The distribution of epistasis on simple fitness landscapes
Christelle Fraïsse, John J Welch

To cite this version:

HAL Id: hal-03821403
https://hal.archives-ouvertes.fr/hal-03821403
Submitted on 21 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The distribution of epistasis on simple fitness landscapes

Christelle Fraïssé$^{1,2,3}$ and John J. Welch$^2$

October 21, 2022

1. Institut des Sciences de l’Evolution, CNRS-UM-IRD, Montpellier, France.
2. Department of Genetics, University of Cambridge, Downing St. Cambridge, CB23EH, UK.
3. Institute of Science and Technology Austria, Am Campus 1, Klosterneuburg 3400, Austria.

* Author for correspondence: christelle.fraise@ist.ac.at
Abstract

Fitness interactions between mutations can influence a population’s evolution in many different ways. While epistatic effects are difficult to measure precisely, important information is captured by the mean and variance of log fitnesses for individuals carrying different numbers of mutations. We derive predictions for these quantities from a class of simple fitness landscapes, based on models of optimizing selection on quantitative traits. We also explore extensions to the models, including modular pleiotropy, variable effects sizes, mutational bias, and maladaptation of the wild-type. We illustrate our approach by reanalysing a large data set of mutant effects in a yeast snoRNA. Though characterized by some large epistatic effects, these data give a good overall fit to the non-epistatic null model, suggesting that epistasis might have limited influence on the evolutionary dynamics in this system. We also show how the amount of epistasis depends on both the underlying fitness landscape, and the distribution of mutations, and so is expected to vary in consistent ways between new mutations, standing variation, and fixed mutations.

Keywords

Fitness landscapes; genetic interactions; Fisher’s geometric model; *Saccharomyces cerevisiae*
Introduction

Fitness epistasis occurs when allelic variation at one locus affects allelic fitness differences at other loci. Epistatic interactions can be used to uncover functional interactions [1], but for other questions, the most important quantity is the complete distribution of epistatic effects. The shape of this distribution can affect a population’s ability to adapt, its genetic load, the outcomes of hybridization, and the evolution of recombination rate, or investment in sexual reproduction [2-13].

To investigate such questions, most research has focused on the mean level of epistasis. This can be estimated from the rate at which mean log fitness declines with the number of mutations carried [7,14-17], which is simple to model [2,4,9,18,19]. But variation around this mean can also affect the evolutionary dynamics [6,7,17].

To understand the complete distribution of effects, one approach is to use Fisher’s geometric model [20], a simple model of optimizing selection on quantitative traits [10,12,21,22]. This is a toy model, but it approximates a broad class of systems biology models, involving metabolic networks [21]. Furthermore, it naturally generates fitness epistasis; and the overall level of epistasis can be “tuned” by adjusting the curvature of the fitness function, that is, the rate at which fitness declines with distance from the optimum [10-12,23-28].

Because it generates a rich spectrum of effects with few parameters, Fisher’s geometric model is particularly suitable for data analysis [24,29-31], including data on fitness epistasis [32-36]. Perhaps most impressively, Martin et al. [32] used the model to successfully predict several properties of the distribution of epistatic effects in the microbes Escherichia coli and Vesicular Stomatitis Virus [15,37]. However, these authors did not directly study the effects of varying the shape of the fitness landscape, nor other possible variants of the model [25,38-41]. Here, following [32], we study properties of fitness epistasis under Fisher’s geometric model. We extend previous results by examining a wider class of fitness landscapes, and also compare the predictions to a recent, large-scale data set of yeast mutants [1].

Models and analysis

Basic notation and a null model without epistasis

Let us denote as $\ln w_d$, the log relative fitness of an individual carrying $d$ mutations. Across many individuals, its scaled mean and variance are

\[ m(d) \equiv \frac{E(\ln w_d)}{E(\ln w_1)} \]

\[ v(d) \equiv \frac{\text{Var}(\ln w_d)}{\text{Var}(\ln w_1)} \]

where, by definition, $m(0) = v(0) = 0$ and $m(1) = v(1) = 1$. These equations use a log scale, because deviations from additivity on a log scale influence the evolutionary dynamics [7].

We can immediately give results for a null model with no epistatic effects. In this case, mutations will contribute identically to the mean and variance in fitness, regardless of how many other mutations are carried. So a collection of individuals carrying two random mutations is expected to have twice the decline in log fitness, and twice the variance in log fitness, as a collection of individuals carrying one mutation. This implies that
\[ m_0(d) = d \quad (3) \]
\[ v_0(d) = m_0(d) \quad (4) \]

where the subscript 0 indicates the non-epistatic null model. These predictions are illustrated by red lines in Figure 1.

To measure epistasis directly, we could measure the pairwise interaction between two mutations, denoted \( a \) and \( b \):

\[ \varepsilon \equiv \ln w_{ab} - \ln w_a - \ln w_b \quad (5) \]

Here, \( w_{ab} \) denotes the relative fitness of the genome carrying the mutation “\( ab \)”, and so on. Though widely used, \( \varepsilon \) can be difficult to work with. For example, if the same mutation appears in multiple double mutants, then the complete distribution of \( \varepsilon \) will entail using the same fitness measurements multiple times, creating complications from pseudoreplication or correlated errors. Furthermore, for a complete picture of epistasis, we would also have to consider higher-order interactions between three or four mutations. For these reasons, we focus on eqs. 1-2, and give some equivalent results for \( \varepsilon \) in Appendix 1. The quantities are also closely related. For example, eq. 3 implies that there is no epistasis on average (i.e., that positive effects exactly match negative effects, such that \( E(\varepsilon) = 0 \)), while eq. 4 implies that all epistatic effects are the same, such that \( \text{Var}(\varepsilon) = 0 \) (see Appendix 1). Together, then, eqs. 3-4 imply that there is no epistasis at all.

**Additive phenotypic models**

Under Fisher’s geometric model, an individual’s fitness depends on its values of \( n \) quantitative traits: \( z = \{z_1, z_2, \ldots, z_n\} \). Fitness depends on the deviation of the phenotype from a single optimal value. A suitable fitness function of this kind is

\[ \ln W(z) \propto - ||z||^k \quad (6) \]

where \( ||z|| \equiv \sqrt{\sum_{i=1}^{n} z_i^2} \) [25, 26]. An alternative, which does not assume identical selection on all traits, is

\[ \ln W(z) \propto - \sum_{i=1}^{n} \lambda_i |z_i|^k \quad (7) \]

where \( \lambda_i \) determines the strength of selection on trait \( i \) [23, 24]. These two fitness functions often give similar results (Figures S1-S2), but they are identical only when \( k = 2 \), and all \( \lambda_i \) are equal.

The simplest models then further assume that: (1) that the wild-type is phenotypically optimal; (2) that mutations are additive with respect to the phenotype, and (3) that the mutant effects on each trait are drawn independently, from a standard normal distribution.

In Appendix 1, we show that these assumptions yield the following results, as illustrated by the black lines in Figure 1:
\[ m(d) = d^{k/2} \]  
\[ v(d) = m^2(d) \]  

Equations 8-9 show how \( k \) affects the level of fitness epistasis [23,26]. When \( k = 2 \), we have no epistasis on average, as with eq. 3 (solid black in lines in Fig. 1a-b). Setting \( k > 2 \) leads to negative epistasis on average (dashed black in lines in Fig. 1a-b), and \( k < 2 \) leads to positive epistasis on average (dotted black in lines in Fig. 1a-b). Note also, that eq. 9 will never agree with eq. 4, because these simple phenotypic models always generate fitness epistasis.

Confronted with data from real quantitative traits [42], many aspects of the models above appear grossly unrealistic. In Appendix 1, we explore several variants of the model, designed to relax its simplifying assumptions. We show that altering the distribution of single mutation effects can have a major effect on results, greatly altering the relationship between \( m(d) \) and \( k \). However, in each case, we also show that this leaves a signature in the variance, such that-

\[ v(d) < m^2(d) \quad d > 1 \]  

This is illustrated by the coloured lines in Figure 1.

Reanalysis of data from a yeast snoRNA

To illustrate the approach above, and compare different measures of epistasis, we now reanalyse the data of Puchta et al. [1], who used saturation mutagenesis of the U3 snoRNA in \textit{Saccharomyces cerevisiae} (see Appendix 2 for full details). Figure 2a confirms that pairwise epistatic interactions are present in these data [1]. Nevertheless, Figure 2c-d show that, considered as a whole, the data give a very good fit to the non-epistatic null model (eqs. 3-4).

Some of this apparent discrepancy can be attributed to the greater robustness of our statistics to measurement error. For example, we show in supplementary Figures S4 and S5, that the inferred variance in epistatic effects decreases with the amount of replication, while patterns in \( m(d) \) and \( v(d) \) are little changed. Furthermore, some reduction in epistasis could have been predicted from other aspects of the data, which violate the assumptions behind eqs. 8-9 (see Appendices for full details). For example, the distribution of single-mutant fitnesses (Figure 2b) is inconsistent with the assumption of normal effects, and the presence of beneficial mutations (346/965 mutations increase growth rate) is inconsistent with the assumption of an optimal wild-type. Nevertheless, we argue in Appendix 1 that the phenotypic models - even in modified form - overestimate the true amount of fitness epistasis in these data. This implies that simple models of independent effects might be sufficient to understand several aspects of the evolutionary dynamics in this system, despite the clear presence of some fitness interactions [1].

Discussion

We have used simple summary statistics to describe levels of fitness epistasis. These statistics are relevant to evolutionary questions [7], and are less sensitive to measurement error than are estimates of individual epistatic effects.

We then developed analytical predictions for these statistics under simple models of quantitative traits selected towards a single optimum. The simplest such model assumes that mutant effects on each trait are i.i.d. normal, and considered as a model of quantitative traits, this seems unrealistic [39,42-44]. Nevertheless, considered as a
fitness landscape, the same model has been shown to give a good fit to fitness data from *E. coli* and VSV [15,32,37]. Our results go further, and show that only this simple model would have fit those data; increasing the realism of the quantitative traits (e.g., by introducing leptokurtic effects), would have underpredicted the amount of epistasis. This reinforces the argument of [21], that the “traits” in Fisher’s geometric model, when considered as a fitness landscape, should not be equated with standard quantitative traits. On a related point, the good fit to the fitness data was obtained by assuming that $k = 2$ [15], and we have shown that no other value of $k$ could have given a comparable fit. This has implications for the evolution of epistasis, because multiple authors have shown that models with no epistasis on average (i.e., with $k = 2$), are vulnerable to invasion by modifiers [26,45,46]. As such, the good fit of $k = 2$ implies that global modifiers of fitness epistasis do not arise in these systems.

Of course, we cannot assume that identical patterns of epistasis will characterise all data sets [47,48], and we have offered two further reasons to doubt this. First, empirically, we have shown that the data of Puclta et al. [1] give a good overall fit to a non-epistatic null model, despite the likely presence of some fitness interactions ([1]; Figure 2). Second, theoretically, we have shown how the observed level of epistasis will depend on both the underlying fitness landscape, and the distribution of mutation effects. For example, a landscape with a high level of curvature (i.e., $k > 2$), might still generate a linear decline in mean log fitness (such that $m(d) \approx d$) if the distribution of mutant effect sizes is highly leptokurtic (see Appendix 1); but this effect should be evident in the reduced levels of variance (eq. 10). Finally, if mutations of very large or very small effect are less likely to contribute to adaptation, then the fixation process acts to restrict the distribution towards mutations of medium size [38]. As such, the levels of observed epistasis should increase steadily for new mutations, standing variation, and differences that are fixed between populations.

Ethics

Not applicable.

Data accessibility

Simulation code is provided as Supplementary Information. The yeast data are available in reference [1].

Authors’ contributions

Both authors designed the study, analysed the data and wrote the manuscript. JW carried out the modelling. All authors agree to be held accountable for the content therein and approve the final version of the manuscript.

Competing interests

We declare no competing interests.

Funding

CF is supported by an IST fellowship (Marie Skłodowska-Curie Co-Funding European program).
Acknowledgements

We are very grateful to Grzegorz Ku kla, Anna Puchta, Elena Kuzmin, Santiago Elena and Rafael Sanjuan for providing their data, and helpful clarifications. We are also grateful to Guillaume Martin, Denis Roze, Nicolas Bierne, Chris Illingworth, and Fyodor Kondrashov for useful discussions. The authors thank the editors and two anonymous reviewers for insightful comments on the article.

References


**Figure Legends**

**Figure 1**

Predictions for mean log fitness (a,c) or the standard deviation in log fitness (b,d). Upper panels show predictions for individuals carrying different numbers of mutations, $\hat{d}$. Lower panels show results for double mutants ($d = 2$), varying the curvature of the fitness landscape, $k$. Results for the null model, with no epistasis, are shown as red.
dashed lines. In this case, the mean and variance in log fitness both change linearly with $d$ (eqs. 3-4). Results for simple phenotypic models are shown as black lines. The upper panels show results with no epistasis on average (solid lines, $k = 2$), negative epistasis on average (dashed lines, $k = 4$), or positive epistasis on average (dotted lines, $k = 1$). Blue lines show results for a model with strongly biased mutations ($\beta = 3$, $k = 2$; eqs. 48 and 50); these can be compared to the dashed line in (a) or the solid line in (b), which correspond to results with very large bias (e.g., eqs. 49 and 51). Green lines show results where the mutations on each trait are drawn from a leptokurtic reflected exponential distribution (eqs. 43).

**Figure 2**

Reanalysis of mutations in *Saccharomyces cerevisiae* U3 snoRNA [1]. (a) shows the distribution of pairwise epistatic effects (eq. 5), compared to the predictions of the simplest phenotypic model with $k = 2$: $\varepsilon \sim N(0, 2\text{Var} (\ln w_1))$ (black line; [32]; Appendix 1), and a normal distribution with matching mean and variance (dotted line). (b) shows the distribution of single mutant log fitnesses, and the best-fit shifted gamma distribution, as predicted by the simplest phenotypic models [29]. (c) shows the mean of the log fitnesses of individuals carrying $d$ mutations (black points with barely visible standard error bars); the median and 90% quantiles (grey points and bars); the analytical prediction, which applies to both the null model and the phenotypic model with $k = 2$ (black line; eqs. 3 and 8); and the best-fit regression for $\ln m(d) \sim \ln d$ (dotted line, which has a slope implying $k = 2.16$). (d) shows the standard deviation in the log fitnesses of individuals carrying $d$ mutations (black points with barely visible standard error bars); analytical predictions from the null model, eq. 4 (dashed line), or the phenotypic model with $k = 2$, eq. 9 (solid line); and the best-fit regression of $\ln v(d) \sim \ln d$ (dotted line, which has slope 0.89).