Biased Estimates of Diminishing-Returns Epistasis? Empirical Evidence Revisited

David Berger*,[1] and Erik Postma†

*Animal Ecology at Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, SE-75236 Uppsala, Sweden, and †Institute of Evolutionary Biology and Environmental Studies, University of Zurich, CH-8057 Zürich, Switzerland

ABSTRACT Empirical evidence for diminishing fitness returns of beneficial mutations supports Fisher's geometric model. We show that a similar pattern emerges through the phenomenon of regression to the mean and that few studies correct for it. Although biases are often small, regression to the mean has overemphasized diminishing returns and will hamper cross-study comparisons unless corrected for.

The Problem

EPISTASIS, *i.e.*, the interactive (nonadditive) effect of coexpressed mutations, is widespread (*e.g.*, Weinreich *et al.* 2006; Flint and Mackay 2009; Huang *et al.* 2012; Corbett-Detig *et al.* 2013) and plays a fundamental role in genetic theories of sex and recombination, mutation load, genetic robustness, response to selection, and speciation (Whitlock *et al.* 1995; Phillips *et al.* 2000; De Visser *et al.* 2011; Olson-Manning *et al.* 2012; Hansen 2013).

A number of recent studies have attempted to demonstrate that as mutations become increasingly beneficial, they are more likely to show negative epistasis for fitness when combined (supporting information, Table S1). The demonstration of such diminishing fitness returns has bearing on evolutionary theory by providing a mechanistic basis for decelerating rates of adaptation, as predicted from Fisher's geometric model (FGM) (Martin *et al.* 2007; Chou *et al.* 2011; Khan *et al.* 2011; Draghi and Plotkin 2013; Szendro *et al.* 2013).

Diminishing-returns epistasis has commonly been inferred from a negative correlation between the additive fitness effects of a pair (or set) of mutations and their epistatic effect in interaction with each other (Table S1). To this end, the relative fitness of mutation i (w_i) is typically measured as

$$w_i = \log\left(\frac{w_{\text{abs},i}}{w_{\text{abs,ref}}}\right),$$

where $w_{\text{abs},i}$ is the absolute fitness of a genotype carrying mutation i and $w_{\text{abs,ref}}$ is the fitness of the wild type used as a reference. The relative fitness of a second mutation j (w_j) and that of mutations i and j combined (w_{ij}) are obtained in the same manner. Subsequently, the epistatic interaction between mutations i and j (E_{ij}) is obtained as

$$E_{ij} = \; w_{ij} - [w_i + w_j] \tag{1}$$

or, in words, by the difference between the *observed* fitness of a genotype carrying both mutations i and j (w_{ij}) and the *expected* fitness of this double mutant if gene action is completely additive ($[w_i + w_j]$) (*e.g.*, da Silva *et al.* 2010). Repeating this for a large number of combinations of mutations, the correlation between E_{ij} and $[w_i + w_j]$ equals

$$r_{E_{ij},[w_i+w_j]} = \frac{\sigma(E_{ij},\; [w_i + w_j])}{\sqrt{\sigma^2(E_{ij})\sigma^2(w_i + w_j)}}. \tag{2}$$

In the presence of diminishing-returns epistasis, mutations with large beneficial effects on fitness show more negative epistasis, resulting in this correlation being negative.

However, by calculating epistasis from expected fitness (Equation 1), the two terms to be correlated will share measurement errors, and a statistical dependence is created artificially. In any empirical study w_i, w_j, and w_{ij} are measured with error. So if $w_i = a_i + e_i$ and $w_j = a_j + e_j$, where

doi: 10.1534/genetics.114.169870
Manuscript received August 18, 2014; accepted for publication October 11, 2014; published Early Online October 13, 2014.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.169870/-/DC1.
[1]Corresponding author: Animal Ecology at Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University, SE-75236 Uppsala, Sweden. E-mail: david.berger@ebc.uu.se

a and e are the additive genetic and residual components of w_i and w_j, respectively, and $w_{ij} = a_i + a_j + i_{ij} + e_{ij}$, where i_{ij} is the epistatic effect of mutations i and j, then

$$E_{ij} = i_{ij} + e_{ij} - \left[e_i + e_j\right]. \tag{3}$$

Assuming uncorrelated measurement errors,

$$r_{E_{ij},[w_i+w_j]} = \frac{\sigma\left(i_{ij},\left[a_i + a_j\right]\right) - \left[\sigma^2(e_i) + \sigma^2(e_j)\right]}{\sqrt{\left[\sigma^2(i_{ij}) + \sigma^2(e_{ij}) + \sigma^2(e_i) + \sigma^2(e_j)\right]\left[\sigma^2(a_i) + \sigma^2(a_j) + \sigma^2(e_i) + \sigma^2(e_j)\right]}}. \tag{4}$$

It follows that measurement error in w_i, w_j, and w_{ij} [*i.e.*, $\sigma^2(e_i) = \sigma^2(e_j) = \sigma^2(e_{ij}) > 0$], appearing in the denominator of Equation 4, weakens the correlation. However, $[\sigma^2(e_i) + \sigma^2(e_j)]$ also appears in the numerator, making negative correlations more negative and positive correlations less positive. The latter is the result of correlating $[w_i + w_j]$ with $w_{ij} - [w_i + w_j]$ and thereby the measurement error in $[w_i + w_j]$ (*i.e.*, $e_i + e_j$) with itself. On the whole, measurement error can thus result in a negative correlation between expected fitness and epistasis, which could erroneously be interpreted as evidence for diminishing-returns epistasis (Figure 1A).

Having knowledge of $\sigma^2(e_i)$, $\sigma^2(e_j)$, and $\sigma^2(e_{ij})$, we are able to obtain the corrected correlation between i_{ij} and $[a_i + a_j]$ that is not biased by measurement error variance, using

$$r_{i_{ij},[a_i+a_j]} = \frac{\sigma\left(E_{ij},\left[w_i + w_j\right]\right) + \sigma^2(e_i) + \sigma^2(e_j)}{\sqrt{\left[\sigma^2(E_{ij}) - \sigma^2(e_{ij}) - \sigma^2(e_i) - \sigma^2(e_j)\right]\left[\sigma^2(w_i) + \sigma^2(w_j) - \sigma^2(e_i) - \sigma^2(e_j)\right]}}. \tag{5}$$

From this it becomes apparent that correcting the variance components in the denominator of Equation 5 for measurement error can lead to both approaching zero whenever error is high relative to additive genetic variance. The latter will be the case whenever correlations are based on statistically nonsignificant variance for epistasis and expected fitness. In such cases, observed correlations run the greatest risk of being inflated. For example, in the extreme scenario when additive genetic variance = 0, it follows from Equation 4 that the uncorrected correlation between epistasis and expected fitness, assuming equal error variances in single and double mutants, $= -2/\sqrt{3*2} = -0.82$ purely due to measurement error. Generally, statistical significance of the correlation therefore needs to be evaluated using data-resampling techniques.

Although we here focus on negative epistasis of beneficial mutations, the described effect would also generate a pattern where combinations of mutations with increasingly deleterious effects show more positive epistasis. Thereby it may partially explain the lack of empirical support for stronger negative epistasis of increasingly deleterious mutations (*i.e.*, the opposite pattern) as a selective agent maintaining sexual reproduction and recombination (Elena and Lenski 1997; Bonhoeffer *et al.* 2004). Furthermore, because the correlation between E_{ij} and $[w_i + w_j]$ is a direct function of the amount of measurement error, variation in the latter can introduce differences in the strength of the correlation among experiments. For example, because the ratio of environmental to genetic variance for fitness often differs between benign and stressful environments (Hoffmann and Merilä 1999; Agrawal and Whitlock 2010), it might erroneously be concluded that epistasic effects are shaped by environmental quality. Similarly, diminishing-returns epistasis might be found to be more pronounced in complex

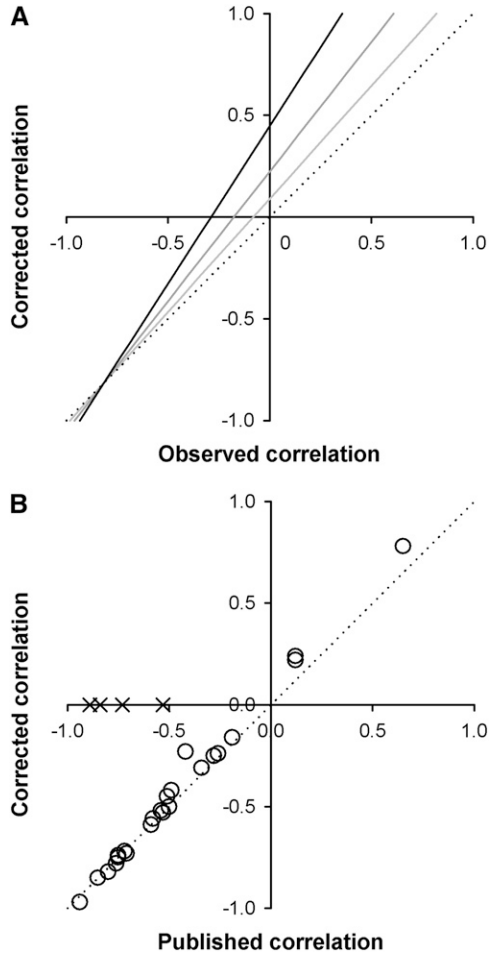

Figure 1 The relationship between observed and corrected estimates of diminishing-returns epistasis. (A) Predicted relationship between $r_{E_{ij},[w_i+w_j]}$ (*i.e.*, the observed correlation; Equation 2) and $r_{i_{ij},[a_i+a_j]}$ (the corrected correlation; Equation 5), showing that for moderately negative and positive correlations, the observed correlation is biased downward, whereas for very strongly negative correlations it is biased upward. For illustrative purposes, $\sigma^2(a_i) = \sigma^2(a_j) = 1$ and $\sigma^2(i_{ij}) = 0.5$. Lines with light shading, lines with dark shading, and solid lines represent $\sigma^2(e_i) = \sigma^2(e_j) = \sigma^2(e_{ij}) = 0.1$, 0.2, and 0.5, respectively. The dotted line designates the 1:1 relationship between observed and corrected estimates. (B**)** Published and corrected estimates of 25 correlations from 15 studies for which variance components were available (see Table S1 and Table S2). The line designates a 1:1 relationship between the published and corrected correlations. An additional four correlations (crosses) could not be corrected due to nonsignificant epistatic variance. Their placement on the *x*-axis shows their published values. Although there is overall evidence for diminishing-returns epistasis from this body of literature (correlations are still strongly negative on average, following correction), regression to the mean has led to downwardly biased estimates and four cases of published negative correlations based on nonsignificant epistatic variance.

organisms, in which fitness is often estimated with less precision.

The effect outlined here is referred to as "regression to the mean" (Galton 1886) and is a common cause of misinterpretation in biology (Kelly and Price 2005; Postma 2006, 2011; Roff 2011; Verhulst *et al.* 2013) and other sciences (*e.g.*, Hotelling 1933; Kahneman and Tversky 1973). Although we have focused on one way the phenomenon can introduce biases, it may raise its head in other ways. First, we note that although Equations 4 and 5 need to be modified if epistasis instead is defined in relative terms or by a multiplicative model and/or if epistasis is regressed on the fitness of the genetic background into which new mutations were introduced, bias remains (Table S1). Second, whenever only mutations with relatively strong effects are selected from a larger sample (as is often the case; Table S1) and if fitness is measured with error, these mutations will on average have lower fitness when measured again and therefore show apparent negative epistasis. Third, even when mutant fitness is estimated without error, estimates may be biased when combinations of beneficial mutations are selected for further investigation from experimental evolution studies (which is common too; Table S1). This is because mutations with large positive epistasis for fitness on the particular genetic background of the experimental population are more likely to fix during experimental evolution and thereby to be selected for introduction into other genetic backgrounds, where they will on average have lower fitness (Draghi and Plotkin 2013; Chou *et al.* 2014; Greene and Crona 2014).

A Brief Literature Survey

Although some authors seem aware of the issue, few have attempted to account for it (Table S1), and the severity of the bias hence remains unknown. We reviewed 30 recent articles that reported results on diminishing-returns epistasis for fitness in microorganisms. In 22 studies, epistasis was directly related to expected fitness, and 18 of these did so without correcting for regression to the mean (Table S1). We note that only one study (Szafraniec *et al.* 2003) looked for diminishing-returns epistasis by regressing observed on predicted fitness, using reduced major axis regression and testing for a slope significantly <1. This method, although not free of problematic assumptions regarding the nature of error variances (see Warton *et al.* 2006; Smith 2009), is more robust to the issue raised here. From 15 studies estimating $r_{E_{ij},[w_i+w_j]}$ we were able to extract estimates of variance components (for details see Table S1, and for a numerical example see Table S2), which allowed us to obtain unbiased estimates of 25 published correlations, using Equation 5. In four additional cases, correlations could not be corrected because of nonsignificant epistatic variance. In these cases, (almost) all variation in E_{ij} is the result of measurement error variance, resulting in corrected correlations taking on values outside the theoretical boundary (see Table S1). The fact that these correlations were strongly negative before correction, together with the fact that most corrected correlations are less negative than the published estimates, shows that regression to the mean introduces directional bias into empirical estimates of diminishing-returns epistasis (Figure 1B, Table S1). In most cases however, corrections did not affect results qualitatively, which can be attributed to mutant fitness typically being estimated with small error.

Conclusion

Here we have shown how biases due to regression to the mean inflate estimates of diminishing-returns epistasis. Although the majority of studies have not corrected for this, biases are in most cases small. Nevertheless, we do observe bias, most notably with four cases of published negative correlations based on nonsignificant epistatic variances, underlining the importance of performing corrections to allow accurate comparative analyses and prevent publication bias. We also note that we may have underestimated the amount of bias by assuming uncorrelated measurement errors, an assumption that is often violated in experiments by uncontrolled temporal or spatial block effects. Crucially, such effects would lead to undetected measurement errors that would overestimate diminishing-returns epistasis further. Application of appropriate statistical corrections in future studies will further increase our understanding of the manifestation and role of diminishing-returns epistasis in evolution.

Acknowledgments

We thank B. Rogell and C. Rueffler for helpful discussions.

Literature Cited

Agrawal, A. F., and M. C. Whitlock, 2010 Environmental duress and epistasis: How does stress affect the strength of selection on new mutations? Trends Ecol. Evol. 25: 450–458.

Bonhoeffer, S., C. Chappey, N. T. Parkin, J. M. Whitcomb, and C. J. Petropoulos, 2004 Evidence for positive epistasis in HIV-1. Science 306: 1547–1550.

Chou, H., H. Chiu, N. F. Delaney, D. Segre, and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science 332: 1190–1192.

Chou, H., N. F. Delaney, J. A. Draghi, and C. J. Marx, 2014 Mapping the fitness landscape of gene expression uncovers the cause of antagonism and epistasis between adaptive mutations. PLoS Genet. 10: e1004149.

Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Hartl, and J. F. Ayroles, 2013 Genetic incompatibilities are widespread within species. Nature 504: 135–137.

da Silva, J., M. Coetzer, R. Nedellec, C. Pastore, and D. E. Mosier, 2010 Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. Genetics 185: 292–303.

De Visser, J. A. G. M., T. F. Cooper, and S. F. Elena, 2011 The causes of epistasis. Proc. Biol. Sci. 278: 3617–3624.

Draghi, J. A., and J. B. Plotkin, 2013 Selection biases the prevalence and type of epistasis along adaptive trajectories. Evolution 67: 3120–3131.
Elena, S. F., and R. E. Lenski, 1997 Test of synergistic interactions among deleterious mutations in bacteria. Nature 390: 395–398.
Flint, C., and T. F. C. Mackay, 2009 Genetic architecture of quantitative traits in mice, flies, and humans. Genome Res. 19: 723–733.
Galton, F., 1886 Regression to mediocrity in hereditary stature. J. Anthropol. Inst. 15: 246–263.
Greene, D., and K. Crona, 2014 The changing geometry of a fitness landscape along an adaptive walk. PLoS Comput. Biol. 10: e1003520.
Hansen, T. F., 2013 Why epistasis is important for selection and adaptation. Evolution 67: 3501–3511.
Hoffmann, A. A., and J. Merilä, 1999 Heritable variation and evolution under favourable and unfavourable conditions. Trends Ecol. Evol. 14: 96–101.
Hotelling, H., 1933 Review of *The Triumph of Mediocrity in Business*. J. Am. Stat. Assoc. 28: 463–465.
Huang, W., S. Richards, M. A. Carbone, D. Zhy, R. R. H. Anholt *et al.*, 2012 Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. Proc. Natl. Acad. Sci. USA 109: 15553–15559.
Kahneman, D., and A. Tversky, 1973 On the psychology of prediction. Psychol. Rev. 80: 237–251.
Kelly, C., and T. D. Price, 2005 Correcting for regression to the mean in behavior and ecology. Am. Nat. 166: 700–707.
Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. Science 332: 1193–1196.
Martin, G., S. F. Elena, and T. Lenormand, 2007 Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nat. Genet. 39: 555–560.
Olson-Manning, C. F., M. R. Wagner, and T. Mitchell-Olds, 2012 Adaptive evolution: evaluating empirical support for theoretical predictions. Nat. Rev. Genet. 13: 867–877.
Phillips, P. C., S. P. Otto, and M. C. Whitlock, 2000 Beyond the average: the evolutionary importance of gene interactions and epistatic effects, pp. 20–38 in *Epistasis and the Evolutionary Process*, edited by J. B. Wolf, E. D. Brodie, and M. J. Wade. Oxford University Press, Oxford.
Postma, E., 2006 Implications of the difference between true and predicted breeding values for the study of natural selection and micro-evolution. J. Evol. Biol. 19: 309–320.
Postma, E., 2011 Comment on "Additive genetic breeding values correlate with the load of partially deleterious mutations". Science 333: 1221-a.
Roff, D. A., 2011 Measuring the cost of plasticity: a problem of statistical non-independence. Proc. R. Soc. Lond. Ser. B 278: 2724–2725; discussion 2726–2727.
Smith, R. J., 2009 Use and misuse of the reduced major axis for line-fitting. Am. J. Phys. Anthropol. 140: 476–486.
Szafraniec, K., D. M. Wloch, P. Sliwa, R. H. Borts, and R. Korona, 2003 Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast Saccharomyces cerevisiae. Genet. Res. 82: 19–31.
Szendro, I. G., M. F. Schenk, J. Franke, J. Krug, and J. A. G. M. de Visser, 2013 Quantitative analyses of empirical fitness landscapes. J. Stat. Mech. 2013: P01005.
Warton, D. I., I. J. Wright, D. S. Falster, and M. Westoby, 2006 Bivariate line-fitting methods for allometry. Biol. Rev. Camb. Philos. Soc. 81: 259–291.
Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. Science 312: 111–114.
Whitlock, M. C., P. C. Phillips, F. B. G. Moore, and S. Tonsor, 1995 Multiple fitness peaks and epistasis. Annu. Rev. Ecol. Syst. 26: 601–629.
Verhulst, S., A. Aviv, A. Benetos, G. S. Berenson, and J. D. Kark, 2013 Do leukocyte telomere length dynamics depend on baseline telomere length? An analysis that corrects for regression to the mean. Eur. J. Epidemiol. 11: 859–866.

Communicating editor: E. A. Stone

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.169870/-/DC1

Biased Estimates of Diminishing-Returns Epistasis? Empirical Evidence Revisited

David Berger and Erik Postma

DOI: 10.1534/genetics.114.169870

Studies reporting relationships between epistasis and background fitness

Table S1 Summary of the 30 reviewed studies reporting an empirical relationship between epistasis and background fitness. These were identified by means of a Google Scholar search for "*diminishing returns epistasis*". We chose for further review the first 30 papers, found either through this search or because they were cited in one of these papers, reporting a relationships between epistasis and expected fitness. Information is given on the type of organism and experimental setting that was used, whether mutations had beneficial or detrimental effects, the locations of the mutations, and under which circumstances selection on mutations was measured. Approximate F- and P-values are given for the epistatic variance components, on which the published correlations were based. The type of test reported in the original publication is given, and if the study corrected for regression-to-the-mean. In a few studies neither point estimates nor significance tests were provided, but instead conclusions in these studies were based on graphical exploration of the relationship between expected fitness and epistasis, and hence they were included here. For all studies for which error variances could be extracted, the published correlation is given along with an estimate of the corrected correlation using Eq. 5 in the main text.

Study	Organism	Setting	Effects	Location	Selection	Appr.F	Appr.P	Published r	Corrected r	Test	Correction
Caudle et al. 2014	ssDNA Bacteriophage ID11	9 isolated mutations tested in novel environment (41°C)	beneficial	different	general	$F_{17,72} = 49.4$	< 0.001	-0.72	-0.72	$r_{E_{ij},[w_i+w_j]}$	n.a
		9 isolated mutations tested in ancestral environment (37°C)	beneficial	different	general	$F_{17,72} = 8.4$	< 0.001	0.12	0.24	$r_{E_{ij},[w_i+w_j]}$	n.a
		9 isolated mutations tested in novel environment (33°C)	both	different	general	$F_{17,72} = 4.6$	< 0.001	-0.54	-0.53	$r_{E_{ij},[w_i+w_j]}$	n.a
Perfeito et al. 2014	*Escherichia coli*	Fitness increase following mutation accumulation in 23 genetic backgrounds	beneficial	different	general	$F_{22,46} = 6.8$	< 0.001	-0.76	-0.78	$r_{w_i,[w_{ij}-w_i]}$	no
		Fitness decline in 23 genotypes during mutation accumulation	deleterious	different	general	$F_{22,46} = 5.9$	< 0.001	-0.34	-0.31	$r_{w_i,[w_{ij}-w_i]}$	no
Elena & Lenski 1997	*Escherichia coli*	Single, double and triple mutants from 225 genotypes generated by mutagenesis	deleterious	different	general						n.a
Trindade et al. 2009	*Escherichia coli*	19 mutations conferring antibiotic resistance, tested in absence of drugs	deleterious	different	general	$F_{102,309} = 14.6$	< 0.001	-0.19	-0.16	$r_{E_{ij},[w_i+w_j]}$	no
Trindade et al. 2012	*Escherichia coli*	25 mutations conferring antibiotic resistance, tested in absence of drugs	deleterious	different	general						n.a
Bonhoeffer et al.	HIV-1	Natural variation in 9466	deleterious	different	general					selection	no

2004		isolated natural strains								
Burch & Chao 2004	RNA bacteriophage Φ6	Fitness decline during mutation accumulation	deleterious	different	general				$r_{w_i,[w_{ij}-w_i]}$	no
Kryazhimskiy et al. 2014	*Saccharomyces cerevisiae*	Fitness increase during experimental evolution, and fitness decline of gene knockouts, in 64 genetic backgrounds	beneficial	different	general				$r_{w_i,[w_{ij}-w_i]}$	yes
Kvitek & Sherlock 2011	*Saccharomyces cerevisiae*	15 mutations isolated during experimental evolution, tested in ancestral environment	beneficial	different	general					n.a
Szafraniec et al. 2003	*Saccharomyces cerevisiae*	74 genotypes generated by mutagenesis	deleterious	different	general				$b_{w_{ij},[w_i+w_j]} < 1$	yes
Jasnos et al 2008	*Saccharomyces cerevisiae*	Multiple single and double deletions generated through genetic engineering, tested in 5 environments	deleterious	different	general					n.a
Xu et al. 2012	*Saccharomyces cerevisiae*	Comparative meta-data	deleterious	different	general				selection	no
Pearson et al. 2012	ssDNA Bacteriophage ID11	Assayed isolated mutation F on 8 fitness backgrounds	beneficial	different	general	$F_{7,32} = 10.0$	< 0.001	-0.71	-0.73	r_{E_{ij},w_i}	no
		Assayed isolated mutation B on 8 fitness backgrounds	beneficial	different	general	$F_{7,32} = 25.1$	< 0.001	-0.53	-0.53	r_{E_{ij},w_i}	no
Rokyta et al. 2011	ssDNA Bacteriophage ID11	9 mutations isolated from directed mutagenesis	beneficial	different	general	$F_{17,72} = 8.7$	< 0.001	0.12	0.22		n.a
Lalic & Elena 2012	Tobacco potyvirus	20 single mutants combined to create 53 double mutants	deleterious	different	general	$F_{42,344} = 2.9$	< 0.001	-0.42	-0.23	$r_{E_{ij},[w_i+w_j]/2}$	no
Sanjuan et al. 2005	Vesicular stomatitis virus	Fitness increase in 12 genetic backgrounds following mutation accumulation	beneficial	different	general	$F_{11,48} = 8.7$	< 0.001	-0.49	-0.42	$r_{w_i,[w_{ij}/w_i]}$	no
Martin et al. 2007	Vesicular stomatitis virus	Multiple mutations generated by mutagenesis, tested in ancestral environment	both	different	general				selection	no
Sanjuan et al. 2004	Vesicular stomatitis virus	47 genotypes generated by mutagenesis	both	different	general				selection	no
Hall & MacLean 2011	*Pseudomonas aeruginosa*	14 mutations in antibiotic resistance gene, tested in absence of antibiotics	deleterious	same gene	general	$F_{13,98} = 3.0$	0.001	-0.51	-0.45	$r_{E_{ij},[w_iw_j]}$	no
		14 mutations in antibiotic resistance gene, tested in absence of antibiotics	deleterious	same gene	general			-0.55		$r_{E_{ij},[w_iw_j]}$	no
		14 mutations in antibiotic resistance gene, tested in	deleterious	same gene	general			-0.45		$r_{E_{ij},[w_iw_j]}$	no

		absence of antibiotics									
Hayden & Wagner 2012	Azoarcus group 1 ribozyme	4 isolated mutations tested in 3 environments	beneficial	different	specific						n.a
Flynn et al. 2013	*Escherichia coli*	5 mutations isolated during experimental evolution, tested in novel environment	beneficial	different	specific	$F_{31,128} = 4.6$	< 0.001	-0.75	-0.74	$r_{E_{ij},[w_i+w_j]}$	no
		5 mutations isolated during experimental evolution, tested in novel environment	beneficial	different	specific	$F_{17,72} = 17.2$	< 0.001	-0.28	-0.25	$r_{E_{ij},[w_i+w_j]}$	no
Khan et al. 2011	*Escherichia coli*	5 mutations isolated during experimental evolution, tested in ancestral environment	beneficial	different	specific	$F_{25,364} = 4.6$	< 0.001	-0.58	-0.56	$r_{E_{ij},[w_i+w_j]}$	no
		rbs-mutation	beneficial	different	specific	$F_{15,224} = 10.5$	< 0.001	-0.26	-0.24	$r_{w_i,[w_{ij}-w_i]}$	no
		topi-mutation	beneficial	different	specific	$F_{15,224} = 13.9$	< 0.001	-0.59	-0.59	$r_{w_i,[w_{ij}-w_i]}$	no
		spoT-mutation	beneficial	different	specific	$F_{15,224} = 16.6$	< 0.001	-0.50	-0.50	$r_{w_i,[w_{ij}-w_i]}$	no
		glmUS-mutation	beneficial	different	specific	$F_{15,224} = 13.0$	< 0.001	-0.50	-0.50	$r_{w_i,[w_{ij}-w_i]}$	no
		pykF-mutation	beneficial	different	specific	$F_{15,224} = 5.2$	< 0.001	0.65	0.78	$r_{w_i,[w_{ij}-w_i]}$	no
Wang et al. 2012	*Escherichia coli*	2 mutations isolated during experimental evolution, tested in ancestral environment on 8 different genetic backgrounds	beneficial	different	specific			-0.81		r_{E_{ij},w_k}	yes
Bull et al 2000	Bacteriophage ΦX174	Benefit of new mutation on 4 genetic backgrounds evolved under heat stress	beneficial	same pathway	specific					$r_{w_i,[w_{ij}-w_i]}$	no
Schenk et al. 2013	*Escherichia coli*	4 antibiotic resistance mutations	beneficial	same pathway	specific	$F_{10,33} > 1000$	< 0.001	-0.85	-0.85	$r_{E_{ij},[w_i+w_j]}$	no
		4 antibiotic resistance mutations	beneficial	same pathway	specific	$F_{10,44} > 1000$	< 0.001	-0.75	-0.75	$r_{E_{ij},[w_i+w_j]}$	no
da Silva et al. 2010	HIV-1	7 mutations generated by directed mutagenesis, tested in ancestral environment	both	same pathway	specific						n.a
Chou et al. 2011	*Methylobacterium extorquens*	4 mutations isolated during experimental evolution, tested in ancestral environment	beneficial	same pathway	specific	$F_{31,64} = 15.8$	< 0.001	-0.8	-0.82	$r_{w_i,[w_{ij}/w_i]}$	no
		fghA-mutation	beneficial	same pathway	specific	$F_{7,16} = 2.2$	0.09	-0.89	undefined	$r_{w_i,[w_{ij}/w_i]}$	no
		pntAB-mutation	beneficial	same pathway	specific	$F_{7,16} = 0.58$	0.76	-0.53	undefined	$r_{w_i,[w_{ij}/w_i]}$	no
		gshA-mutation	beneficial	same pathway	specific	$F_{7,16} = 5.0$	0.004	-0.94	-0.97	$r_{w_i,[w_{ij}/w_i]}$	no

		GB-mutation	beneficial	same pathway	specific	$F_{7,16} = 2.2$	0.09	-0.84	undefined	$r_{w_i,[w_{ij}/w_i]}$	no
Chou et al. 2014	*Methylobacterium extorquens*	4 mutations isolated during experimental evolution, tested in ancestral environment	beneficial	same pathway	specific	$F_{27,56} = 289.7$	< 0.001	-0.53	-0.53	$r_{E_{ij},[\,w_i w_j]}$	yes
MacLean 2009	*Pseudomonas aeruginosa*	Experimental activation of two antibiotic resistance pathways	both	same pathway	specific	$F_{7,21} = 1.9$	0.12	-0.73	undefined	$r_{E_{ij},[\,w_i w_j]}$	no
MacLean et al. 2010	*Pseudomonas aeruginosa*	Antibiotic resistance increase in three different genetic backgrounds	beneficial	same pathway	specific			-1		$r_{w_i,[w_{ij}/w_i]}$	no

Setting: See original publications for more detail.

Location: *same gene* refers to the authors reporting that most studied mutations were located physically in the same gene, *same pathway* refers to mutations affecting the same physiological pathway (as often was the case for beneficial mutations isolated during experimental evolution on a specific growth medium), and *different* refers to mutations that were random and often their effects and locations were not known a priori.

Selection: *direct* corresponds to when fitness of mutations was assessed under the same specific selective conditions as they first were identified in (as for fitness of mutations conferring antibiotic resistance tested on a growth medium containing the antibiotic), and *general* corresponds to when fitness of mutations were scored in less specific conditions (as for fitness of antibiotic resistance mutations on a growth medium not containing the antibiotic, or for random mutations acquired by mutagenesis or through mutation accumulation experiments).

Appr.F: an approximation of the F-ratio for the epistatic interaction variance [i.e. $\sigma^2(E_{ij}) / \sigma^2(e_i+e_j+e_{ij})$ for the case when absolute epistasis had been estimated]. The first term in the subscript of F gives the degrees of freedom for the effect of mutant genotype ($n_{genotypes}$-1), and the second term gives the degrees of freedom for the error term ($n_{tot} – n_{genotypes}$).

Appr.P: the accompanying approximation of the P-value for the epistatic interaction variance. Note that no P-values were calculated for the corrected correlations as it would require simulation and resampling using the original datasets (see "Correcting the correlations for measurement error" below for further details).

Test: $r_{E_{ij},[w_i+w_j]}$, $r_{E_{ij},[w_i+w_j]/2}$ and $r_{E_{ij},[\,w_i w_j]}$ refer to a correlation between epistasis and expected fitness, r_{E_{ij},w_i} refers to a correlation between epistasis and background fitness, $r_{w_i,[w_{ij}/w_i]}$ and $r_{w_i,[w_{ij}-w_i]}$ refer to a correlation between background fitness and fitness improvement or decline, $b_{w_{ij},[w_i+w_j]} < 1$ refers to a test of a major axis (MA) regression slope of observed fitness on expected fitness of double mutants being significantly

below 1, *selection* refers to subsampling of mutations based on their fitness effects and subsequent comparisons of the strength of epistasis between mutations with above and below average fitness effects.

Correction: whether the statistical method itself, or any additional measure was taken to reduce or correct for the effect of regression-to-the-mean. *n.a* refers to cases where relationships with epistasis were not directly tested by regressions using background or expected fitness. Many of these studies instead compared epistasis across high and low quality environments, or between high and low order gene interaction.

Correcting the correlations for measurement error

As outlined in the main text, the correlation between [$w_i + w_j$] and E_{ij} is a function of not only $\sigma^2(a_i + a_j)$, $\sigma^2(i_{ij})$ and $\sigma(i_{ij}, [a_i+a_j])$, but also of $\sigma^2(e_i + e_j)$ and $\sigma^2(e_{ij})$ (see eqn. 4). Having knowledge of $\sigma^2(e_i + e_j)$ and $\sigma^2(e_{ij})$, we are able to obtain the corrected correlation between i_{ij} and [$a_i + a_j$] that is not biased by measurement error variance, using eqn. 5 (see main text for more details). In the main text we derive eq. 5 for the case when the expected fitness of double mutants assuming purely additive effects ($w_i + w_j$) is correlated with the absolute amount of epistasis ($E = w_{ij} - [w_i + w_j]$). However, some studies used a relative measure of epistasis (i.e. $w_{ij} / [w_i * w_j]-1$), or they used the absolute (i.e. $w_{ij} - w_i$) or relative (i.e. w_{ij} / w_i-1) fitness improvement associated with introducing mutation *j* into a genetic background containing mutation *i* (see table S1 for further details). To accommodate this, eq. 5 was modified appropriately.

In cases where the true variance between mutants is low (small $\sigma^2(a_i)$, $\sigma^2(a_j)$ and $\sigma^2(a_{ij})$ relative to measurement error ($\sigma^2(e_i)$, $\sigma^2(e_j)$ and $\sigma^2(e_{ij})$), it becomes clear that correcting the variance components in the denominator inflates the corrected correlation to take on extreme values. Hence, correlations based on non-significant variance components will be erroneous, and confidence limits and significance of correlations needs to be estimated using data resampling techniques. Therefore we calculated the F-ratio and accompanying P-value for the epistatic variance component using $\sigma^2(E_{ij}) / \sigma^2(e_i+e_j+e_{ij})$ as an indicator of the reliability of the published estimate of the correlation. Indeed, in four cases the epistatic variance was approximated to be non-significant (Table S1), and correction lead to the correlation taking on values outside the theoretical boundary ($r < -1$) due to the corrected epistatic variance (i_{ij}), present in the denominator of Eq. 5, approaching zero.

Approximate measurement error variances were derived from the mean squared standard errors of mean fitness for both single and double mutants. In cases where we could not find separate estimates of error variance for both single and double mutants, equal error variances were assumed. Estimates of errors in expected fitness ($\sigma^2(e_i + e_j)$) and epistasis ($\sigma^2(e_i+e_j+e_{ij})$) were obtained through error propagation of single and double mutant estimation errors.

It is clear that these corrected correlations and F-values only serve as approximations. Indeed, in all our corrections we estimated measurement error variance from the mean standard error across all measured genotypes for one class of mutant (single or double). Thereby we assumed that measurement error was the same for all genotypes of one class of mutant. In addition, in Khan et al. (2011), Chou et al. (2011;2014), Flynn et al. (2013) and Schenk et al. (2013), complex higher-order epistatic interactions between 2 to 5 mutations were studied, requiring more elaborate statistical corrections using resampling methods to arrive at exact estimates of the correlations and their statistical significance. Nevertheless, our analysis shows that for the majority of studies measurement error variance is relatively small, and as a consequence, correction has little effect on the qualitative conclusions drawn from the combined body of literature reviewed here concerning diminishing returns epistasis (Table S1).

References

Bonhoeffer S, Chappey C, Parkin NT, Whitcomb JM, Petropoulos CJ. 2004. Evidence for positive epistasis in HIV-1. Science 306: 1547-1550.

Bull JJ, Badgett MR, Wichman HA. 2000. Big-benefit mutations in a bacteriophage inhibited with heat. Molecular Biology and Evolution 17: 942-950.

Burch CL, Chao L. 2004. Epistasis and its relationship to canalizations in the RNA virus Φ6. Genetics 167: 559-567.

Caudle SB, Miller CR, Rokyta DR. 2014. Environment determines epistatic patterns for a ssDNA virus. Genetics 196: 267-279.

Chou H, Chiu H, Delaney NF, Segre D, Marx CJ. 2011. Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science 332: 1190-1192.

Chou H, Delaney NF, Draghi JA, Marx CJ. 2014. Mapping the fitness landscape of gene expression uncovers the cause of antagonism and epistasis between adaptive mutations. PLoS Genetics 10: e1004149.

da Silva J, Coetzer M, Nedellec R, Pastore C, Mosier DE. 2010. Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. Genetics 185: 292-303.

Elena SF, Lenski RE. 1997. Test of synergistic interactions among deleterious mutations in bacteria. Nature 395-398.

Flynn KM, Cooper TF, Moore FGB, Cooper VS. 2013. The environment affects epistatic interactions to alter the topology of an empirical fitness landscape. PLoS Genetics 9: e1003426.

Hall AR, MacLean RC. 2011. Epistasis buffers the fitness effects of Rifampicin-resitance mutations in Pseudomonas aeruginosa. Evolution 65: 2370-2379.

Jasnos L, Tomala K, Paczesniak D, Korona R. 2008. Interactions between stressful environment and gene deletions alleviate the expected average loss of fitness in yeast. Genetics 178: 2105-2111.

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.

Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014. Global epistasis makes adaptation predictable despite sequence-level stochasticity. bioRxiv, doi: 10.1101/001784.

Kvitek DJ, Sherlock G. 2011. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genetics 7: e1002056.

Lalic J, Elena SF. 2012. Magnitude and sign epistasis among deleterious mutations in a positive-sense plant RNA virus. Heredity 109: 71-77.

MacLean RC. 2009. Predicting epistasis: an experimental test of metabolic control theory with bacterial transciption and translation. Journal of Evolutionary Biology 23: 488-493.

MacLean RC, Perron GG, Gardner A. 2010. Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for Rifampicin resistance in Pseudomonas aeruginosa. Genetics 186: 1345-1354.

Martin G, Elena SF, Lenormand T. 2007. Distributions of epistasis in microbes fit predictions from a fitness landscape model. Nature Genetics 39: 555-560.

Martinez JP, Bocharov G, Ignatovich A, Reiter J, Dittmar MT, Wain-Hobson S, Meyerhans A. 2011. Fitness ranking of individual mutants drives patterns of epistatic interactions in HIV-1. PLoS ONE 6: e18375.

Pearson VM, Miller CR, Rokyta DR. 2012. The constancy of beneficial fitness effects of mutations across diverse genetic backgrounds. PLOS ONE 7: e43864.

Perfeito L, Sousa A, Bataillon T, Gordo I. 2013. Rates of fitness decline and rebound suggest pervasive epistasis. Evolution 68: 150-162.

Rokyta DR, Joyce P, Caudle SB, Miller C, Beisel CJ, Wichman HA. 2011. Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. PLoS Genetics 7: e1002075.

Sanjuan R, Cuevas JM, Moya A, Elena SF. 2005. Epistasis and the adaptability of an RNA virus. Genetics 170: 1001-1008.

Sanjuan R, Moya A, Elena SF. 2004. The contribution of epistasis to the architecture of fitness in an RNA virus. Proceedings of the National Academy of Sciences USA 101: 15376-15379.

Schenk MF, Szendro IG, Salverda MLM, Krug J, de Visser JAGM. 2013. Patterns o fepistasis between beneficial mutations in an antibiotic resistance gene. Molecular Biology and Evolution 30: 1779-1787.

Szafraniec K, Wloch DM, Sliwa P, Borts RH, Korona R. 2003. Small fitness effects and weak genetic interactions between deleterious mutations in heterozygous loci of the yeast Saccharomyces cerevisiae. Genetical Reseach of Cambridge 82: 19-31.

Trindade S, Sousa A, Xavier KB, Dionisio F, Ferreira MG, Gordo I. 2009. Positive epistasis drives the acquisition of multidrug resistance. PLoS Genetics 5: e 1000578.

Trindade S, Sousa A, Gordo I. 2012. Antibiotic resitance and stress in the light of Fisher's model. Evolution 66: 3815-3824.

Wang Y, Arenas CD, Stoebel DM, Cooper TF. 2012. Genetic background affects epistatic interactions between two beneficial mutations. Biology Letters 9: 20120328.

Xu L, Barker B, Gu Z. 2012. Dynamic epistasis for different alleles of the same gene. Proceedings of the National Academy of Sciences USA 109: 10420-10425.

Table S2 Numerical example of (post-hoc) correction for measurement error variance

Available for download as an Excel file at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.114.169870/-/DC1