Appendices for: "The distribution of epistasis on simple fitness landscapes"

Christelle Fraïsse[1,2,3*] and John J. Welch[2]

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.fraisse@ist.ac.at

Appendix 1: Derivations

In this Appendix, we derive the key results in the main text. We also present results for extensions to the simplest phenotypic model, and for direct measures of pairwise epistasis (eq. 5).

1. The distribution of pairwise epistatic effects:

Martin et al. [1] examined the scaled moments of the distribution of pairwise epistatic effects (eq. 5). These moments are closely related to the scaled moments of the genotypic fitness values, $m(d)$ and $v(d)$, that we use in the main text (eqs. 1-2). To see this, let us consider individuals with a wild-type phenotype $\mathbf{z}$. The relative fitness of an individual carrying a single mutation, with phenotypic effects $\mathbf{x}$, is

$$\ln w_1 \equiv \ln W(\mathbf{z}+\mathbf{x}) - \ln W(\mathbf{z}) \tag{11}$$

This is closely related to the selection coefficient of the mutation, s, because when s is small in magnitude, $s \approx \ln(1+s) = \ln w_1$. This is why the quantity shown in eq. 11 is denoted as s by Martin et al. [1]. From eq. 5, the pairwise epistatic effect for two mutations, a and b, is then

$$\varepsilon = \ln W(\mathbf{z}+\mathbf{x}_a+\mathbf{x}_b) - \ln W(\mathbf{z}+\mathbf{x}_a) - \ln W(\mathbf{z}+\mathbf{x}_b) + \ln W(\mathbf{z}) \tag{12}$$

[1,2]. We can now use eqs. 1-2 to write the mean and variance of epistatic effects, scaled by the same quantities for single mutations:

$$\frac{E(\varepsilon)}{E(\ln w_1)} = m(2) - 2 \tag{13}$$

$$\frac{\mathrm{Var}(\varepsilon)}{\mathrm{Var}(\ln w_1)} = v(2) + 2 - 4\sqrt{v(2)}r_{12} \tag{14}$$

Here, we have defined

$$r_{12} \equiv \mathrm{Cor}\left(\ln W(\mathbf{x}_a+\mathbf{x}_b+\mathbf{z}), \ln W(\mathbf{x}_a+\mathbf{z})\right) \tag{15}$$

as the correlation coefficient between the log fitnesses of genotypes carrying a single mutation alone, and in combination with a second mutation. Under the null model, with no epistasis, the double mutant log fitness must be the sum of two i.i.d. random variables, describing the effects of each of the two mutations. Since $\mathrm{Cor}(x+y, y) = \sqrt{1/2}$ if x and y are i.i.d., it follows that $r_{12} = \sqrt{1/2}$ under the null model. With this value, $\mathrm{Var}(\varepsilon) = 0$ when $v(2) = 2$, justifying the assertion in the main text, that eq. 4 implies no variation in epistatic effects. The value $r_{12} = \sqrt{1/2} \approx 0.707$ for the null model can also be compared to results from other models below.

2. Results for the simplest phenotypic model:

Let us first consider results for the simplest model, when the wild-type is phenotypically optimal ($\mathbf{z} = \mathbf{0}$), and the effects of each mutation on each trait are drawn from independent standard normal distributions. In this case, the phenotype of an individual carrying d mutations can be written as

$$\mathrm{z} = \left\{ \sum_{j=1}^{d} x_{j1}, \sum_{j=1}^{d} x_{j2}, \ldots, \sum_{j=1}^{d} x_{jn} \right\} \tag{16}$$

where

$$x_{ji} \sim N(0,1). \tag{17}$$

If we use the fitness function of eq. 6 [2,3], which assumes equal selection on all n traits, then the quantities we require for eqs. 1-2 are simply moments of the Chi-squared distribution, with n degrees of freedom:

$$-E\left(\ln w_d\right) = (2d)^{k/2} \frac{\Gamma\left(\frac{k+n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \tag{18}$$

$$\mathrm{Var}\left(\ln w_d\right) = (2d)^{k} \left(\frac{\Gamma\left(\frac{2k+n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} - \frac{\Gamma^2\left(\frac{k+n}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)} \right) \tag{19}$$

The results of eqs. 8-9 follow directly.

If we allow for variation in the strength of selection between traits, and use the fitness function of eq. 7 [4], then the key quantities are now the moments of a folded normal distribution (i.e., the absolute value of a normally-distributed random variable).

$$-E\left(\ln w_d\right) = (2d)^{k/2} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi}} \sum_{i=1}^{n} \lambda_i \tag{20}$$

$$\mathrm{Var}\left(\ln w_d\right) = (2d)^{k} \left(\frac{\Gamma\left(\frac{2k+1}{2}\right)}{\sqrt{\pi}} - \frac{\Gamma^2\left(\frac{k+1}{2}\right)}{\pi} \right) \sum_{i=1}^{n} \lambda_i^2 \tag{21}$$

and again, eqs. 8-9 follow directly. Figure S1a confirms, with simulations, that the two fitness functions give identical results.

2.1 Pairwise epistatic effects

To calculate the variance in pairwise epistatic effects (eq. 14), we also require the correlation coefficient of eq. 15. For the fitness function of eq. 7, this is maximized at $k = 2$, where it takes the value:

$$r_{12} = \mathrm{Cor}\left(\left| x_{ai} + x_{bi} \right|^2, \left| x_{ai} \right|^2 \right) = \frac{1}{2}, \quad k = 2 \tag{22}$$

and so the correlation between single- and double-mutant fitnesses is always lower than under the null model. The same value holds approximately for other values of k, and for the alternative fitness function of eq. 6. As such, we have the results

$$-\frac{E(\varepsilon)}{E(\ln w_1)} = 2 - 2^{k/2} \tag{23}$$

$$= 0, \quad k = 2 \tag{24}$$

$$\frac{\mathrm{Var}(\varepsilon)}{\mathrm{Var}(\ln w_1)} \approx 2(1 + 2^{k-1} - 2^{k/2}) \tag{25}$$

$$= 2, \quad k = 2 \tag{26}$$

These results are compared to simulation in Figure S2. When $k = 2$, eqs. 24 and 26 reproduce the results of Martin et al. [1], while increasing k above this value makes expected levels of epistasis more negative ($E(\varepsilon) < 0$), and increases the variance in epistatic effects ($\mathrm{Var}(\varepsilon) > 2\mathrm{Var}(\ln w_1)$).

The complete distribution of ε is also derivable for $k = 2$, since we have

$$\varepsilon \propto \sum_i^n \lambda_i \xi_i, \qquad k = 2 \tag{27}$$

where $\xi_i \equiv x_{ai} x_{bi}$, and this has the pdf

$$\mathrm{pdf}(\xi) = \int_0^\infty \frac{\cos(|\xi| t)}{\pi\sqrt{t^2+1}} dt$$

which has a vanishing mean and unit variance. As shown in Figure S2, the mode of the distribution remains close to zero for all k values, meaning that variation in the curvature of the fitness landscape acts to skew the distribution of epistatic effects.

3. Extensions to the simplest phenotypic model

As stated in the main text, many assumptions of the model above appear highly unrealistic. For example, unless the number of traits is very small, the i.i.d. normal model suppresses mutations of overall small effect, and yet there is good reason to think that such mutations are very common [5-8].

More concerns arise if we take the model seriously as a model of quantitative traits, and consider data from real traits. For example, there is clear evidence that both selection and mutation are correlated among traits [9,10], and that mutations are characterised by highly leptokurtic distributions, with stronger concentrations of very small and very large effects; and bias, with a tendency to change traits in a particular direction [11,12]. Furthermore, there is some evidence of appreciable epistasis at the level of phenotype [13,14]; and restricted or modular pleiotropy, where mutations affect only a subset of traits ([5,15]; though see [16]). Finally, there is often evidence of beneficial mutations, which implies that the wild-types are suboptimal. None of this is consistent with eqs. 16-17.

Some of the simplifying assumptions are only apparent. For example, the major effect of correlations can often be transformed away, by redefining the axes, and considering a smaller "effective number of traits" [9,17,18]. Nonetheless, other assumptions are certainly restrictive. In the following sections, we explore several extensions of the model, building on the results of several previous studies [1,5,7,9,18-20], but focussing only on assumptions that can be relaxed in a general way. In particular, we consider variable distributions of effect sizes, restricted pleiotropy, mutational bias, and suboptimal wild-types. Together, the results support the assertions in the main text, namely that the relationship between $m(d)$ and k is altered by relaxing the assumptions, and secondly, that this also acts

to reduce the variance in epistasis, such that eq. 10 holds. Furthermore, we will show that when $k \geq 2$, results for the null model act as lower bound, such that $v(d) \geq m(d)$.

3.1. Modular pleiotropy and variable effects sizes

The first set of extensions are most easily made with the isotropic fitness function of eq. 6.

Let us first consider the effects of restricting pleiotropy. Instead of assuming that each mutation affects all n traits, we now assume that pleiotropy is modular ([11]; see also [20,21]), such that each new mutation affects a distinct "module" containing n' traits, which are under selection independently of other modules. To treat this case, consider the total length of the phenotypic effect for a double mutant. This can be written as:

$$\|\mathbf{x}_a + \mathbf{x}_b\| = \sqrt{\sum_i^n (x_{ai} + x_{bi})^2} = \sqrt{\|\mathbf{x}_a\|^2 + \|\mathbf{x}_b\|^2 + 2\|\mathbf{x}_a\|\|\mathbf{x}_b\|\cos(\theta)} \tag{28}$$

where θ is the angle in radians between the two mutational vectors, in the n-dimensional trait space [11]. If the mutations affect different modules, then their individual vectors will be orthogonal, such that $\cos(\theta) = 0$. Since the sum of Chi-squared random variables is also Chi-squared distributed, we require the moments of a Chi-squared distribution, with dn' degrees of freedom:

$$-E(\ln w_d) = 2^{k/2} \frac{\Gamma\left(\frac{k+dn'}{2}\right)}{\Gamma\left(\frac{dn'}{2}\right)}, \tag{29}$$

$$= dn', \qquad k = 2$$

$$\mathrm{Var}(\ln w_d) = 2^k \left(\frac{\Gamma\left(\frac{2k+dn'}{2}\right)}{\Gamma\left(\frac{dn'}{2}\right)} - \frac{\Gamma^2\left(\frac{k+dn'}{2}\right)}{\Gamma^2\left(\frac{dn'}{2}\right)} \right) \tag{30}$$

$$= 2dn', \qquad k = 2$$

When $k = 2$, these results immediately reproduce the null model (eqs. 3-4). We also have the approximation

$$\frac{v(d)}{m^2(d)} = \frac{\Gamma(dn'/2 + k)\,\Gamma(dn'/2)\,\Gamma^{-2}(dn'/2 + k/2) - 1}{\Gamma(n'/2 + k)\,\Gamma(n'/2)\,\Gamma^{-2}(n'/2 + k/2) - 1} \approx \frac{1}{d} \tag{31}$$

which is exact when $k = 2$, or in the limit as $n' \to \infty$. Using the Beta function, we also have the limits:

$$m(d) = \frac{B\left(\frac{n'}{2}, \frac{k}{2}\right)}{B\left(\frac{dn'}{2}, \frac{k}{2}\right)} \tag{32}$$

$$\to d^{k/2}, \quad n' \to \infty \tag{33}$$

$$\to d, \quad n' \to 0 \tag{34}$$

More complete models would have to specify the probability that a pair of mutations appears in the same

module, and also consider modules of different sizes. However, the results above are sufficient to show that $m(d)$ will be intermediate between the simple phenotypic model (eq. 8), and the null model (eq. 3), so that the influence of k is relaxed. The results also confirm that eq. 10 will hold. In both cases, this makes intuitive sense: epistatic interactions will be reduced when mutations sometimes affect independent sets of traits. Simulations with intermediate values of n' are shown in Figure S1b, and confirm the results above.

The influence of modular pleiotropy seems inherently implausible for the yeast data reanalysed here [22], because all of the mutations affect sites in the same snoRNA. Nevertheless, we now show that the effects of modular pleiotropy can also be replicated in a model with universal pleiotropy, if we allow for mutations of very different sizes. This is equivalent to assuming a highly leptokurtic distribution of effects on the overall size of mutations, and thereby on each trait. This is easiest to demonstrate by considering pairwise epistatic effects, when $k = 2$. In this case, we have

$$\varepsilon = 2 \left\| \mathbf{x}_a \right\| \left\| \mathbf{x}_b \right\| \cos(\theta), \qquad k = 2 \tag{35}$$

As shown by Fisher [23], when the number of traits, n, is not very small, then an unbiased distribution of mutation directions leads to $2\cos(\theta) \sim N(0, 4/n)$ [5,24]. As such, we have

$$E\left(\varepsilon\right) = 0, \qquad k = 2$$

$$\mathrm{Var}\left(\varepsilon\right) \approx \frac{4}{n} E\left(\left(\left\| \mathbf{x}_a \right\| \left\| \mathbf{x}_b \right\| \right)^2 \right), \qquad k = 2 \tag{36}$$

If we follow Lourenço et al. [20] and draw the squared mutation magnitudes from a Chi-squared distribution with n' degrees of freedom, then it follows that $\mathrm{Var}\left(\varepsilon\right) = \frac{4}{n} n'^2$. The excess kurtosis of the Chi-squared distribution is $12/n^{'}$ and so decreasing n' increases the kurtosis, and decreases the variance in epistatic effects. Simulation results, shown in Figure S1c and Figure S2c-d, show that the same general pattern holds for other values of k, and for other leptokurtic distributions of mutation sizes.

3.2. Varying the distribution of effects on each trait

In the previous section, we used a "top-down" approach to mutation, in which the vector size and direction were independently calculated [25]. The alternative, "bottom-up" approach is to directly specify the distribution of effects on individual traits. This is simplest with the fitness function of eq. 7, where analytical results can be obtained for double mutants, with $d = 2$.

Because the distribution of mutations on quantitative traits is often leptokurtic, let us first consider results when mutational effects are drawn from a reflected exponential distribution, with parameter μ. In this case, the absolute effect on a single trait, $|x|$, is exponentially distributed, such that

$$E[|x|^k] \quad = \quad k!\mu^k \tag{37}$$

$$\mathrm{Var}\left(|x|^k\right) \quad = \quad \mu^{2k}\left((2k)! - (k!)^2\right) \tag{38}$$

For quantities involving two mutations ($d = 2$), if their effects have the same sign, then we have an Erlang distribution:

$$E\left(|x_a + x_b|^k \,|x_a x_b > 0\right) \quad = \quad \mu^k \frac{\Gamma(2+k)}{\Gamma(2)} = \mu^k (k+1)! \tag{39}$$

If they have different signs, we have a difference in exponentials, whose pdf is

$$f(\delta) = \frac{2}{\mu^2} e^{-|\delta|/\mu} \tag{40}$$

$$E\left(|x_a + x_b|^k \,|x_a x_b < 0\right) = k!\mu^k \tag{41}$$

The signs differ with 50% probability, and so, combining these results, we have

$$\begin{aligned} E[|x_a + x_b|^k] &= \mu^k \left(\frac{(k+1)! + k!}{2}\right) = \mu^k k! \left(\frac{k+2}{2}\right) \\ \mathrm{Var}[|x_a + x_b|^k] &= \mu^{2k} \left[(2k)!\,(k+1) - \left(k! \left(\frac{k+2}{2}\right)\right)^2\right] \end{aligned} \tag{42}$$

and so, we find:

$$m(2) = 1 + \frac{k}{2}$$

$$v(2) = \frac{(2k)!\,(k+1) - (k!)^2 \left(\frac{k+2}{2}\right)^2}{(2k)! - (k!)^2} \approx 1 + k \tag{43}$$

where the approximate expression for $v(2)$ uses Stirling's approximation: $k! \approx \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$, such that $(2k)!/(k!)^2 \approx 2^{2k}/\sqrt{\pi k}$. The results are supported by simulations shown in Figure S1d. The important point is that the introduction of kurtosis reduces the curvature in $m(d)$, taking it closer to the null model, while for the variance, $v(d)$, we have $m^2\,(2)\,/v(2) \approx 1 + k^2/(4(1+k))$, such that eq. 10 holds.

For completeness, and to highlight the role of kurtosis, let us now assume a platykurtic distribution of effects, such that the effect on each trait is assumed to be uniformly distributed with mean zero: $x_i \sim U\,(-u/2, u/2)$. The key quantities can now be found by direct integration for $d = 1$ and $d = 2$.

$$
\begin{aligned}
-E\left(\ln w_d\right) &= \frac{(u/2)^k}{k+1}\sum_{i=1}^{n}\lambda_i, \qquad d=1 \\
&= \frac{u^k}{\binom{k+2}{2}}\sum_{i=1}^{n}\lambda_i, \qquad d=2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}\left(\ln w_d\right) &= (u/2)^{2k}\left((2k+1)^{-1}-(k+1)^{-2}\right)\sum_{i=1}^{n}\lambda_i^2, \qquad d=1 \\
&= u^{2k}\left(\binom{2k+2}{2}^{-1}-\binom{k+2}{2}^{-2}\right)\sum_{i=1}^{n}\lambda_i^2, \qquad d=2
\end{aligned} \tag{44}
$$

and so

$$
m(2) = \frac{2^{k+1}}{k+2}
$$

$$
v(2) = \frac{2^{2k}(k+5)}{(k+2)^2} \tag{45}
$$

Simulations of this model are shown in Figure S1e. The results show that reducing the kurtosis of the mutational effects acts to increase the effects of epistasis on the mean fitness (i.e., exaggerating the effects of k on $m(d)$), and also increases the variance, such that $v(2) > m^2(2)$.

The two previous sections confirm that introducing leptokurtic distributions of mutation effects alter predictions in a similar way to introducing modular pleiotropy, producing results that are intermediate between the simplest phenotypic model, and the null model with no epistasis. This is also illustrated in the green lines in Figure 1c-d. These results are relevant for interpretting the yeast data set [22], because, as shown in Figure 2b, the observed single-mutant effects are highly leptokurtic. However, our results also show that levels of kurtosis must be extreme for results to approach the predictions of the null model, as observed for the yeast data (Figure 2c-d). As such, high kurtosis cannot be the sole explanation of these results.

3.3. Biased mutations, and suboptimal wild-type

In this section, we allow for bias in the effects of mutations (i.e. a non-vanishing mean effect), and relax the assumption that the wild-type genotype, carrying no mutations, is phenotypically optimal. In both cases, this is easiest if we assume the isotropic fitness function of eq. 6.

For bias, we assume that the effects of the j^{th} mutation on the i^{th} trait is distributed as

$$
x_{ij} \sim N(\beta_i, 1) \tag{46}
$$

For suboptimality, we denote as z_i, the deviation from the optimum for the i^{th} trait in the wild-type. In this case, the sum of squared trait values follows a non-central Chi-squared distribution, whose noncentrality parameter is given by the sum of the squared deviations for each trait, namely $\alpha \equiv \sum_i^n \left(z_i + d\beta_i\right)^2$. The P^{th} moment of log fitness is the $(Pk/2)^{th}$ moment of this distribution, and so

$$E\left((-\ln W_d)^P\right) = (2d)^{Pk/2} e^{-\alpha/(2d)} \frac{\Gamma\left(\frac{Pk+n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} K\left(\frac{Pk+n}{2}, \frac{n}{2}, \frac{\alpha}{2d}\right) \tag{47}$$

$$= dn + \alpha, \qquad Pk/2 = 1$$

$$= 2d(dn + 2\alpha) + (dn + \alpha)^2, \qquad Pk/2 = 2$$

where

$$K(a, b, z) \equiv \sum_{i=0}^{\infty} \frac{\binom{a}{i}}{\binom{b}{i}} \frac{z^i}{i!}$$

is Kummer's confluent hypergeometric function [26]. Simple results now follow for $k = 2$, namely, $E(\ln w_d) = -(dn + \alpha + \ln W_0)$ and $\mathrm{Var}(\ln w_d) = 2d(dn + 2\alpha)$. For general k, well defined limits [26], allow us to derive results where maladaptation, or bias, are large.

First, let us consider results with biased mutation, but an optimal wildtype ($z_i = 0$). If we define $\beta \equiv \sum \beta_i^2$, then we find:

$$m_\beta(d) = d\frac{1 + d\beta}{1 + \beta}, \quad k = 2 \tag{48}$$

$$\to d^k, \quad \beta \to \infty \tag{49}$$

$$v_\beta(d) = d^2\frac{1 + 2d\beta}{1 + 2\beta}, \quad k = 2 \tag{50}$$

$$\to d^{2k-1} = \frac{m_\beta^2(d)}{d}, \quad \beta \to \infty \tag{51}$$

The decline of the mean fitness is now more rapid than in a model without bias (compare eqs. 8 and 48-49). This reflects the fact that epistasis becomes more negative when mutations have a tendency to modify traits in a consistent direction. Nevertheless, there is still a reduction in the variance compared the simplest model, such that eq. 10 still holds. These effects are illustrated by the blue lines in Figure 1a-b, which show the effects of bias when $k = 2$, and compared with simulations in Figure S1f.

Now let us consider the case where mutations are unbiased ($\beta_i = 0$), but where the wild-type is suboptimal. If we define $\xi = \sum z_i^2$, and note that $\ln W_0 = -\xi^{k/2}$, then we find

$$m_\xi(d) = d, \quad k = 2,$$

$$\to d, \quad \xi \to \infty \tag{52}$$

$$v_\xi(d) = d^2 \frac{1 + 2\xi/d}{1 + 2\xi}, \qquad k = 2$$

$$\to d = \frac{m_\xi^2(d)}{d}, \quad \xi \to \infty \tag{53}$$

These results show that the non-epistatic null model is approached as the wild-type becomes very maladapted [27]. This reflects the maladaptation reduces "overshoots" of the optimum, which cause sign epistasis.

Note also that eqs. 53 and 51, are equivalent to eq. 31, showing that extreme levels of maladaptation, modularity and bias have identical effects on the variance. As with the effects of high kurtosis, the results with maladaptation are relevant for interpretting the yeast data, where beneficial mutations were observed at a high rate (Figure 2b). Nevertheless, as with high kurtosis, levels of maladaption must be extreme to reproduce the predictions of the null model, and so, as we state in the main text, it seems that the phenotypic models, even in modified form, overpredict levels of epistasis in these data.

4. Simulation procedure

In Figures S1 and S2, analytical predictions are compared to simulations written in R. The simulations made various assumptions about the fitness function, and the distribution of mutant effects, and these are described in the text and Figure legends. For Figure S1, each increase in d was simulated by adding a 10^6 new mutations to the existing backgrounds. As such, each point in each Figure S1 represents the scaled mean or variance in fitness among 10^6 mutant individuals. For Figure S2, we generated 2×10^6 single mutations at random, and then combined these in pairs to calculate the 10^6 epistatic effects. As such, the larger points in Figure S2 represent the mean or variance in epistatic effects among 10^6 pairs of mutations, scaled by the mean or variance among the 2×10^6 single mutants. The smaller points in Figure S2a and c show estimated modal values. These were calculated using the half-range mode estimator of Bickel [28] with a bandwidth of 0.95, as implement in the R package *modeest* v. 2.1 [29]. When simulations used the fitness function of eq. 7, to generate the λ_i parameters, we followed [4,18], and used the eigenvalues of selection and mutation matrices, which were random Wishart matrices with n degrees of freedom.

Appendix 2: Details of data reanalysis

We searched the literature for data sets combining replicated measures of fitness for multiple mutations, chosen without regard for their fitness consequences. We rejected many excellent data sets where the trait measured was not a plausible proxy for fitness [30,31], or which contained no genotypes carrying four or more mutations [32,33], or mutations that were known in advance to be beneficial [34,35], or were otherwise biased [33], or which contained clear edge effects that could not be easily corrected [33,36]. Moreover, we did not consider mutation accumulation lines, where the number of mutations was not measured directly, so that estimates can be confounded by changes in mutation rate [37].

For the data set of Puchta et al. [22], a 333-nucleotide long U3 snoRNA gene in *Saccharomyces cerevisiae* was the target of a saturation mutagenesis experiment. The wild-type was a D343 strain, in which the U3 gene was

transformed to allow the yeast to survive on a selected environment containing glucose (otherwise U3 is down-regulated and growth arrested). Libraries of U3 mutated strains were constructed using "doped oligonucleotides" that randomly mutated any possible site between position 7 to 333 of the gene (327/333 sites, with an approximately 1% mutation rate per position). All possible point mutations of the U3 gene were represented in the libraries, which contained single-nucleotide polymorphisms (SNPs) and short insertions and deletions (indels). To measure fitness, competition experiments were performed in an environment containing glucose. Following Puchta et al. [22], our main text reports results from the "env. 1" condition, which was kept at 30°C.

Due to the mutagenesis procedure, many mutation combinations were present multiple times, and where this was the case, we took the mean of the log fitness estimates. Figure S3a compares the mean and standard deviation of the log fitness estimates for replicated strains. The plot shows a clear trend for heteroscedasticity, with larger fitness effects associated with greater measurement uncertainty (or higher environmental variance). Such heteroscedasticity should increase $v(d)$ above its true value, militating against a fit to the null model, and therefore making our conclusions conservative.

The data set of Puchta et al. [22] also includes additional replication, because fitness estimation was repeated in a second environment at 37°C ("env. 2"), and a third environment, also at 30°C ("env. 3"). As shown in Figure S4a-b, results for the two identical environments were highly correlated. Considering these replicate experiments, clarifies a disadvantage of using direct estimates of pairwise epistasis, eq. 5, because the estimates of this quantity, as shown in Figure S4c, are much less precisely replicated than the estimates of single- or double-mutant effects (Figure S4a-b). Furthermore, the estimated variance in epistatic effects, which was the subject of predictions by Martin et al. [1], is highly sensitive to the amount of replication. This is shown in Figure S4d. By contrast, as shown in Figure S5, the patterns evident in the moments of $\ln w_d$, are relatively robust between the three experiments (Figure S5a), and even more so, when multiple experiments are treated as replicates (Figure S5b). This remains true when we consider only Single Nucleotide Polymorphism mutations (i.e., excluding small insertions and deletions), of the kind that are used in the calculation of pairwise epistasis measures (Figure S5c). As is clear from Figure S5a,c, the experiment "env. 1", which we report in the main text, shows the largest deviations from expectations under the null model, again making our conclusions conservative.

A final consequence of the saturation mutagenesis procedure was that around half of the strains contained more than $d = 4$ mutations, and some contained as many as $d = 57$. We did not reanalyze these highly mutated strains, due to experimental difficulties in measuring very low fitness values. In particular, Puchta et al. [22] truncated their fitness measurements at $\ln w = -3$. This leads to edge effects that are clearly visible in Figure S3b (where log fitness values were averaged across all three replicate experiments). The edge effects are also visible in Figure S6, where we replicate Figure 1a-b, but retaining strains carrying up to $d = 12$ mutations (thereby including 93% of the data set). These edge effects explain our conservative choice to restrict the reanalysis to strains carrying $d \leq 4$ mutations in the main text.

Supplementary Figures

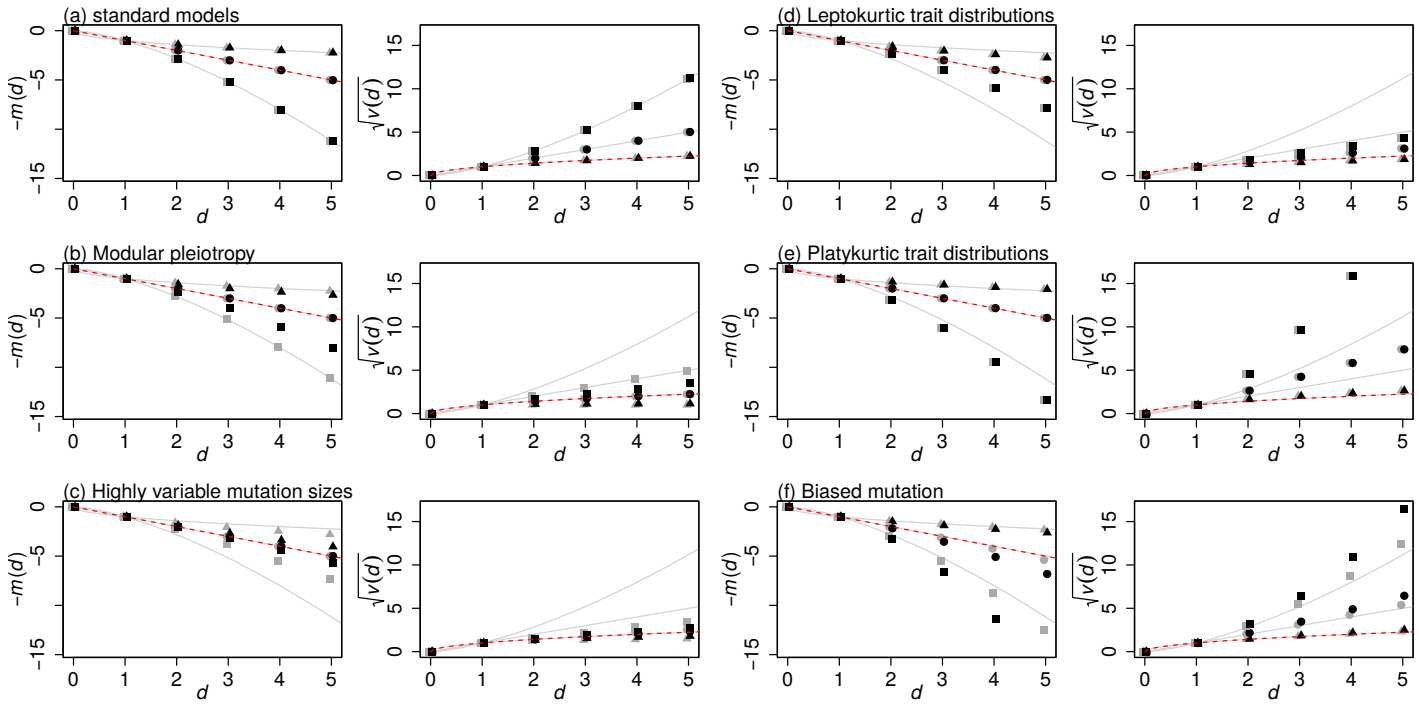

Figure S1: Properties of fitness epistasis between mutations under simple phenotypic models, based on Fisher's geometric model. The left-hand panel of each pair shows the mean log fitness of individuals carrying d mutations (eq. 1), and right-hand panel shows the equivalent standard deviation in log fitnesses for individuals carrying d mutations (eq. 2). For all plots, simulations are compared with $k = 1$ (triangles), $k = 2$ (circles) and $k = 3$ (squares). The lines show predictions for the simplest phenotypic model (eqs. 8-9), and the null model (eqs. 3-4 shown as dashed red lines). Each pair of panels shows results from two simulation conditions shown in either black or grey points. The conditions differ between panels as follows. In panel (a) results are compared for the simplest phenotypic models (eqs. 16-17) with the two different fitness function, each with $n = 5$ traits (black points: eq. 7; grey points: eq. 6). In panel (b), results use the fitness function of eq. 6, but with each mutation affecting either a distinct trait (black points: $n' = 1$), or a distinct set of 50 traits (grey points: $n' = 50$). In panel (c) the fitness function of eq. 6, was used with randomly orientated mutations on $n = 5$ traits; their magnitudes were drawn from either a Chi distribution with 0.1 degrees of freedom (black points), or an exponential distribution (grey points). In panel (d), the fitness function of eq. 7 was used, with the effects on each trait drawn from a reflected gamma distribution, with scale parameter 1, and shape parameter $(\sqrt{5} - 1)/2 \approx 0.61$ (i.e., a distribution with vanishing mean, unit variance, and a high kurtosis); results are compared with $n = 5$ traits (black points), and $n = 50$ traits (grey points). In panel (e), all details are as for panel (d), but the effects on each trait were drawn from a uniform distribution, on the range, $[-0.5, 0.5]$. In panel (f), the fitness function of eq. 6 was used with $n = 5$ traits, each with a non-zero mean effect; results are compared for biases of $\beta_i = 0.5$ (black points), and $\beta_i = 0.1$ (grey points). Other details of the simulations are given in the text.

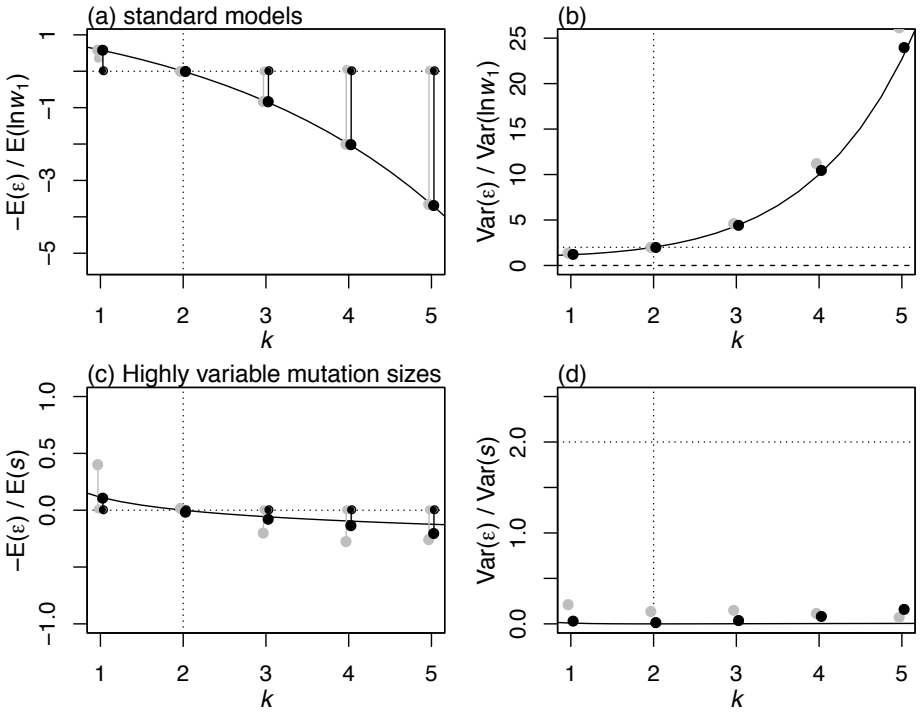

Figure S2: Simulations and analytical predictions for the distribution of pairwise epistatic fitness effects (eq. 5), under the additive phenotypic models. Each panel shows the scaled mean or variance in epistasis (eqs. 13-14), as a function of k, the curvature of the fitness landscape (eqs. 6-7), and compares predictions (curves) to simulations (points). In panels (a)-(b), mutation effect sizes were normal (eq. 17); curves show eqs. 23-25, and simulations and colours match Figure S1a. In panels (c)-(d), mutation sizes have a highly leptokurtic distribution; curves use eqs. 13, 14, 31 and 32; and simulations and colours match those used in Figure S1c. In panels (a) and (c), larger dots show means, and smaller dots show modal values.

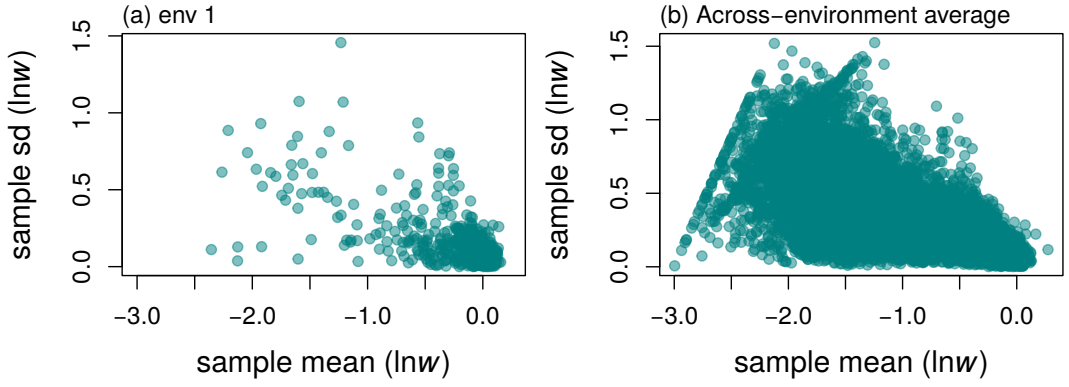

Figure S3: The correlation between the mean and standard deviation of replicate measures of mutant fitness for the dataset of Puchta et al. [22]. Results are for all individuals carrying up to $d = 12$ mutations. Panel (a) shows fitness measurements in environment 1, and includes only mutations that were replicated due to multiple hits during the random mutagensis. Panel (b) shows results for all mutations, by treating the 3 environments as replicated measures. The visible lines show the edge effects caused by inability to measure very small fitness values.

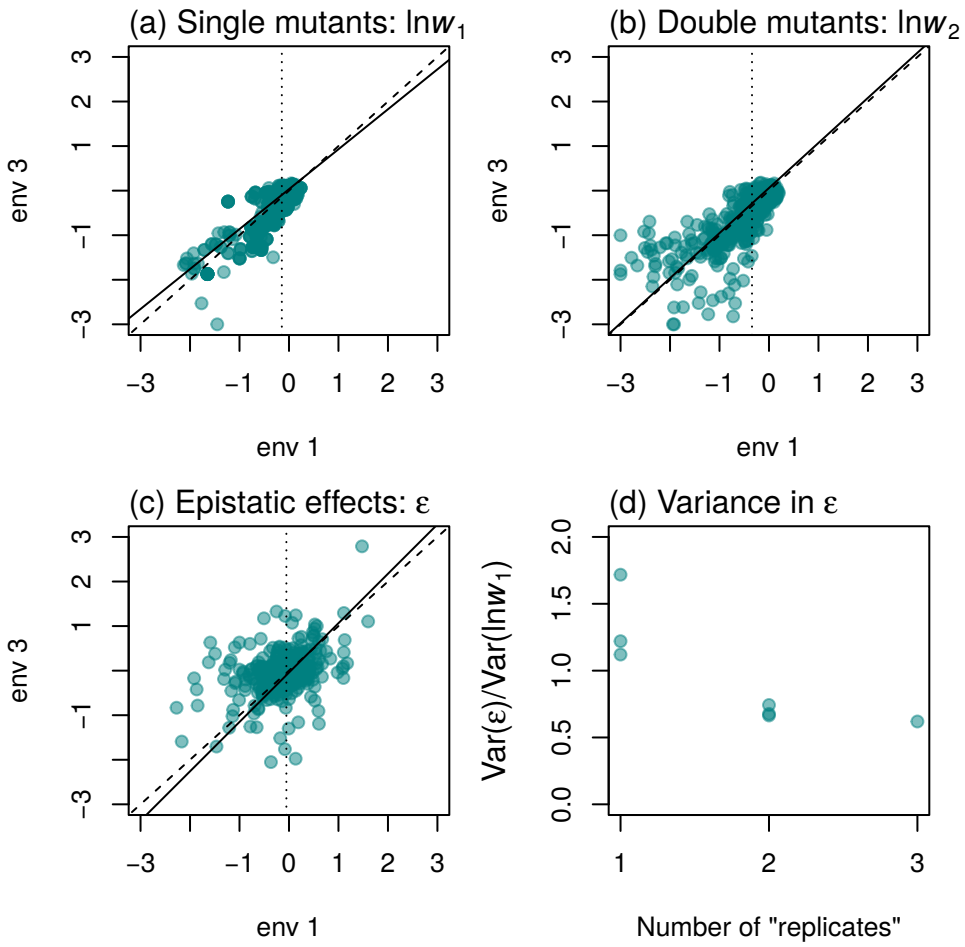

Figure S4: *Saccharomyces cerevisiae* snoRNA mutants generated by Puchta et al. [22]. Fitness measurements are shown for the same mutant strains, assayed in two environments, env 1 and env 3 (both containing glucose at 30°C). Results are shown only for Single Nucleotide Polymorphism mutations that were present as both single and double mutants (i.e., discarding all insertions and deletions, and mutations appearing only a singletons). Panel (a) shows the single mutants; panel (b) the double mutants, and panel (c) shows the corresponding epistatic effects (eq. 5). In each case, the best-fit Standardized Major Axis regression (solid line) is compared to the 1:1 slope (dashed line). Panel (d) shows the scaled variance in epistatic effects (eq. 14), when the log fitness values were either measured in a single environment, or averaged over 2 or 3 environments. Increasing the level of replication decreases the inferred variance in epistatic effects.

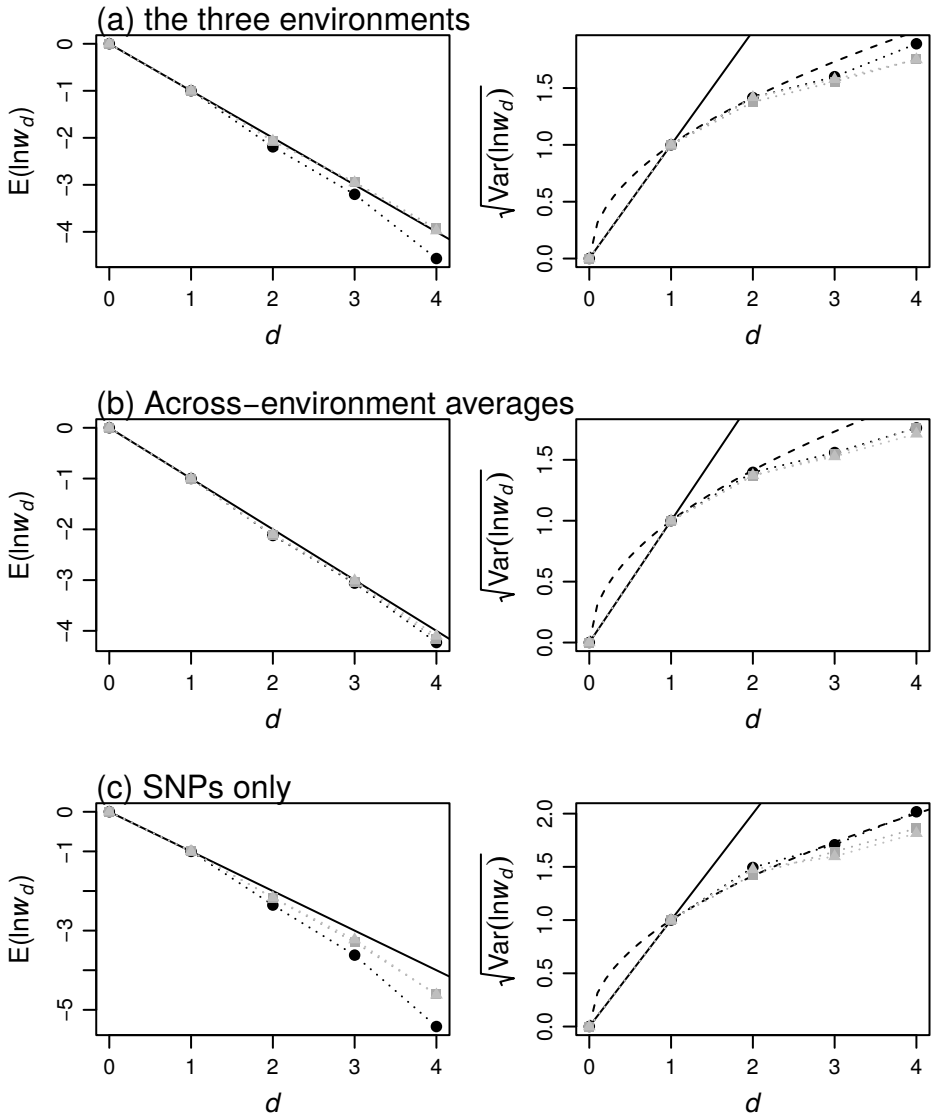

Figure S5: *Saccharomyces cerevisiae* snoRNA mutants generated by Puchta et al. [22], and assayed in competition experiments in three environments (env. 1 and 3 in glucose at 30°C, and env. 2 in glucose at 37°C). All plots show the mean and standard deviation in the log fitnesses of individuals carrying d mutations, as in Figure 2. Panel (a) shows results for the three environments separately (env. 1: black circles, env. 2: dark grey squares, and env. 3: lighter grey triangles). Panel (b) shows results when log fitness measurements were averaged across environments: (env. 1 and 3: black points, env. 1 and 2: dark grey squares, and all three environments: lighter grey triangles). Panel (c) is identical to panel (a), but shows only Single Nucleotide Polymorphism mutations (i.e., discarding small insertions and deletions).

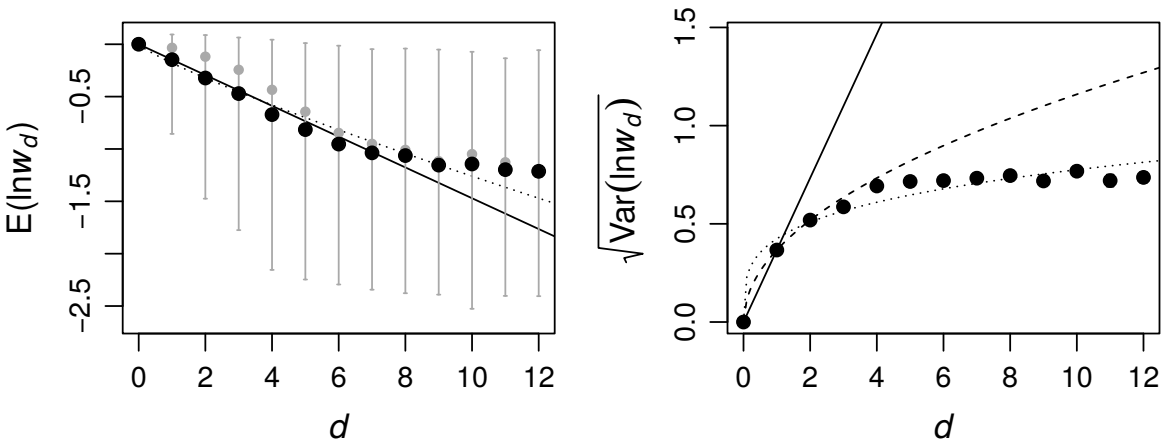

Figure S6: *Saccharomyces cerevisiae* snoRNA mutants generated by Puchta et al. [22]. Plots are identical to Figure 2c-d, but show results for individuals carrying up to $d = 12$ mutations. Edge effects, caused by the inability to measure fitness accurately below a certain value, have a visible effect after the first few mutations. This explains why our main results were truncated at $d = 4$.

Supplementary References

1. Martin G, Elena SF, Lenormand T. 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. *Nat. Genet.* **39**, 555–60. (doi: 10.1038/ng1998)

2. Gros PA, Nagard HL, Tenaillon O. 2009 The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. *Genetics* **182**, 277–293. (10.1534/genetics.108.099127)

3. Tenaillon O, Silander OK, Uzan JP, Chao L. 2007 Quantifying Organismal Complexity using a Population Genetic Approach. *PLoS One* **2**, e217. (doi: 10.1371/journal.pone.0000217)

4. Martin G, Lenormand T. 2006 The fitness effect of mutations across environments: a survey in light of fitness landscape models. *Evolution* **60**, 2413–2427. (doi: 10.1111/j.0014-3820.2006.tb01878.x)

5. Welch JJ, Waxman D. 2003 Modularity and the cost of complexity. *Evolution* **57**, 1723–1734. (doi: 10.1111/j.0014-3820.2003.tb00581.x)

6. Orr HA. 2000 Adaptation and the cost of complexity. Evolution 54, 13-20. (doi: 10.1111/j.0014-3820.2000.tb00002.x)

7. Wingreen NS, Miller J, Cox EC. 2003 Scaling of mutational effects in models for pleiotropy. Genetics 164, 1221-1228.

8. Rockman MV. 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**, 1-17. (doi: 10.1111/j.1558-5646.2011.01486.x)

9. Waxman D, Welch JJ. 2005 Fisher's Microscope and Haldane's Ellipse. *Am. Nat.* **166**, 447–457. (doi: 10.1086/444404)

10. Estes S, Ajie BC, Lynch M, Phillips PC. 2005 Spontaneous mutational correlations for life-history, morphological and behavioral characters in *Caenorhabditis elegans*. *Genetics* **170**, 645–653. (doi: 10.1534/genetics.104.040022)

11. Welch JJ, Waxman D. 2002 Non-equivalent loci and the distribution of mutant effects. *Genetics* **161**, 897-904.

12. Waxman D, Peck JR. 2003 The anomalous effects of biased mutation. *Genetics* **164**, 1615–1626.

13. Damerval C, Maurice A, Josse JM, de Vienne D. 1994 Quantitative trait loci underlying gene product variation: a novel perspective by analyzing regulation of genome expression. *Genetics* **137**, 289-301.

14. Kroymann J, Mitchell-Olds T. 2005 Epistasis and balanced polymorphism influencing complex trait variation. *Nature* **435**, 95–98. (doi: 10.1038/nature03480)

15. Wagner GP, Zhang J. 2011 The pleiotropic structure of the genotype-phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12**, 204. (doi: 10.1038/nrg2949)

16. Hill WG, Zhang XS. 2012 On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. *Genetics* **190**, 1131-1137. (doi:10.1534/genetics.111.135681)

17. Martin G. 2014 Fisher's geometrical model emerges as a property of complex integrated phenotypic networks. Genetics 197, 237–255. (doi: 10.1534/genetics.113.160325)

18. Martin G, Lenormand T. 2006 A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**, 893–907. (doi: 10.1111/j.0014-3820.2006.tb01169.x)

19. Orr HA. 1998 The Population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. Evolution 52, 935–949. (doi: 10.1111/j.1558-5646.1998.tb01823.x)

20. Lourenço J, Galtier N, Glémin S. 2011 Complexity, pleiotropy, and the fitness effect of mutations. *Evolution* **65**, 1559–1571. (doi: 10.1111/j.1558-5646.2011.01237.x)

21. Chevin LM, Martin G, Lenormand T. 2010 Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**, 3213-31. (doi: 10.1111/j.1558-5646.2010.01058.x)

22. Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016 Network of epistatic interactions within a yeast snoRNA. *Science* **352**, 840-844. (doi: 10.1126/science.aaf0965)

23. Fisher R. 1930 The Genetical Theory of Natural Selection. Clarendon Press, Oxford, U.K.

24. Leigh EG. 1987 Ronald Fisher and the development of evolutionary theory. II. Influences of new variation on evolutionary processes. Pp. 212–263 in P. H. Harvey, and L. Partridge, eds. Oxford surveys in evolutionary biology. Vol. 4. Oxford Univ. Press, Oxford, U.K.

25. Poon A, Otto SP. 2000 Compensating for our load of mutations: Freezing the mutational meltdown. *Evolution* **54**, 1467–1479. (doi: 10.1111/j.0014-3820.2000.tb00693.x)

26. Abramowitz M, Stegun IA. 1964. Handbook of Mathematical Functions. Dover, New York.

27. Blanquart F, Achaz G, Bataillon T, Tenaillon O. 2014 Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* **68**, 3537-3554. (doi: 10.1111/evo.12545)

28. Bickel DR. 2002 Robust estimators of the mode and skewness of continuous data. *Comput. Stat. Data Anal.* **39**, 153-163. (doi: 10.1016/S0167-9473(01)00057-3)

29. Poncet P. 2012 modeest: Mode Estimation. R package version 2.1. http://CRAN.R-project.org/package=modeest.

30. Sarkisyan KS, Bolotin DA, Meer MV, Usmanova DR, Mishin AS, Sharonov GV, Ivankov DN, Bozhanova NG, Baranov MS, Soylemez O, Bogatyreva NS, Vlasov PK, Egorov ES, Logacheva MD, Kondrashov AS, Chudakov DM, Putintseva EV, Mamedov IZ, Tawfik DS, Lukyanov KA, Kondrashov FA. 2016 Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397. (doi: 10.1038/nature17995)

31. Olson CA, Wu NC, Sun R. 2014 A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643-2651. (doi: 10.1016/j.cub.2014.09.072)

32. Kemble HE, Eisenhauer C, Couce A, Chapron A, Magnan M, Gautier G, Le Nagard H, Nghe P, Tenaillon O. 2018 Flux, toxicity and protein expression costs shape genetic interaction in a metabolic pathway. *bioRxiv* 362327. (doi:10.1101/362327)

33. Kuzmin E, VanderSluis B, Wang W, Tan G, Deshpande R, Chen Y, Usaj M, Balint A, Mattiazzi Usaj M, van Leeuwen J, Koch EN, Pons C, Dagilis AJ, Pryszlak M, Wang JZY, Hanchard J1, Riggi M, Xu K, Heydari H, San Luis BJ, Shuteriqi E, Zhu H, Van Dyk N, Sharifpoor S, Costanzo M, Loewith R, Caudy A, Bolnick D, Brown GW, Andrews BJ, Boone C, Myers CL. 2018 Systematic analysis of complex genetic interactions. *Science* **360**, eaao1729. (doi:10.1126/science.aao1729)

34. Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ. 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190-1192. (doi: 10.1126/science.1203799)

35. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF. 2011. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193-1196. (doi: 10.1126/science.1203801)

36. Baryshnikova A, Costanzo M, Kim Y, Ding H, Koh J, Toufighi K, Youn JY, Ou J, San Luis BJ, Bandyopadhyay S, Hibbs M, Hess D, Gingras AC, Bader GD, Troyanskaya OG, Brown GW, Andrews B, Boone C, Myers CL. 2010 Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7**, 1017 (doi: 10.1038/nmeth.1534)

37. Sharp NP, Agrawal AF. 2012 Evidence for elevated mutation rates in low-quality genotypes. *Proc. Natl. Acad. Sci. USA* **109**, 6142-6146. (doi: 10.1073/pnas.1118918109)