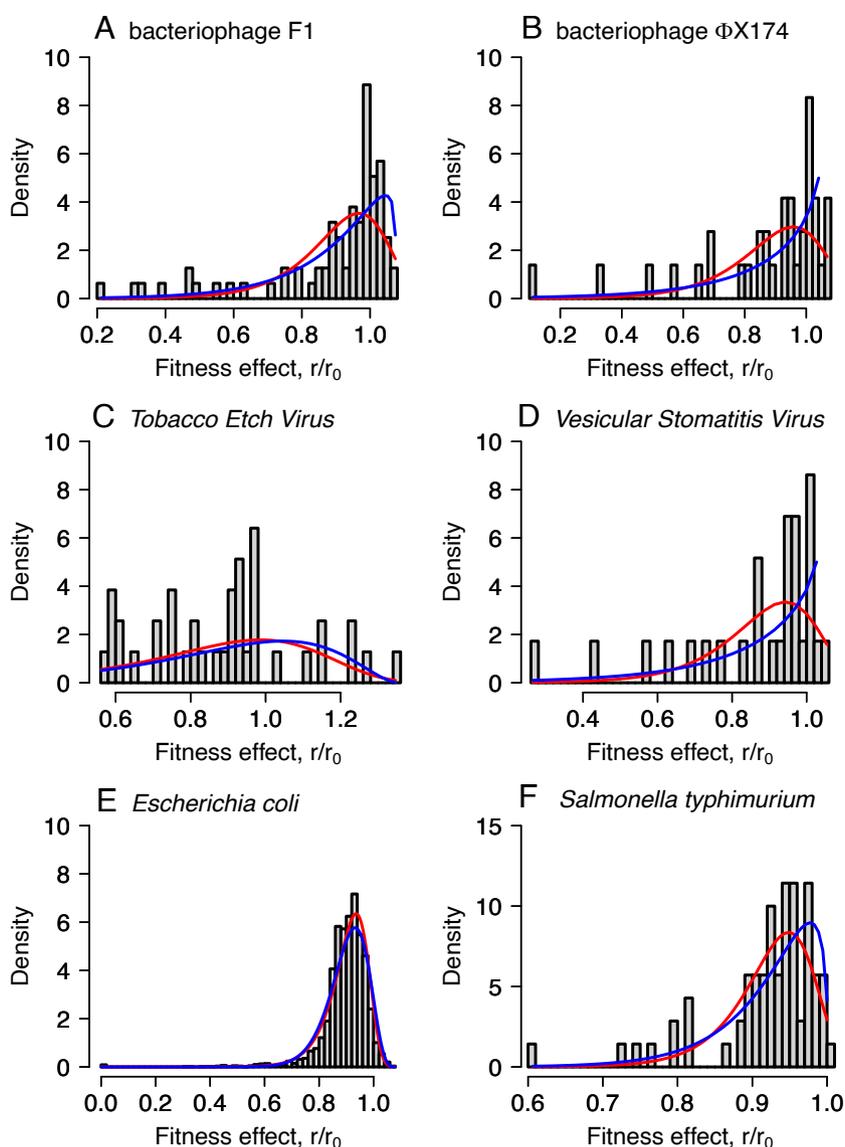
## 2. Results

We now examine how the DFE predicted from the null model compares with empirical DFEs obtained through site-directed mutagenesis in several microorganisms. We also examine how these predictions compare with DFEs simulated using Fisher's geometric model, as well as how the patterns of epistasis predicted by the null model compare with data.

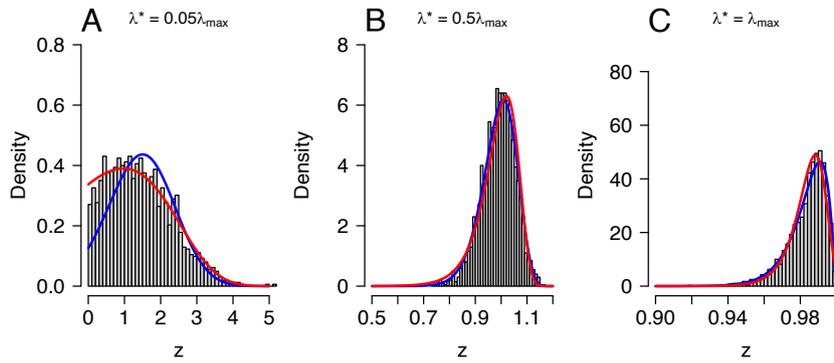
**2.1 Fitting Empirical DFEs to the Null Model.** We found that the null model matches empirical DFEs from microorganisms very well. Fig. 2 shows the fitted null model for  $z$  (i.e., Eq. 3) along with data for several organisms. The datasets include viruses and bacteria where the fitness of mutations has been measured in different environments. Further examples are provided in

*SI Appendix, Appendix 4*. For comparison, we also fitted a Gamma distribution to these data (blue curves in Fig. 2), because this is the prediction from Fisher's geometric model (20, 24). Consistent with previous studies, the Gamma distribution also fits very well. In some cases, however, we observe important discrepancies between the observed DFE and that predicted from both Fisher's geometric model and the null DFE, including some examples of bimodality in the bacterial data under certain growth conditions (*SI Appendix, Appendix 4*).

**2.2 Fitting DFE Simulated from the Fisher's Geometric Model (FGM).** We further investigated how easy it is to distinguish the true fitness landscape from the null model when the former is known. To do so, we simulated DFE under the FGM following the method of ref. 20 (*SI Appendix, Appendix 5*). Briefly, the FGM considers a  $n$ -dimensional phenotype in a landscape where fitness decreases smoothly (in its classical form, see ref.



**Fig. 2.** Fit of model predictions to empirical DFE. In all cases, the relative fitness was measured as  $z = r/r^*$ . Red: prediction of the null model (Gompertz distribution from Eq. 3), blue: shifted negative Gamma distribution (*SI Appendix, Eq. S20*). The distributions (parameters  $\alpha$  and  $\beta$  in Eq. 3 and  $y_0$ ,  $\alpha$ , and  $\beta$  in *SI Appendix, Eq. S20*) were fitted by maximum likelihood using the Nelder-Mead algorithm from "nmkn" function in package dfoptim (31) in R (32). (A–D) Compilation of DFE in viruses from ref. 7. (E) DFE of 3985 nonessential deletions from the Keio collection in *Escherichia coli* from ref. 33. (F) DFE of random base pair substitutions on the ribosomal protein S20 (growth media M9) in *Salmonella typhimurium* from ref. 34. See *SI Appendix, Appendix 4* for the full results.



**Fig. 3.** DFE of mutations simulated under the FGM, as measured by  $z = r/r^*$ . Simulation of 2,000 mutants from the reference genotype. Blue: Distribution expected under the FGM (*SI Appendix, Eq. S22*). Red: expectation under the null model, corresponding to a Gompertz distribution (Eq. 3). The parameters of the distribution were fitted by maximum likelihood similarly to Fig. 3. We varied the fitness of the reference phenotype  $\lambda^*$  relative to that of the optimum phenotype (set at 0)  $\lambda_{max} = 3.3$  (the value of  $\lambda_{max}$  is chosen from ref. 7), as indicated in each panel, while keeping the net strength of selection constant ( $S = 0.05$ ). (A–C) Depict increasing levels of adaptedness of the reference genotype from which mutations were generated. The number of dimensions of the phenotype is 5, with weak correlation ( $m = 1,000$ , see ref. 20).

24) with the distance to an optimum phenotype (*SI Appendix, Appendix 5*). In the FGM theory, selection is usually measured as  $s = \lambda/\lambda^* - 1$  e.g., ref. (20). However, we here choose to follow empirical studies (see above) by measuring relative fitness of mutants as  $z = r/r^*$ .

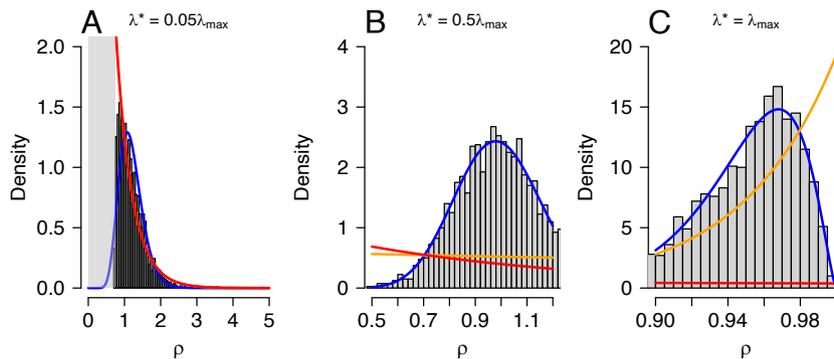
When relative fitness is measured by  $z$ , we found that the DFE expected under the null model (Eq. 3) often matches DFE simulated from the FGM very well (Fig. 3, red curves). For comparison, Fig. 3 also shows the fit with a shifted negative Gamma distribution as expected under the FGM (blue curves). Interestingly, a Gompertz distribution (Eq. 3) can be better fitted to the simulated DFE obtained when the reference phenotype is poorly adapted than a Gamma distribution [predicted under the FGM when the reference phenotype is well adapted (20); Fig. 3A].

As a further investigation of the relationship between the null model predictions and FGM, we estimated the fitness of the reference genotype and the mean relative fitness in the population from the fitted Gompertz distribution parameters (*SI Appendix, Table S2*). These closely match those used in the simulated FGM data when the reference genotype is poorly adapted (Fig. 3A). The estimated values from the null model diverge from those used in the FGM simulations when the reference phenotype gets closer to the optimum (Fig. 3 B and C). However, the null model nevertheless captures qualitatively the expectation that the mean

relative fitness decreases with increasing fitness of the reference genotype (Fig. 3 A–C and *SI Appendix, Table S2*).

Some discrepancies between the null model and the FGM arise when considering the ratio of the multiplicative growth rates,  $\rho$ , instead of the ratio of the Malthusian growth rates,  $z$  (Fig. 4). When the wild type is very far from the optimum in FGM the resulting DFE for relative multiplicative fitness tends to be monotonically decreasing and the null model from Eq. (2) captures this pattern very well (Fig. 4A). Similarly, when the wild type is at the optimum in FGM the resulting DFE for relative multiplicative fitness tends to be monotonically increasing, and then the (general) null model with both minimum and maximum fitness bounds from *SI Appendix, Appendix 3 and Eq. S14* can capture this pattern (Fig. 4C, orange curve). However, when the wild type is an intermediate distance from the optimum in FGM, the resulting DFE tends to have a bell shape as a result of the assumption that mutations deviate from the wild type according to a Gaussian distribution (Fig. 4B). The general null DFE for *multiplicative* relative fitness never has a bell shape (Fig. 4).

This large qualitative difference between model fits for multiplicative fitness measures versus the ratio of Malthusian growth rates might at first be surprising since the two are related to one another via a simple log transform. However, it can be shown that the ratio of Malthusian growth rates tends to be unimodal



**Fig. 4.** DFE of mutations simulated under the FGM, as measured by  $\rho = \lambda/\lambda^*$  and corresponding to Fig. 3. Blue: Distribution expected under the FGM (*SI Appendix, Eq. S21*, the parameters are fitted by maximum likelihood similarly to Fig. 2). Red: expectation under the null model without an upper fitness bound, which is a shifted exponential distribution (Eq. 2, where  $a$  is calculated from  $\lambda^*$  and  $\bar{\rho}$  is calculated from the simulated distribution). Orange: expectation under the null model bounded to the maximum fitness (*SI Appendix, Appendix 3 and Eq. S14*) in the simulated Fisher's landscape, where  $a = \min(\rho)$ ,  $b = \max(\rho)$  (see Fig. 1), and we numerically calculated the constant  $\kappa_2$  in *SI Appendix, Eq. S14* such that  $\int \rho f(\rho) d\rho = \bar{\rho}$  with  $\bar{\rho}$  calculated from the simulated distribution. In (A), the red and orange curves are identical.

under a wide variety of different forms for multiplicative fitness, not just the exponential form used here. The nonlinearity of the log transform stretches out the density lying in the interval (0, 1) over the entire negative axis and this (apparently) tends to make distinguishing between different distributions more difficult on this scale.

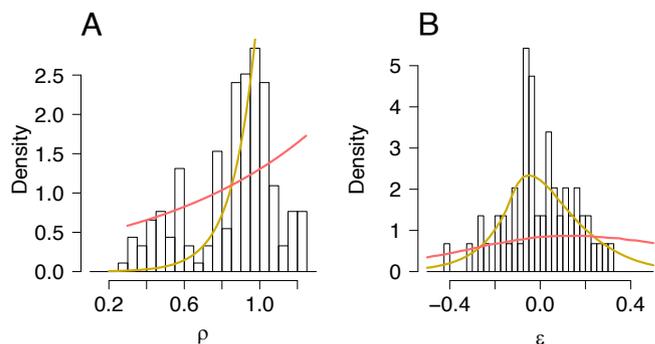
**2.3 Null Model Predictions for Epistasis.** Landscape models, like Fisher's geometric model, have also been used to make predictions about the pattern of epistasis among new mutations. For example, Fisher's geometric model predicts that pairwise epistasis among new mutations should follow a bell-shaped distribution with a mean close to zero. Moreover, it predicts that the mean should be negative when the wild type is poorly adapted and it should tend toward positive when the wild type is well adapted (21). These predictions tend to agree with documented empirical patterns (e.g., refs. 35 and 36).

We can also use the results presented here to generate null model predictions for the patterns of epistasis. To do so, we first take two draws from the null DFE [2] to represent the relative multiplicative fitness of two single mutations (denoted  $\rho_A$  and  $\rho_B$ ). Then, because the null model does not impose any relationship among the genotypes within the set  $G_m$ , the double mutant is simply a third independent draw,  $\rho_{AB}$ , from this same distribution. A commonly used measure of epistasis is then  $\epsilon = \ln(\frac{\rho_{AB}}{\rho_A\rho_B}) = \ln \rho_{AB} - \ln \rho_A - \ln \rho_B$ . It can be shown

that the probability density of  $x = \ln \rho$  is  $p(x) = \frac{e^x e^{(a-e^x)/(\bar{\rho}-a)}}{\bar{\rho}-a}$  for  $x \in (\ln a, \infty)$  and  $p(x) = 0$  otherwise. Denoting the pdf of  $\epsilon$  by  $h(\epsilon)$  we then have  $h(\epsilon) = \int_{\Omega} p(x)p(y)p(\epsilon+x+y)dx dy$  where  $\Omega$  is the set of all values of  $x$  and  $y$ . The null distribution of epistasis,  $h(\epsilon)$ , has a bell shape, and it can be proved that the mean epistasis  $\mathbb{E}[\epsilon]$  is positive if the average fitness of mutations is less than the reference genotype (i.e.,  $\bar{\rho} < 1$ ), and it is negative if the average fitness of mutations is large enough relative to the reference genotype. The reason for this is most easily understood by considering an alternative definition of epistasis that is not log-transformed:  $\hat{\epsilon} = \rho_{AB} - \rho_A\rho_B$ . By independence, the mean of  $\hat{\epsilon}$  is  $\mathbb{E}[\hat{\epsilon}] = \bar{\rho}(1 - \bar{\rho})$ , showing that epistasis is positive if  $\bar{\rho} < 1$  and it is negative if  $\bar{\rho} > 1$ . Also notice that if the reference genotype used for fitness comparison with the mutations is the wild type, then the distribution of deleterious mutations will tend to display positive epistasis, whereas the distribution of beneficial mutations will tend to display negative epistasis, in agreement with empirical studies.

The above findings about epistasis can be intuited as follows. Under the null model, the fitness of the double mutant is randomly assigned in a way that is independent and identically distributed to that of the fitness of each single mutant. As a result, if we consider the set of deleterious mutations (i.e., those mutations whose fitness is smaller than that of the wild type; i.e.,  $\rho = \lambda/\lambda_{WT} < 1$ ), then the fitness of the double mutant will (on average) be higher than that expected based on the multiplied effects of the two single deleterious mutants (i.e., there will be positive epistasis). On the other hand, if we consider the set of beneficial mutations (i.e., those mutations whose fitness is larger than that of the wild type; i.e.,  $\rho = \lambda/\lambda_{WT} > 1$ ), then the fitness of the double mutant will (on average) be lower than that expected based on the multiplied effects of the two single beneficial mutants (i.e., there will be negative epistasis).

Fig. 5 illustrates how the null model compares with data from the vesicular stomatitis virus for both the DFE of relative multiplicative fitness  $\rho$  and the distribution of epistasis,  $\hat{\epsilon}$ . For



**Fig. 5.** Fit of the null model to the DFE (A) and epistasis (B) generated with 47 mutations in the vesicular stomatitis virus (data from ref. 37, and details therein). Fitness of mutant  $i$  is measured relative to a reference genotype by  $\rho_i$  and epistasis between mutations  $i$  and  $j$  is measured by  $\hat{\epsilon}_{ij} = \rho_{ij} - \rho_i\rho_j$ . However, the data from ref. 37 are for the values of  $\rho^{1/\tau}$  for a characteristic generation time  $\tau$  (SI Appendix, Appendix 1). We have no way to determine  $\tau$  and so we have arbitrarily chosen  $\tau = 2$  for the purpose of illustration. (A) The fitness effects of mutations from ref. 37. Red: SI Appendix, Eq. S14 is fitted to the distribution by maximum likelihood using  $a = \min(\rho)$  and  $b = \max(\rho)$  ( $\kappa_2 = 1.145$ , see caption of Fig. 4). Yellow: DFE from SI Appendix, Eq. S14 when the distribution of epistasis is better matched to data ( $a = 0.2$ ,  $b = 1.1$ , and  $\kappa_2 = 8$ ). (B) Observed distribution of epistasis and two distributions of epistasis generated from the null model corresponding to the curves in panel (A). Distributions of  $\hat{\epsilon}$  from the null model are generated by drawing three fitness values from the red and yellow DFE in (A) and using the above definition of  $\hat{\epsilon}$  (repeated  $10^6$  times).

this dataset, it appears that the wild type has an intermediate level of adaptation, and so, the DFE is unimodal. As a result, the null DFE does not fit the data particularly well (red curves in Fig. 5). On the other hand, it is possible to fit the null distribution of epistasis to the data remarkably well (Fig. 5B, yellow). Doing so, however, requires that the fit of the DFE be altered substantially (Fig. 5A, yellow).

### 3. Discussion

Fitness landscape models have been developed to predict the properties of empirically observed DFEs. Some of these models (e.g., Fisher's geometric model) produce predictions that match some of the general properties of empirical DFEs very well (but see, e.g., ref. 26). Although such models often depend on parameters that do not have a clear biological interpretation, this fit between theory and data has been taken as an indication that the models capture the fundamentally important biology involved in mapping the process of mutation to fitness.

But how good of a fit is good enough? For example, should an excellent fit between Fisher's geometric model and the empirical DFE from *E. coli* in Fig. 2E be taken as evidence that this model captures important features of the map between mutation and fitness in *E. coli*? In other words, does this excellent fit mean that when mutations occur in *E. coli* they cause small changes in a set of quantitative traits, that intermediate values of each trait have the highest fitness according to a Gaussian function, and that there are no other significant fitness consequences of mutation? Informally, if the model predictions fit the empirical DFE "surprisingly well" then we might have strong confidence in this conclusion. On the other hand, if the model predictions did not fit the data any better than if the mapping of mutation to fitness were completely random, then we likely would have much less confidence that the model predictions were informative. The purpose of this paper is to provide a quantitative analysis of this "completely random" null model for the DFE in order to

help assess the strength of evidence supporting various fitness landscape models.

The null model constructed here assumes that the genotype-fitness map is chosen randomly, in an unbiased way, from the set of all possible maps. Under this assumption, we then proved that the distribution of fitness values that results (i.e., the DFE) is the distribution that maximizes information entropy. We have also then shown that if the mean fitness value of the set of mutations generated from a wild type is fixed by the environment and the reference genotype against which fitness is measured, then the null DFE for a commonly used index of relative fitness is a Gompertz distribution. Intuitively, under the null model, the relative fitness of a genotype generated randomly by mutation can take any value with equal probability, subject to the constraint that the mean relative fitness of all such genotypes is fixed. In a sense, mutant genotypes are thus spread maximally on the relative fitness surface subject to the constraint that they have a fixed mean.

We compared published data on the DFE of several different microorganisms including viruses and bacteria with the predictions of the null model. Overall the null model matched the data remarkably well, and typically just as well as landscape models such as Fisher's geometric model. This suggests that DFEs alone contain limited information for inferring the nature of the genotype-fitness map. We also showed that the null model matches the predictions of simulations from Fisher's geometric model very well, meaning that the DFE predictions from Fisher's geometric model are very similar to what would be expected if, instead, fitness values were assigned randomly to genotypes.

Interestingly, landscape models have also been used to make predictions about other properties of new mutations such as epistasis. It has been shown, for example, that the pairwise epistasis of new mutations is expected to follow a bell-shaped distribution with a mean close to zero, but that is somewhat negative when the wild type is far from the optimum and somewhat positive when the wild type is close to the optimum (21). We have shown that the null model makes these same predictions as well. Under the null model, the fitness of a double mutant is randomly assigned in a way that is independent and identically distributed to that of the fitnesses of each single mutant. As a result, if the wild type is close to the optimum then new mutations will, on average, be deleterious, and so the fitness of the double mutant will be higher than that expected based on the combined effects of the two single mutants (yielding positive epistasis). Likewise, if the wild type is far from the optimum then new mutations will, on average, be beneficial, and so the fitness of the double mutant will be lower than that expected based on the combined effects of the two single mutants (yielding negative epistasis). This null expectation for patterns of epistasis arises for reasons closely related to the phenomenon of "regression to the mean."

What implications do these results have for the study of DFEs? Perhaps most importantly, our results suggest that an examination of DFEs alone is not a very powerful approach for understanding how mutations get mapped to fitness. Several different genotype-to-fitness maps are consistent with observed DFEs, meaning that broad patterns of the fitness effects of mutations are not diagnostic of underlying mechanisms. However, this should not be taken as an undermining of the utility or value of landscape models. Indeed, the purpose of the null model is to help determine whether a fit between a particular fitness landscape model and data is better than one might expect solely through chance (i.e., solely by assigning fitness values to genotypes through a randomly chosen genotype-fitness map).

Another benefit of having an explicit, quantitative, null model is that it can be used to help determine the kinds of analyses that are better suited for assessing the fit of landscape models. To this end, our results suggest that patterns in the DFE are likely to be most informative when fitness is measured on a multiplicative scale and when the wild type has an intermediate level of adaptation. Furthermore, fitness landscape models such as FGM also make predictions about how the DFE changes during the process of adaptation. For example, the amount of epistasis is predicted to change during adaptation, as the fitness of the wild type increases (23). Empirical investigations focused on such changes might provide a more powerful way to distinguish among different models (see, e.g., ref. 38). Lastly, the results in Fig. 5 also suggest that examining the fit of models to multiple different quantities simultaneously, like both the DFE and the distribution of epistasis, is a more powerful approach. Measuring the DFE in different environments can also provide more information on the underlying fitness landscape than measurements in a single environment (39). Looking forward, however, it also seems likely that studies aimed at directly testing different mechanisms will be needed to unambiguously determine the causes driving patterns of fitness in mutations.

Ideas related to entropy from statistical physics have been used previously by evolutionary biologists in several different contexts. For example, such ideas have been used to analyze problems in multilocus population genetics, where the dimensionality of the system being studied is very large (see for example a review in refs. 40 and 41). A common analogy in this context between statistical physics and population genetics is that phenotypes (including fitness) are observable "macrostates" that arise from a combination of numerous and often unknown factors (e.g., genes) referred to as "microstates." Barton and Coe (42) have shown that, all else equal, we expect populations to evolve to a phenotypic (macro)state that is produced by the largest number of combinations of allelic (micro)states. In other words, subject to any external constraints (e.g., selection) the population ends up displaying a phenotype corresponding to the set of microstates that are most numerous. This is similar to our characterization of the null DFE as the fitness distribution that results from the largest number of genotype-fitness maps, subject to the constraint that this DFE has a fixed mean relative fitness. Indeed, a similar approach could be developed for the mapping of genotype to any phenotype of interest. Our derivation showing that the null distribution is that which maximizes information entropy would still hold, but the resulting distribution might differ from that of the null DFE depending on the values that are possible for the phenotype of interest, as well as the restrictions on the statistical properties of  $G_m$  that are imposed. Interestingly, empirical genotype-to-phenotype maps often find that the dimensionality of genotype space is much larger than that of phenotype space (e.g., ref. 43). If many genotypes express the same phenotype, they thus have the same fitness in a given environment. Such genotype-to-fitness redundancies therefore support the argument for building a null model which relies on the observation that the number of GP maps grows faster than the number of fitness (phenotype) distributions when the number of genotypes increases.

In summary, we have derived a null model for the DFE and showed that the null DFE maximizes information entropy. This null model is a potentially useful tool for generating reference predictions about what the DFE should be like if the genotype-fitness map were chosen at random from the set of all possible maps. This, in turn, can help to determine how empirical DFEs

can be used to distinguish among various mechanistic fitness landscape models.

**Data, Materials, and Software Availability.** All study data are included in the article and/or [supporting information](#). Previously published data were used for this work (7, 33, 34, 44).

1. D. M. Weinreich, N. F. Delaney, M. A. DePristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
2. R. D. Barrett, D. Schluter, Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44 (2008).
3. W. G. Hill, Understanding and using quantitative genetic variation. *Philos. Trans. R. Soc. B: Biol. Sci.* **365**, 73–85 (2010).
4. S. T. Schultz, M. Lynch, Mutation and extinction: The role of variable mutational effects, synergistic epistasis, beneficial mutations, and degree of outcrossing. *Evolution* **51**, 1363–1371 (1997).
5. S. P. Otto, T. Lenormand, Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**, 252–261 (2002).
6. T. Bataillon, S. F. Bailey, Effects of new mutations on fitness: Insights from models and data. *Ann. N. Y. Acad. Sci.* **1320**, 76–92 (2014).
7. R. Sanjuán, Mutational fitness effects in RNA and single-stranded DNA viruses: Common patterns revealed by site-directed mutagenesis studies. *Philos. Trans. R. Soc. B: Biol. Sci.* **365**, 1975–1982 (2010).
8. A. Couce, M. Magnan, R. E. Lenski, O. Tenaillon, The evolution of fitness effects during long-term adaptation in bacteria. [bioRxiv \[Preprint\] \(2022\)](https://doi.org/10.1101/2022.05.17.492360). <https://doi.org/10.1101/2022.05.17.492360> (Accessed 9 January 2023).
9. D. Aggeli, Y. Li, G. Sherlock, Changes in the distribution of fitness effects and adaptive mutational spectra following a single first step towards adaptation. *Nat. Commun.* **12**, 1–14 (2021).
10. A. Eyre-Walker, P. D. Keightley, The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).
11. R. MacLean, G. Perron, A. Gardner, Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* **186**, 1345–1354 (2010).
12. D. L. Halligan, P. D. Keightley, Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Syst.* **40**, 151–172 (2009).
13. J. Chen, T. Bataillon, S. Glémin, M. Lascoux, What does the distribution of fitness effects of new mutations reflect? Insights from plants. *New Phytol.* **233**, 1613–1619 (2021).
14. I. Fragata, A. Blanckaert, M. A. D. Louro, D. A. Liberles, C. Bank, Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* **34**, 69–82 (2019).
15. J. F. Kingman, A simple model for the balance between selection and mutation. *J. Appl. Prob.* **15**, 1–12 (1978).
16. S. A. Kauffman, E. D. Weinberger, The NK model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989).
17. T. Aita *et al.*, Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: Application to prolyl endopeptidase and thermolysin. *Biopolym.: Orig. Res. Biomol.* **54**, 64–79 (2000).
18. H. A. Orr, The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. *Evolution* **52**, 935–949 (1998).
19. H. A. Orr, Adaptation and the cost of complexity. *Evolution* **54**, 13–20 (2000).
20. G. Martin, T. Lenormand, A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**, 893–907 (2006).
21. G. Martin, S. F. Elena, T. Lenormand, Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* **39**, 555–560 (2007).
22. P. A. Gros, H. Le Nagard, O. Tenaillon, The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. *Genetics* **182**, 277–293 (2009).
23. F. Blanquart, G. Achaz, T. Bataillon, O. Tenaillon, Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* **68**, 3537–3554 (2014).
24. O. Tenaillon, The utility of Fisher's geometric model in evolutionary genetics. *Annu. Rev. Ecol. Syst.* **45**, 179–201 (2014).
25. G. Martin, Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. *Genetics* **197**, 237–255 (2014).
26. F. Blanquart, T. Bataillon, Epistasis and the structure of fitness landscapes: Are experimental fitness landscapes compatible with Fisher's geometric model? *Genetics* **203**, 847–862 (2016).
27. S. A. Frank, Generative models versus underlying symmetries to explain biological pattern. *J. Evol. Biol.* **27**, 1172–1178 (2014).
28. N. J. Gotelli, B. J. McGill, Null versus neutral models: What's the difference? *Ecography* **29**, 793–800 (2006).
29. I. Sanov, On the probability of large deviations of random variables. *Mat. Sb.* **42** (84), 11–44 (1957).
30. T. M. Cover, J. A. Thomas, Information theory and statistics. *Elem. Inf. Theory* **1**, 279–335 (1991).
31. R. Varadhan *et al.*, *dfoptim: Derivative-Free Optimization* (R package version 2020.10-1, 2020) <https://cran.r-project.org/web/packages/dfoptim/index.html>.
32. R Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2021).
33. G. Chevereau *et al.*, Quantifying the determinants of evolutionary dynamics leading to drug resistance. *PLoS Biol.* **13**, e1002299 (2015).
34. P. A. Lind, O. G. Berg, D. I. Andersson, Mutational robustness of ribosomal protein genes. *Science* **330**, 825–827 (2010).
35. H. H. Chou, H. C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
36. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011).
37. R. Sanjuán, A. Moya, S. F. Elena, The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15376–15379 (2004).
38. A. Couce, O. A. Tenaillon, The rule of declining adaptability in microbial evolution experiments. *Front. Genet.* **6**, 99 (2015).
39. H. Kemble *et al.*, Flux, toxicity, and expression costs generate complex genetic interactions in a metabolic pathway. *Sci. Adv.* **6**, eabb2236 (2020).
40. G. Sella, A. E. Hirsh, The application of statistical physics to evolutionary biology. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9541–9546 (2005).
41. H. P. De Vladar, N. H. Barton, The contribution of statistical physics to evolutionary biology. *Trends Ecol. Evol.* **26**, 424–432 (2011).
42. N. H. Barton, J. Coe, On the application of statistical physics to evolutionary biology. *J. Theor. Biol.* **259**, 317–324 (2009).
43. S. E. Ahnert, Structural properties of genotype-phenotype maps. *J. R. Soc. Interface* **14**, 20170275 (2017).
44. J. D. Mehlhoff *et al.*, Collateral fitness effects of mutations. *Proc. Natl. Acad. Sci. U.S.A.* **117**, e1918680117 (2020).

**ACKNOWLEDGMENTS.** We thank Mike Whitlock and Aneil Agrawal for helpful discussions and Sylvain Gandon and François Blanquart for many important comments on a previous version of this manuscript. We also thank Olivier Tenaillon and two anonymous reviewers for helping us to clarify our ideas and arguments. This work was funded by a Natural Sciences and Engineering Research Council of Canada grant to T.D. and O.C.