The distribution of epistasis on simple fitness landscapes

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

March 8, 2019

1. Université de Montpellier, Sète, France; 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

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 about the overall distribution 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 it 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 focussed 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 [22], a simple model of optimizing selection acting on quantitative traits [10,12,20,21]. Though a toy model, this approach is closely related to a broad class of systems biology models, involving metabolic networks [21]. Furthermore, it naturally generates fitness epistasis, even when mutations are additive on the phenotype; 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 fitting to data [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 curvature of the fitness landscape, and neither did they explore other possible variants of Fisher's geometric 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, the scaled mean and standard deviation of this quantity are

$$m(d) \equiv \frac{E\left(\ln w_d\right)}{E\left(\ln w_1\right)} \tag{1}$$

$$\sqrt{v(d)} \equiv \frac{\mathrm{sd}\left(\ln w_d\right)}{\mathrm{sd}\left(\ln w_1\right)} \tag{2}$$

where, by definition, $m(0) = v(0) = 0$ and $m(1) = v(1) = 1$. These equations use a log scale, because deviations from multiplicativity (i.e. 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 are 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

$$m_0(d) = d \tag{3}$$

$$v_0(d) = m_0(d) \tag{4}$$

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:

$$\varepsilon \equiv \ln w_{(ab)} - \ln w_{(a)} - \ln w_{(b)} \tag{5}$$

Here, $w_{(a)}$ denotes the relative fitness of the genome carrying the mutation "a", and so on. Though widely used, ε can be difficult to work with. For example, if the same mutation appears in multiple double mutants, then the complete distribution of ε will entail using the samefitness 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, in the main text, we will focus on the simpler quantities of eqs. 1-2, and give some equivalent results for ε 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 $\mathrm{Var}(\varepsilon) = 0$ (see Appendix 1). Together, then, eqs. 3-4 imply that there is no epistasis at all.

Additive phenotypic models

We now examine results under Fisher's geometric model. Here, an individual's fitness depends on its phenotype, described as an n-dimensional vector, $\mathbf{z} = \{z_1, z_2, ..., z_n\}$, whose components, z_i, are the value of each trait. Fitness depends on the deviation of the phenotype from a single optimal value. A suitable fitness function of this kind uses the Euclidean distance of the phenotype from the origin, raised to the k^{th} power.

$$\ln W(\mathbf{z}) \propto -\|\mathbf{z}\|^k \tag{6}$$

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

$$\ln W(\mathbf{z}) \propto -\sum_{i=1}^{n} \lambda_i |z_i|^k \tag{7}$$

where λ_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 λ_i are equal.

The simplest versions of the model make three further assumptions: (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 this case, the phenotype of an individual carrying d mutations can be written as

$$\mathrm{z} = \left\{ \sum_{j=1}^{d} x_{j1}, \sum_{j=1}^{d} x_{j2}, \ldots, \sum_{j=1}^{d} x_{jn} \right\} \tag{8}$$

where

$$x_{ji} \sim N(0,1). \tag{9}$$

In Appendix 1, we show that, for both fitness functions, these assumptions yield the following results, as illustrated by the black lines in Figure 1:

$$m(d) = d^{k/2} \tag{10}$$

$$v(d) = m^2(d) \tag{11}$$

Eqs. 10-11 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. 11 will never agree with eq. 4, because these simple phenotypic models always generate fitness epistasis.

Extensions to the phenotypic model

Confronted with data from real quantitative traits [42], many aspects of the model above appear grossly 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 [39,43-45].

Furthermore, there is clear evidence that both selection and mutation are correlated among traits [46,47], 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 [48,49]. Furthermore, there is some evidence of appreciable epistasis at the level of phenotype [50,51]; and restricted or modular pleiotropy, where mutations affect only a subset of traits ([39,52]; though see [53]). Finally, there is often evidence of beneficial mutations, which implies that the wild-types are suboptimal. None of this is consistent with eqs. 8-9.

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" [21,29,46]. Nonetheless, other assumptions are certainly restrictive. In Appendix 1, we explore several extensions of the model, building on the results of several previous studies [29,32,38,39,41,44,46], 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. Despite their heterogeneity, most of these extensions act to reduce mean levels of epistasis. With modular pleiotropy, this is because mutations affecting different traits will interact less; with high kurtosis, it is because epistasis is reduced when any of the mutations is very small in magnitude; finally, parental maladaptation reduces "overshoots" of the optimum, which cause sign epistasis [27]. In all cases, the predicted $m(d)$ is intermediate between predictions from the simplest phenotypic models (eq. 10) and the null model (eq. 3). This is illustrated by the green lines in Figure 1c, which show results with a leptokurtic distribution of effects on each trait. Only one of the modifications has a qualitatively different effect. When mutations are biased, their tendency to modify traits in a consistent direction makes epistasis more negative. To illustrate this, let us assume that mutational effects have a non-zero mean, β_i, such that, $x_{ij} \sim N(\beta_i, 1)$. When the bias is large, we find that

$$m_\beta(d) \approx d^k, \qquad \beta \gg 1 \tag{12}$$

where $\beta \equiv \sum \beta_i^2$ (see Appendix 1 for details). The decline of the mean fitness is now more rapid than in a model without bias (compare eqs. 10 and 12), and this is illustrated by the blue lines in Figure 1a, which show the effects of bias when $k = 2$.

For the variance in log fitness, the effects are even more consistent. For all of the extensions, we find a reduction, compared to simplest phenotypic model, such that

$$v(d) < m^2(d), \qquad d > 1 \tag{13}$$

and when $k \geq 2$, results for the null model act as lower bound, such that $v(d) \geq m(d)$. This is illustrated by the green and blue lines in Figure 1b and d.

To summarize, modifying the phenotypic model, to reflect data from real quantitative traits, has two main effects. First, it erases information about the true curvature of the fitness landscape, so that the form of $m(d)$ cannot easily be used to estimate k. Second, it reduces the variance in log fitness, below $m^2(d)$.

Reanalysis of data from a yeast snoRNA

To illustrate the approach above, we now reanalyse a published data set, examining its fit to the predictions above, and comparing different measures of epistasis. In particular, we examine data from Puchta et al. [1], who used saturation mutagenesis of the U3 snoRNA in *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, relative to simple phenotypic model, could have been predicted from other aspects of the data. For example, the distribution of single-mutant fitnesses (Figure 2b), shows that the distribution is highly leptokurtic, and indicates the presence of beneficial mutations (346/965 mutations increase growth rate). Nevertheless, kurtosis and wild-type maladaptation both need to be extreme for predictions to converge to the null model (see Appendix 1). Furthermore, the hypothesis of modularity, whereby mutations each affect different sets of traits, seems inherently implausible for these data, where all mutations affect sites in the same snoRNA. As such, we conclude that the phenotypic models - even in modified form - overestimate the true amount of fitness epistasis in these data. This implies that the simplest population genetic models, which ignore epistasis altogether, 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 samemodel 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, or restricting pleiotropy), 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 [26,54,55]. As such, the good fit of $k = 2$ implies that global modifiers of fitness epistasis do not arise in these systems.

Of course, there is no reason to assume that identical patterns of epistasis will characterise all data sets [56,57], and we have offered two further reasons to doubt this. First, empirically, we have shown that the data of [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; but this effect should be evident in the reduced levels of variance (such that $v(d) < m^2(d)$ for $d > 1$). 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

Not applicable.

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 Sklodowska-Curie Co-Funding European program).

Acknowledgements

We are very grateful to Grzegorz Kudla, 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

1. 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)

2. Kimura M, Maruyama T. 1966 The mutational load with epistatic gene interactions in fitness. *Genetics* **54**, 1337–1351.

3. Lewontin RC. 1974 The genetic basis of evolutionary change. Columbia University Press, London.

4. Kondrashov AS. 1988 Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440. (doi: 10.1038/336435a0)

5. Kondrashov AS. 1995 Contamination of the genome by very slightly deleterious mutations: why have we not died 100 times over? *J. Theor. Biol.* **175**, 583-594. (doi: 10.1006/jtbi.1995.0167)

6. Otto SP, Feldman MW. 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* **51**, 134-47. (doi: 10.1006/tpbi.1997.1301)

7. Phillips PC, Otto SP, Whitlock MC. 2000 Beyond the average: the evolutionary importance of gene interactions and variability of epistatic effects. Ch. 2 In: Wolf, J. B., E. D. Brodie III, and M. J. Wade (Eds.). Epistasis and the Evolutionary Process, Oxford University Press, New York (2000), pp. 20-38.

8. Kouyos RD, Silander OK, Bonhoeffer S. 2007 Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol. Evol.* **22**, 308–315. (doi: 10.1016/j.tree.2007.02.014)

9. Peck JR, Waxman D, Welch JJ. 2012 Hidden epistatic interactions can favour the evolution of sex and recombination. *PLoS One* **7**, e48382. (doi: 10.1371/journal.pone.0048382)

10. Charlesworth B. 2013 Why we are not dead one hundred times over. *Evolution* **67**, 3354-3361. (doi: 10.1111/evo.12195)

11. Fraïsse C, Gunnarsson PA, Roze D, Bierne N, Welch JJ. 2016 The genetics of speciation: Insights from Fisher's Geometric Model. *Evolution* **70**, 1450-1464. (doi: 10.1111/evo.12968)

12. Barton NH. 2017 How does epistasis influence the response to selection? *Heredity* **118**, 96-109. (doi: 10.1038/hdy.2016.109)

13. Simon A, Bierne N, Welch JJ. 2018 Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. *Evolution Lett.* **2**, 472-498. (doi:10.1002/evl3.66)

14. Mukai T. 1969 The genetic structure of natural populations of *Drosophila melanogaster*. VII. Synergistic interaction of spontaneous mutant polygenes controlling viability. *Genetics* **61**, 749–761.

15. Elena SF, Lenski RE. 1997 Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398. (doi: 10.1038/37108)

16. West SA, Peters AD, Barton NH. 1998 Testing for epistasis between deleterious mutations. *Genetics* **149**, 435–444.

17. Halligan DL, Keightley PD. 2009 Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Ann. Rev. Ecol. Syst.* **40**, 151–172. (doi.org/10.1146/annurev.ecolsys.39.110707.173437)

18. Charlesworth B. 1990 Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genet. Res.* **55**, 199–221. (doi: 10.1017/s0016672300025532)

19. Wilke CO, Adami C. 2001 Interaction between directional epistasis and average mutational effects. *Proc. Roy. Soc. B. Biol. Sci.* **268**, 1469-1474. (doi: 10.1098/rspb.2001.1690)

20. Lande R. 1980 The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* **94**, 203–215.

21. 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)

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

23. Peck JR, Barreau G, Heath S. 1997 Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* **145**, 1171–99.

24. Martin G, Lenormand T. 2006b 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)

25. 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)

26. 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. (doi: 10.1534/genetics.108.099127)

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. Roze D, Blanckaert A. 2014 Epistasis, pleiotropy, and the mutation load in sexual and asexual populations. *Evolution* **68**, 137–149. (doi: 10.1111/evo.12232)

29. Martin G, Lenormand T. 2006a 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)

30. Manna F, Martin G, Lenormand T. 2011 Fitness landscapes: an alternative theory for the dominance of mutation. *Genetics* **189**, 923–937. (doi: 10.1534/genetics.111.132944)

31. Sousa A, Magalhaes S, Gordo I. 2011 Cost of Antibiotic Resistance and the Geometry of Adaptation. *Mol. Biol. Evol.* **29**, 1417-1428. (doi: 10.1093/molbev/msr302)

32. 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)

33. Maclean RC, Perron GG, Gardner A. 2010 Diminishing Returns From Beneficial Mutations 1011 and Pervasive Epistasis Shape the Fitness Landscape for Rifampicin Resistance in 1012 Pseudomonas aeruginosa. *Genetics* **186**, 1345-1354. (doi: 10.1534/genetics.110.123083)

34. Perfeito L, Sousa A, Bataillon T, Gordo I. 2013 Rates of fitness decline and rebound suggest pervasive epistasis. *Evolution* **68**, 150-162. (doi:10.1111/evo.12234)

35. Weinreich DM, Knies JL. 2013 Fisher's geometric model of adaptation meets the functional synthesis: data on pairwise epistasis for fitness yields insights into the shape and size of phenotype space. *Evolution* **67**, 2957–2972. (doi:10.1111/evo.12156)

36. Blanquart F, Bataillon T. 2016 Epistasis and the structure of fitness landscapes: are experimental fitness landscapes compatible with Fisher's model? *Genetics* **203**, 847–862. (doi: 10.1534/genetics.115.182691)

37. Sanjuàn R, Moya A, Elena SF. 2004 The contribution of epistasis to the architecture of fitness in an RNA virus. *Proc. Natl. Acad. Sci. USA* **101**, 15376–15379. (doi: 10.1073/pnas.0404125101)

38. 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)

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

40. Chevin LM, Martin G, Lenormand T. 2010 Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogenous mutation, and parallel evolution. *Evolution* **64**, 3213-3231. (doi: 10.1111/j.1558-5646.2010.01058.x)

41. 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)

42. Lynch M, Walsh B. 1998 Genetics and Analysis of Quantitative Traits. Sinauer Associates, Sunderland MA.

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

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

45. 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)

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

47. 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)

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

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

50. 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.

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

52. 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)

53. 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)

54. Liberman U, Feldman MW. 2006 Evolutionary theory for modifiers of epistasis using a general symmetric model. *Proc. Natl. Acad. Sci. USA* **103**, 19402–19406. (doi: 10.1073/pnas.0608569103)

55. Desai MM, Weissman D, Feldman MW. 2007 Evolution can favor antagonistic epistasis. *Genetics* **177**, 1001-1010. (doi:10.1534/genetics.107.075812)

56. Sanjuàn R, Elena SF. 2006 Epistasis correlates to genomic complexity. *Proc. Natl. Acad. Sci. USA* **103**, 14402–14405. (doi: 10.1073/pnas.0604543103)

57. Belshaw R, Gardner A, Rambaut A, Pybus OG. 2008 Pacing a small cage: mutation and RNA viruses. *Trends Ecol. Evol.* **23**, 188-93. (doi:10.1016/j.tree.2007.11.010)

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, 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. 51-52); 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., eq. 12). Green lines show results where the mutations on each trait are drawn from a leptokurtic reflected exponential distribution (eqs. 44).

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 $k = 2$: $\varepsilon \sim N(0, 2\mathrm{Var}(\ln w_1))$ (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 10); and the best-fit regression for $\ln m(d) \sim \ln d$ (dotted line, which has a slope implying $\hat{k} = 2.16$). (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. 11 (solid line); and the best-fit regression of $\ln v(d) \sim \ln d$ (dotted line, which has slope 0.89).

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

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

1. Université de Montpellier, Sète, France; 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, and justify claims about the extensions to the simplest phenotypic model. We will also present results 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

$$\ln w_1 \equiv \ln W(\mathbf{z}+\mathbf{x}) - \ln W(\mathbf{z}) \tag{14}$$

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. 14 is denoted as s by Martin et al. [1]. From eq. 5, the pairwise epistatic effect for two mutations, a and b, is then

$$\varepsilon = \ln W(\mathbf{z}+\mathbf{x}_a+\mathbf{x}_b) - \ln W(\mathbf{z}+\mathbf{x}_a) - \ln W(\mathbf{z}+\mathbf{x}_b) + \ln W(\mathbf{z}) \tag{15}$$

[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:

$$\frac{E(\varepsilon)}{E(\ln w_1)} = m(2) - 2 \tag{16}$$

$$\frac{\mathrm{Var}(\varepsilon)}{\mathrm{Var}(\ln w_1)} = v(2) + 2 - 4\sqrt{v(2)}r_{12} \tag{17}$$

Here, we have defined

$$r_{12} \equiv \mathrm{Cor}\left(\ln W(\mathbf{x}_a+\mathbf{x}_b+\mathbf{z}), \ln W(\mathbf{x}_a+\mathbf{z})\right) \tag{18}$$

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 (eq. 9).

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:

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

$$\mathrm{Var}\left(\ln w_d\right) = (2d)^{k}\left(\frac{\Gamma\left(\frac{2k+n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} - \frac{\Gamma^2\left(\frac{k+n}{2}\right)}{\Gamma^2\left(\frac{n}{2}\right)}\right) \tag{20}$$

The results of eqs. 10-11 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).

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

$$\mathrm{Var}\left(\ln w_d\right) = (2d)^{k}\left(\frac{\Gamma\left(\frac{2k+1}{2}\right)}{\sqrt{\pi}} - \frac{\Gamma^2\left(\frac{k+1}{2}\right)}{\pi}\right)\sum_{i=1}^{n}\lambda_i^2 \tag{22}$$

and again, eqs. 10-11 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. 17), we also require the correlation coefficient of eq. 18. For the fitness function of eq. 7, this is maximized at $k = 2$, where it takes the value:

$$r_{12} = \mathrm{Cor}\left(\left|x_{ai} + x_{bi}\right|^2, \left|x_{ai}\right|^2\right) = \frac{1}{2}, \quad k = 2 \tag{23}$$

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

$$-\frac{E\left(\varepsilon\right)}{E\left(\ln w_1\right)} = 2 - 2^{k/2} \tag{24}$$

$$= 0, \quad k = 2 \tag{25}$$

$$\frac{\mathrm{Var}\left(\varepsilon\right)}{\mathrm{Var}\left(\ln w_1\right)} \approx 2(1 + 2^{k-1} - 2^{k/2}) \tag{26}$$

$$= 2, \quad k = 2 \tag{27}$$

These results are compared to simulation in Fig. S2. When $k = 2$, eqs. 25 and 27 reproduce the results of Martin et al. [1], while increasing k above this value makes expected levels of epistasis more negative ($E\left(\varepsilon\right) < 0$), and increases the variance in epistatic effects ($\mathrm{Var}\left(\varepsilon\right) > 2\mathrm{Var}\left(\ln w_1\right)$).

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

$$\varepsilon \propto \sum_{i}^{n} \lambda_i \xi_i, \qquad k = 2 \tag{28}$$

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 Fig. 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

In this section, we consider various extensions to the simplest phenotypic model. These analyses support eqs. 12 and 13 and statements in the main text.

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 ([5]; see also [6,7]), 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:

$$\|\mathbf{x}_a + \mathbf{x}_b\| = \sqrt{\sum_{i}^{n} (x_{ai} + x_{bi})^2} = \sqrt{\|\mathbf{x}_a\|^2 + \|\mathbf{x}_b\|^2 + 2\|\mathbf{x}_a\| \|\mathbf{x}_b\| \cos(\theta)} \tag{29}$$

where θ is the angle in radians between the two mutational vectors, in the n-dimensional trait space [5]. 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:

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

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

$$\mathrm{Var}\left(\ln w_d\right) = 2^k \left(\frac{\Gamma\left(\frac{2k+dn'}{2}\right)}{\Gamma\left(\frac{dn'}{2}\right)} - \frac{\Gamma^2\left(\frac{k+dn'}{2}\right)}{\Gamma^2\left(\frac{dn'}{2}\right)} \right) \tag{31}$$

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

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

$$\frac{v(d)}{m^2(d)} = \frac{\Gamma(dn'/2+k)\,\Gamma(dn'/2)\,\Gamma^{-2}(dn'/2+k/2)-1}{\Gamma(n'/2+k)\,\Gamma(n'/2)\,\Gamma^{-2}(n'/2+k/2)-1} \approx \frac{1}{d} \tag{32}$$

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

$$m(d) = \frac{B\left(\frac{n'}{2}, \frac{k}{2}\right)}{B\left(\frac{dn'}{2}, \frac{k}{2}\right)} \tag{33}$$

$$\to d^{k/2}, \quad n' \to \infty \tag{34}$$

$$\to d, \quad n' \to 0 \tag{35}$$

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. 10), and the null model (eq. 3), and that eq. 13 will hold. Simulations with intermediate values of n' are shown in Figure S1b, and confirm these claims.

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

$$\varepsilon = 2\,\|\mathbf{x}_a\|\,\|\mathbf{x}_b\|\cos(\theta), \qquad k = 2 \tag{36}$$

As shown by Fisher [8], when the number of traits, n, is not very small, then an unbiased distribution of mutation directions leads to $2\cos(\theta) \sim N(0, 4/n)$ [9,10]. As such, we have

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

$$\mathrm{Var}(\varepsilon) \approx \frac{4}{n} E\left(\left(\|\mathbf{x}_a\|\,\|\mathbf{x}_b\|\right)^2\right), \qquad k = 2 \tag{37}$$

If we follow Lourenço et al. [7] and draw the squared mutation magnitudes from a Chi-squared distribution with n' degrees of freedom, then it follows that $\mathrm{Var}(\varepsilon) = \frac{4}{n}n'^2$. 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 [11]. 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

404 effect on a single trait, $|x|$, is exponentially distributed, such that

$$E[|x|^k] = k!\mu^k \tag{38}$$

$$\mathrm{Var}\left(|x|^k\right) = \mu^{2k}\left((2k)! - (k!)^2\right) \tag{39}$$

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

$$E\left(|x_a + x_b|^k \,|x_a x_b > 0\right) = \mu^k \frac{\Gamma(2+k)}{\Gamma(2)} = \mu^k (k+1)! \tag{40}$$

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

$$f(\delta) = \frac{2}{\mu^2} e^{-|\delta|/\mu} \tag{41}$$

$$E\left(|x_a + x_b|^k \,|x_a x_b < 0\right) = k!\mu^k \tag{42}$$

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

$$\begin{aligned} E[|x_a + x_b|^k] &= \mu^k \left(\frac{(k+1)! + k!}{2}\right) = \mu^k k! \left(\frac{k+2}{2}\right) \\ \mathrm{Var}[|x_a + x_b|^k] &= \mu^{2k}\left[(2k)!\,(k+1) - \left(k!\left(\frac{k+2}{2}\right)\right)^2\right] \end{aligned} \tag{43}$$

409 and so, we find:

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

$$v(2) = \frac{(2k)!\,(k+1) - (k!)^2\left(\frac{k+2}{2}\right)^2}{(2k)! - (k!)^2} \approx 1 + k \tag{44}$$

410 where the approximate expression for $v(2)$ uses Stirling's approximation: $k! \approx \sqrt{2n\pi}\left(\frac{n}{e}\right)^n$, such that $(2k)!/(k!)^2 \approx$
411 $2^{2k}/\sqrt{\pi k}$. The results are supported by simulations shown in Figure S1d. The important point is that the intro-
412 duction of kurtosis reduces the curvature in $m(d)$, taking it closer to the null model, while for the variance, $v(d)$,
413 we have $m^2(2)/v(2) \approx 1 + k^2/(4(1+k))$, such that eq. 13 holds.
414 For completeness, and to highlight the role of kurtosis, let us now assume a platykurtic distribution of effects,
415 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
416 key quantities can now be found by direct integration for $d = 1$ and $d = 2$.

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

$$
\begin{aligned}
\mathrm{Var}\left(\ln w_d\right) &= (u/2)^{2k}\left((2k+1)^{-1}-(k+1)^{-2}\right)\sum_{i=1}^{n}\lambda_i^2, \qquad d=1 \\
&= u^{2k}\left(\binom{2k+2}{2}^{-1}-\binom{k+2}{2}^{-2}\right)\sum_{i=1}^{n}\lambda_i^2, \qquad d=2
\end{aligned} \tag{45}
$$

and so

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

$$
v(2) = \frac{2^{2k}(k+5)}{(k+2)^2} \tag{46}
$$

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 $v(2) > m^2(2)$.

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

$$
x_{ij} \sim N(\beta_i, 1) \tag{47}
$$

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

$$
\begin{aligned}
E\left((-\ln W_d)^P\right) &= (2d)^{Pk/2}e^{-\alpha/(2d)}\frac{\Gamma\left(\frac{Pk+n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}K\left(\frac{Pk+n}{2},\frac{n}{2},\frac{\alpha}{2d}\right) \\
&= dn+\alpha, \qquad Pk/2=1 \\
&= 2d(dn+2\alpha)+(dn+\alpha)^2, \qquad Pk/2=2
\end{aligned} \tag{48}
$$

where

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

is Kummer's confluent hypergeometric function [12]. Simple results now follow for $k = 2$, namely, $E\left(\ln w_d\right) = -(dn + \alpha + \ln W_0)$ and $\mathrm{Var}\left(\ln w_d\right) = 2d(dn + 2\alpha)$. For general k, well defined limits [12], allow us to derive results where maladaptation, or bias, are large.

First, let us consider the case where mutations are unbiased ($\beta_i = 0$), but 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

$$\begin{aligned} m_\xi(d) &= d, \quad k = 2, \\ &\to d, \quad \xi \to \infty \end{aligned} \tag{49}$$

$$\begin{aligned} v_\xi(d) &= d^2 \frac{1 + 2\xi/d}{1 + 2\xi}, \qquad k = 2 \\ &\to d = \frac{m_\xi^2(d)}{d}, \quad \xi \to \infty \end{aligned} \tag{50}$$

These results show that the non-epistatic null model is approached as the wild-type becomes very maladapted [13].

Results with bias, but an optimal wildtype ($z_i = 0$), follow in the same way. If we define $\beta \equiv \sum \beta_i^2$, then we find:

$$\begin{aligned} m_\beta(d) &= d \frac{1 + d\beta}{1 + \beta}, \quad k = 2 \\ &\to d^k, \quad \beta \to \infty \end{aligned} \tag{51}$$

$$\begin{aligned} v_\beta(d) &= d^2 \frac{1 + 2d\beta}{1 + 2\beta}, \quad k = 2 \\ &\to d^{2k-1} = \frac{m_\beta^2(d)}{d}, \quad \beta \to \infty \end{aligned} \tag{52}$$

Note that eqs. 50 and 52, are equivalent to eq. 32, showing that extreme levels of maladaptation, modularity and bias have identical effects on the variance. Simulation results with mutational bias are shown in Figure S1f.

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 [24] with a bandwidth of 0.95, as implement in the R package *modeest* v. 2.1 [25]. When simulations used the fitness function of eq. 7, to generate the λ_i parameters, we followed [4,23], 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 [14,15], or which contained no genotypes carrying four or more mutations [16,17], or mutations that were known in advance to be beneficial [18,19], or were otherwise biased [17], or which contained clear edge effects that could not be easily corrected [17,20]. 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 [21].

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 Figure Legends

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. 10-11), 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. 8-9) 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. 16-17), 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. 9); curves show eqs. 24-26, and simulations and colours match Figure S1a. In panels (c)-(d), mutation sizes have a highly leptokurtic distribution; curves use eqs. 16, 17, 32 and 33; 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. 17), 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. 2006b 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. 2002 Non-equivalent loci and the distribution of mutant effects. *Genetics* **161**, 897-904.

6. 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)

7. 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)

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

9. 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.

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

11. 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)

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

13. 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)

14. 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)

15. 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)

16. 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)

17. 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)

18. 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)

19. 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)

20. 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)

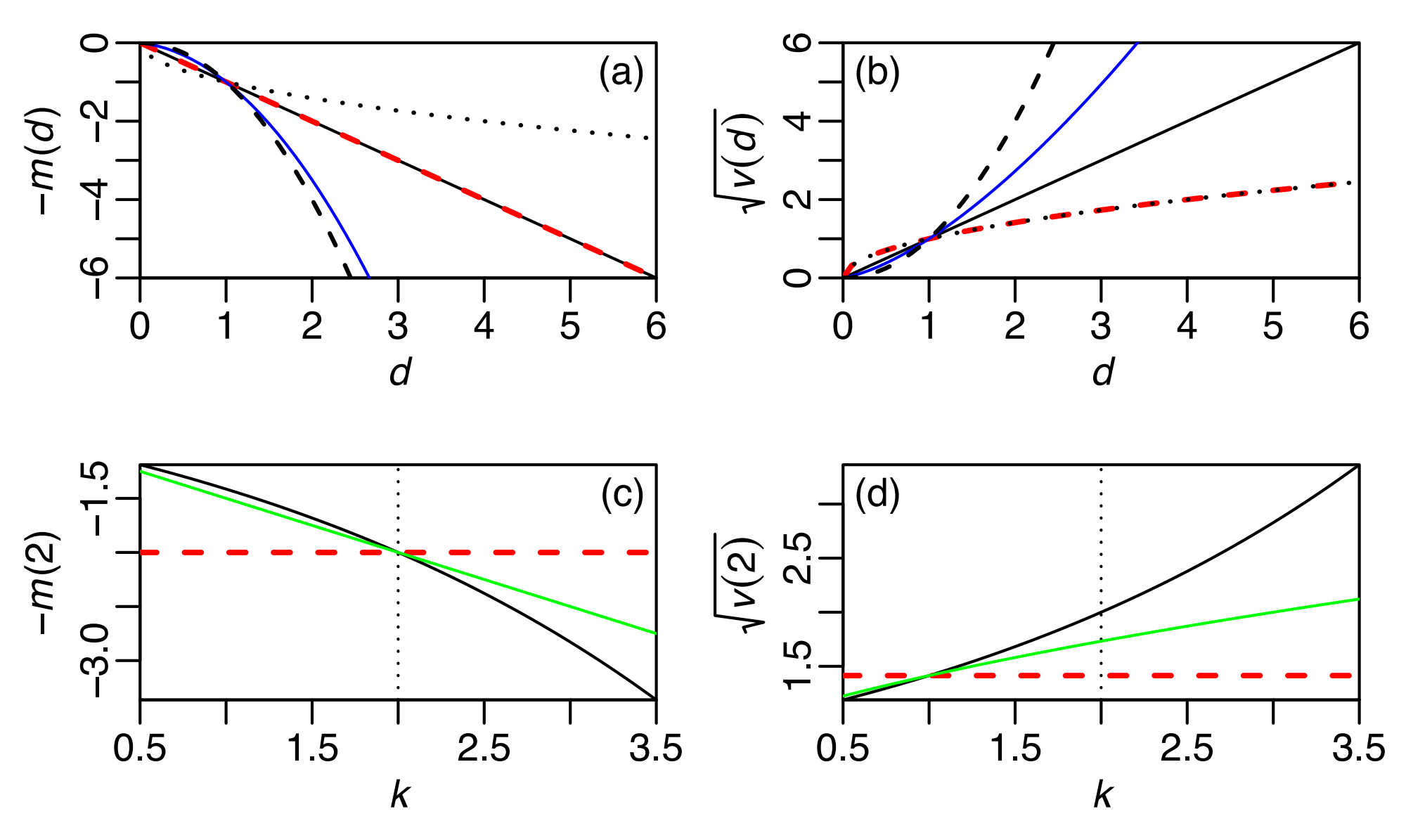
21. Sharp NP, Agrawal AF. 2012 Evidence for elevated mutation rates in low-quality genotypes. *Proc. Natl. Acad. Sci. USA* **109**, 6142-6146. (doi: 10.1073/pnas.1118918109)

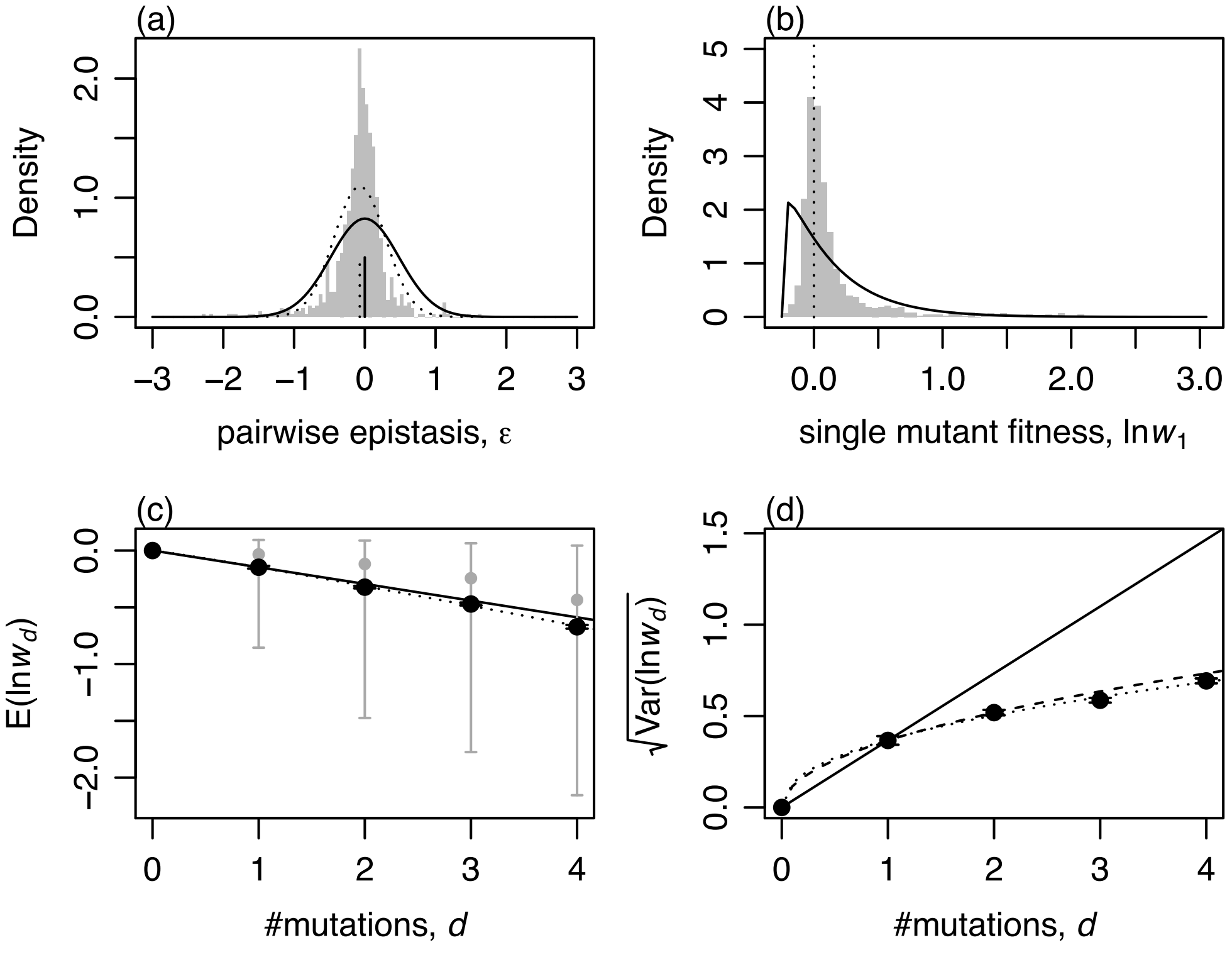
22. Puchta O, Cseke B, Czaja H, Tollervey D, Sanguinetti G, Kudla G. 2016 Network of epistatic interactions within a yeast snoRNA. *Science* **352**, 840-844. (doi: 10.1126/science.aaf0965)

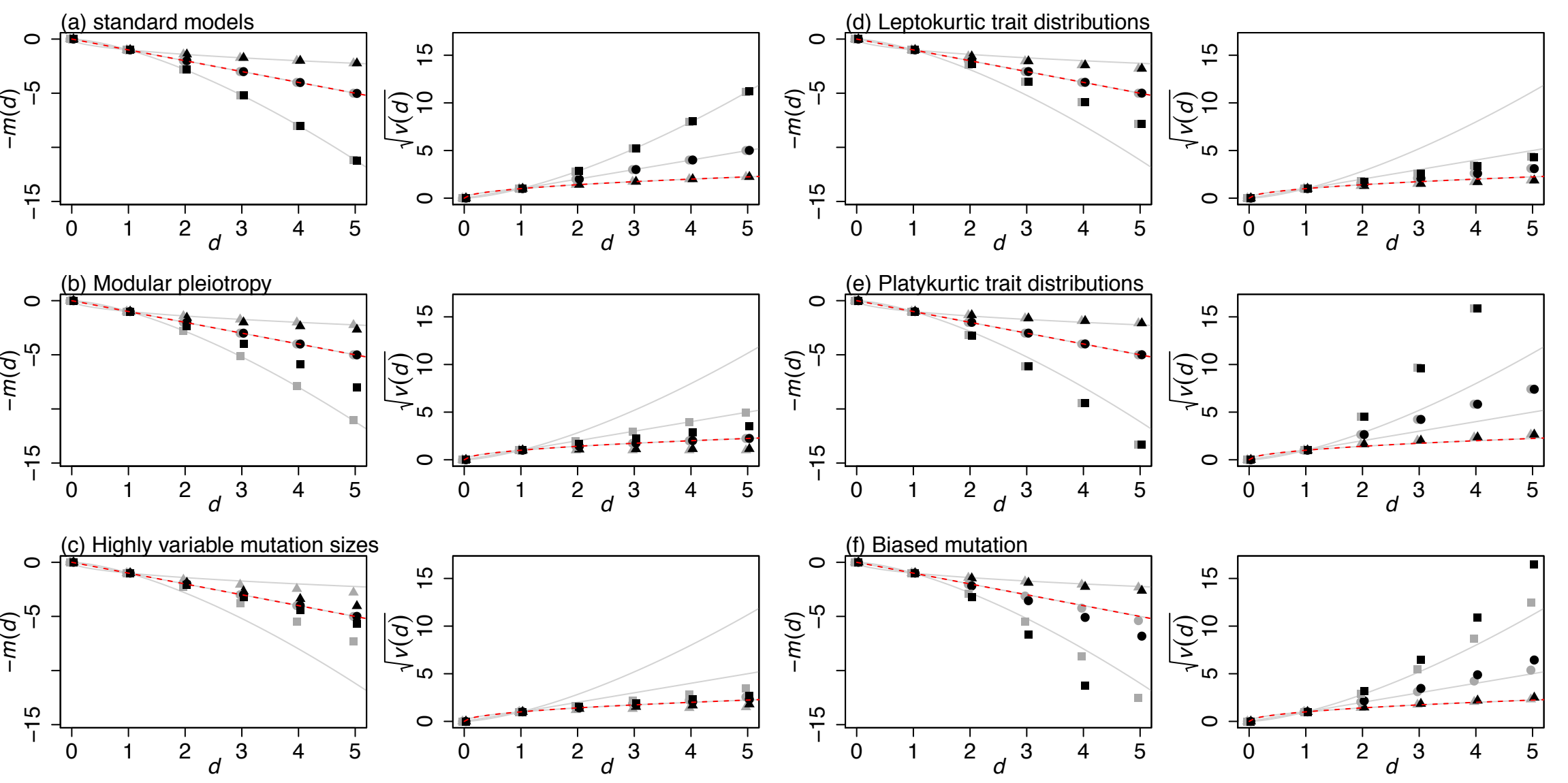
23. Martin G, Lenormand T. 2006a A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**, 893–907. (doi: 10.1111/j.0014-3820.2006.tb01169.x)

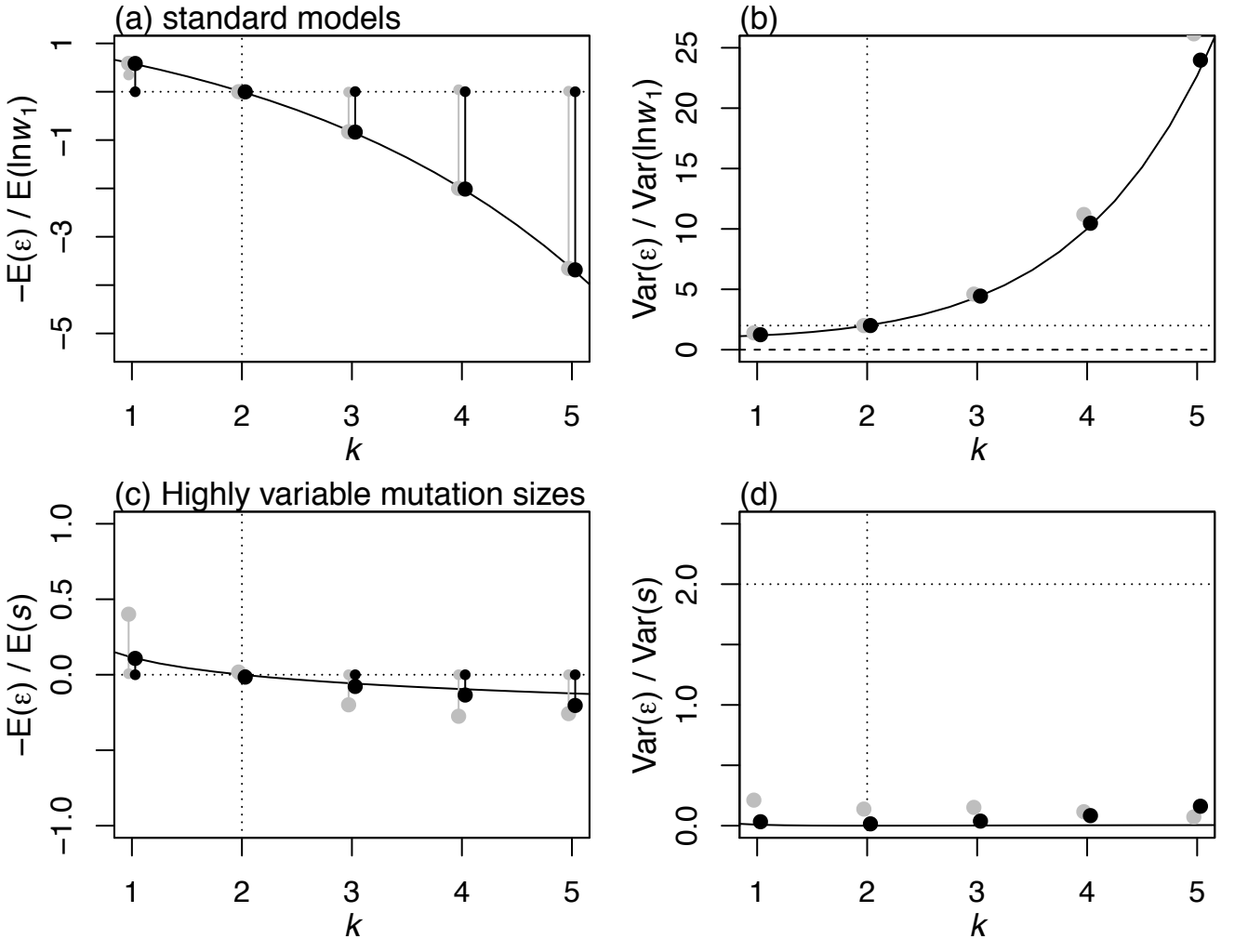
24. Bickel DR. 2002 Robust estimators of the mode and skewness of continuous data. *Comput. Stat. Data Anal.* **39**, 153-163. (doi: 10.1016/S0167-9473(01)00057-3)

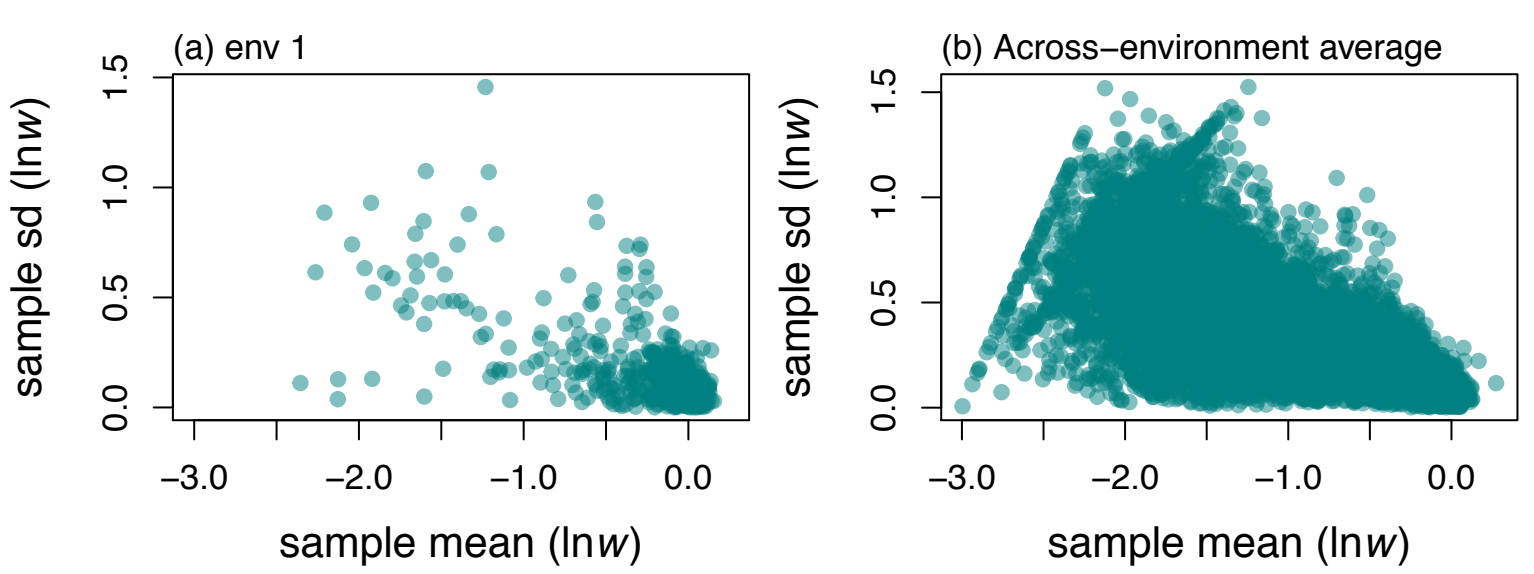
25. Poncet P. 2012 modeest: Mode Estimation. R package version 2.1. http://CRAN.R-project.org/package=modeest.

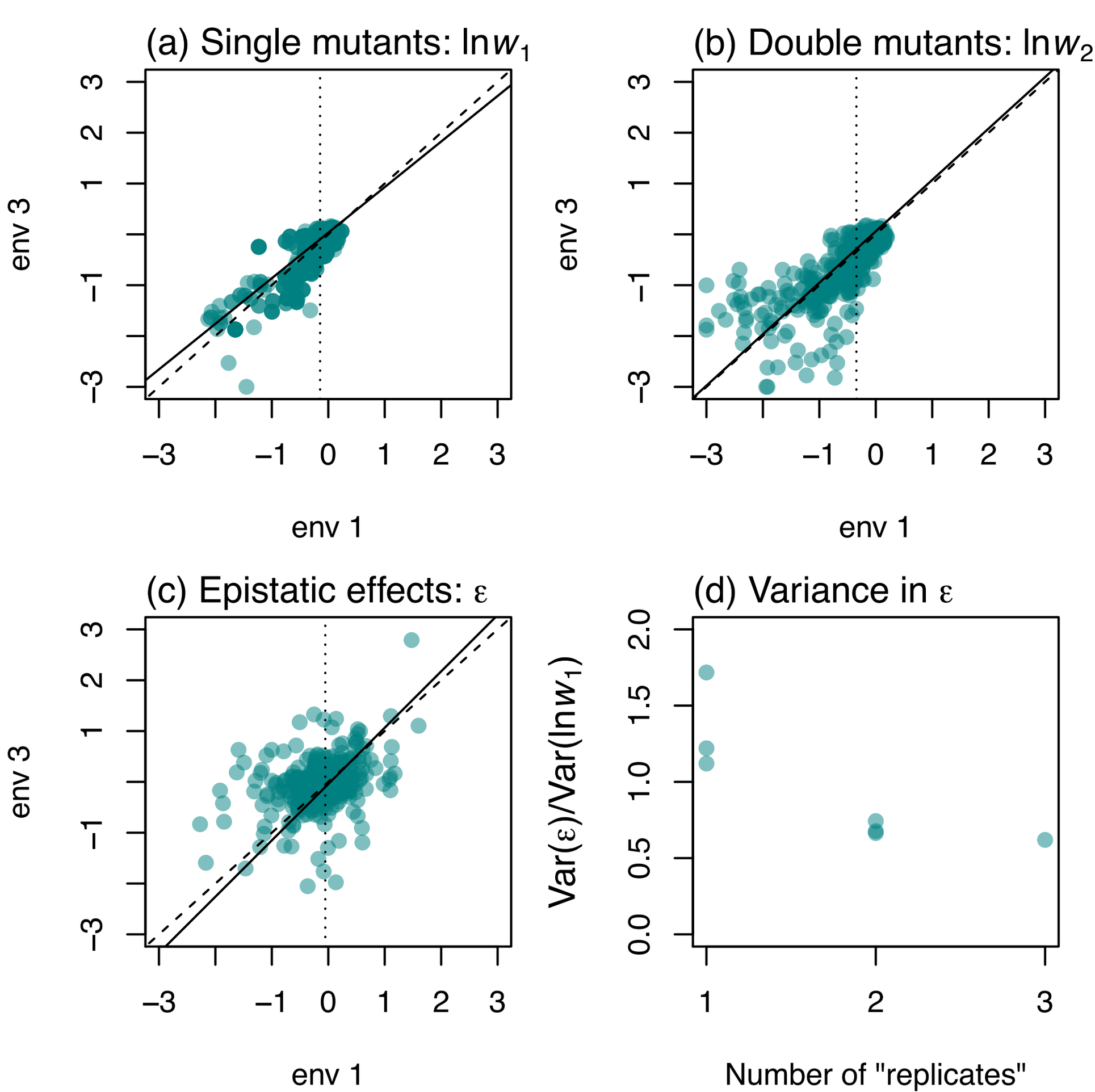

(a)
−m(d)
0
−2
−4
−6
0
1
2
3
4
5
6
d
(b)
√v(d)
0
2
4
6
0
1
2
3
4
5
6
d
(c)
−m(2)
−1.5
−3.0
0.5
1.5
2.5
3.5
k
(d)
√v(2)
1.5
2.5
0.5
1.5
2.5
3.5
k

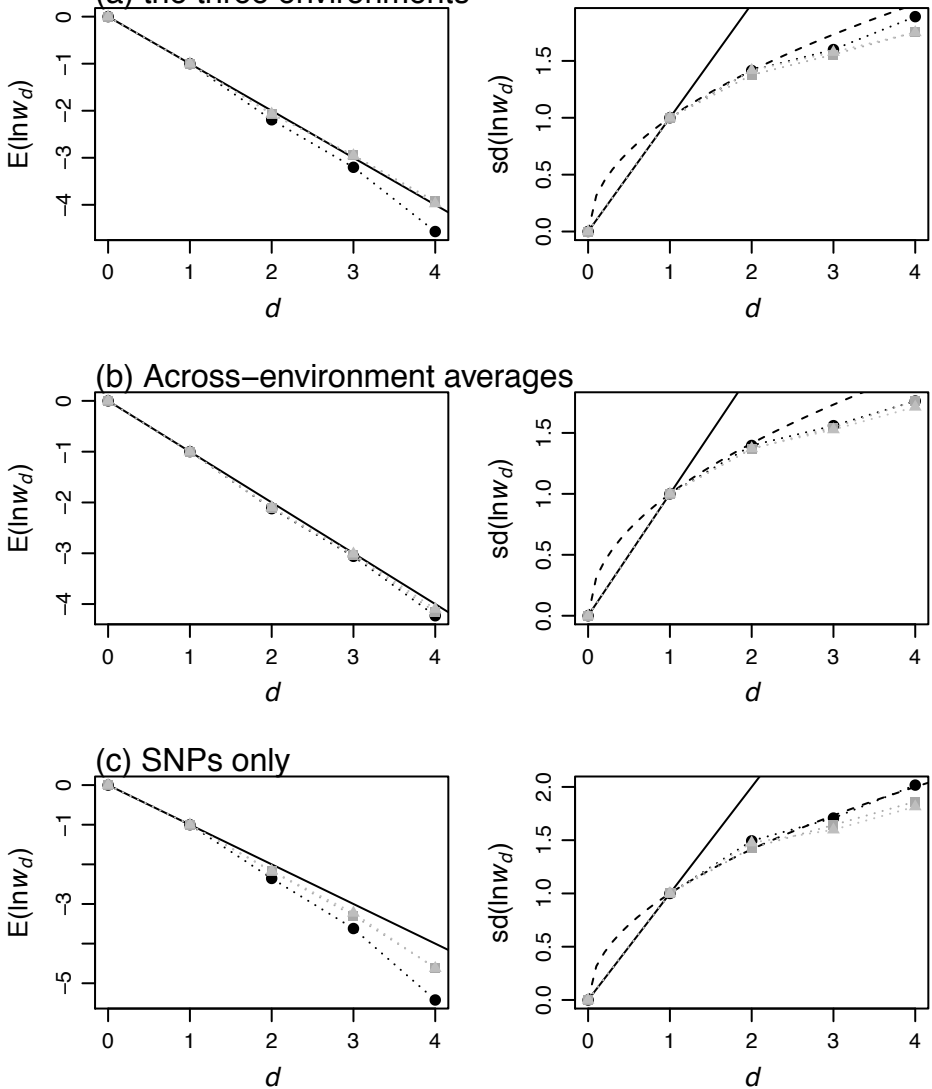

(a)
Density
0.0
1.0
2.0
−3
−2
−1
0
1
2
3
pairwise epistasis, ε
(b)
Density
0
1
2
3
4
5
0.0
1.0
2.0
3.0
single mutant fitness, lnw_1
(c)
E(lnw_d)
0.0
−1.0
−2.0
0
1
2
3
4
#mutations, d
(d)
$\sqrt{\text{Var}(\ln w_d)}$
0.0
0.5
1.0
1.5
0
1
2
3
4
#mutations, d

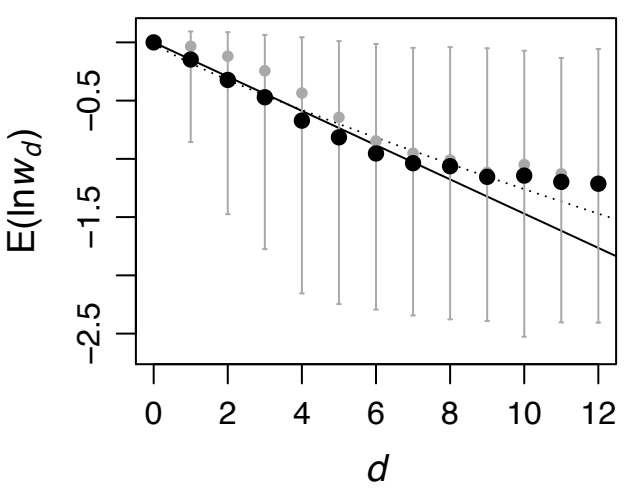

(a) standard models
(b) Modular pleiotropy
(c) Highly variable mutation sizes
(d) Leptokurtic trait distributions
(e) Platykurtic trait distributions
(f) Biased mutation
−m(d)
√v(d)
d
0
−5
−15
5
10
15
1
2
3
4

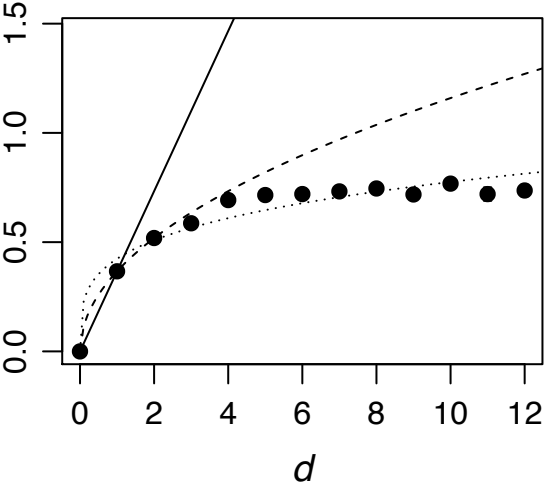

(a) standard models
−E(ε) / E(lnw1)
1
−1
−3
−5
(b)
Var(ε) / Var(lnw1)
0
5
10
15
20
25
(c) Highly variable mutation sizes
−E(ε) / E(s)
1.0
0.5
0.0
−1.0
(d)
Var(ε) / Var(s)
0.0
1.0
2.0
1
2
3
4
5
k


(a) env 1
sample sd (ln$w$)
0.0
0.5
1.0
1.5
−3.0
−2.0
−1.0
0.0
sample mean (ln$w$)
(b) Across−environment average
sample sd (ln$w$)
0.0
0.5
1.0
1.5
−3.0
−2.0
−1.0
0.0
sample mean (ln$w$)


(a) Single mutants: lnw_1
(b) Double mutants: lnw_2
(c) Epistatic effects: ε
(d) Variance in ε
env 3
env 1
Var(ε)/Var(lnw_1)
Number of "replicates"
−3
−1
0
1
2
3
0.0
0.5
1.0
1.5
2.0


(a) the three environments
E(lnw_d)
sd(lnw_d)
d
(b) Across−environment averages
E(lnw_d)
sd(lnw_d)
d
(c) SNPs only
E(lnw_d)
sd(lnw_d)
d


E(ln w_d)
−0.5
−1.5
−2.5
0
2
4
6
8
10
12
d


sd(ln w_d)
1.5
1.0
0.5
0.0
0
2
4
6
8
10
12
d