bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Detecting high-order epistasis in nonlinear genotype-phenotype maps

Zachary R. Sailer[1,2], Michael J. Harms[1,2,*]

1. Institute of Molecular Biology, University of Oregon, Eugene, OR, USA

2. Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, USA

* harms@uoregon.edu

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

High-order epistasis has been observed in many genotype-phenotype maps. These multi-way interactions between mutations may be useful for dissecting complex traits and could have profound implications for evolution. Alternatively, they could be a statistical artifact. High-order epistasis models assume the effects of mutations should add, when they could in fact multiply or combine in some other nonlinear way. A mismatch in the "scale" of the epistasis model and the scale of the underlying map would lead to spurious epistasis. In this paper, we develop an approach to estimate the nonlinear scales of arbitrary genotype-phenotype maps. We can then linearize these maps and extract high-order epistasis. We investigated seven experimental genotype-phenotype maps for which high-order epistasis had been reported previously. We find that five of the seven maps exhibited nonlinear scales. Interestingly, even after accounting for nonlinearity, we found statistically significant high-order epistasis in all seven maps. The contributions of high-order epistasis to the total variation ranged from 2.2% to 31.0%, with an average across maps of 12.7%. Our results provide strong evidence for extensive high-order epistasis, even after nonlinear scale is taken into account. Further, we describe a simple method to estimate and account for nonlinearity in genotype-phenotype maps.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Recent analyses of genotype-phenotype maps have revealed "high-order" epistasis—that is, interactions between three, four, and even more mutations (Ritchie et al. 2001; Segrè et al. 2005; Xu et al. 2005; Tsai et al. 2007; Imielinski and Belta 2008; Matsuura et al. 2009; da Silva et al. 2010; Pettersson et al. 2011; Wang et al. 2012; Weinreich et al. 2013; Hu et al. 2013; Sun et al. 2014; Anderson et al. 2015; Yokoyama et al. 2015). The importance of these interactions for understanding biological systems and their evolution is the subject of current debate (Poelwijk et al. 2016; Weinreich et al. 2013). Can they be interpreted as specific, biological interactions between loci? Or are they misleading statistical correlations?

We set out to tackle one potential source of spurious epistasis: a mismatch between the "scale" of the map and the scale of the model used to dissect epistasis (Fisher 1918; Rothman et al. 1980; Frankel and Schork 1996; Cordell 2002; Phillips 2008; Szendro et al. 2013). The scale defines how to combine mutational effects. On a linear scale, the effects of individual mutations are added. On a multiplicative scale, the effects of mutations are multiplied. Other, arbitrarily complex scales, are also possible (Rokyta et al. 2011; Schenk et al. 2013; Blanquart 2014).

Application of a linear model to a nonlinear map will lead to apparent epistasis (Fisher 1918; Rothman et al. 1980; Frankel and Schork 1996; Cordell 2002; Phillips 2008; Szendro et al. 2013). Consider a map with independent, multiplicative mutations. Analysis with a multiplicative model will give no epistasis. In contrast, analysis with a linear model will give epistatic coefficients to account for the multiplicative nonlinearity (Cordell 2002; Phillips 2008). Epistasis arising from a mismatch in scale is mathematically valid, but obscures a key feature of the map: its scale. It is also not parsimonious, as it uses many coefficients to describe a potentially simple nonlinear function. Finally, it can be misleading because these epistatic coefficients partition global information about the nonlinear scale into (apparently) specific interactions between mutations.

Most high-order epistasis models assume a linear scale (or a multiplicative scale transformed onto a linear scale) (Heckendorn and Whitley 1999; Szendro et al. 2013; Weinreich et al. 2013; Poelwijk et al. 2016). These models sum the independent effects of mutations to predict multi-mutation

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

phenotypes. Epistatic coefficients account for the difference between the observed phenotypes and the phenotypes predicted by summing mutational effects. The epistatic coefficients that result are, by construction, on the same linear scale (Poelwijk et al. 2016; Weinreich et al. 2013; Heckendorn and Whitley 1999).

Because the underlying scale of genotype-phenotype maps is not known *a priori*, the interpretation of high-order epistasis extracted on a linear scale is unclear. If a nonlinear scale can be found that removes high-order epistasis, it would suggest that high-order epistasis is spurious: a highly complex description of a simple, nonlinear system. In contrast, if no such scale can be found, high-order epistasis provides a window into the profound complexity of genotype-phenotype maps.

In this paper, we set out to estimate the nonlinear scales of experimental genotype-phenotype maps. We then account for these scales in the analysis of high-order epistasis. We took our inspiration from the treatment of multiplicative maps, which can be transformed into linear maps using a log transform. Along these same lines, we set out to transform genotype-phenotype maps with arbitrary, nonlinear scales onto a linear scale for analysis of high-order epistasis. We develop our methodology using simulations and then apply it to experimentally measured genotype-phenotype maps.

Materials and Methods

Experimental data sets

We collected a set of published genotype-phenotype maps for which high-order epistasis had been reported previously. Measuring an Lth-order interaction requires knowing the phenotypes of all binary combinations of L mutations—that is, 2^L genotypes. The data sets we used had exhaustively covered all 2^L genotypes for five or six mutations. These data sets cover a broad spectrum of genotypes and phenotypes. Genotypes included point mutations to a single protein (Weinreich et al. 2006), point mutations in both members of a protein/DNA complex (Anderson et al. 2015), random genomic mutations (Khan et al. 2011; de Visser et al. 2009), and binary combinations of alleles within a biosynthetic network (Hall et al. 2010). Measured phenotypes included selection

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

coefficients (Weinreich et al. 2006; Khan et al. 2011; de Visser et al. 2009), molecular binding affinity (Anderson et al. 2015), and yeast growth rate (Hall et al. 2010). (For several data sets, the "phenotype" is a selection coefficient. We do not differentiate fitness from other properties for our analyses; therefore, for simplicity, we will refer to all maps as genotype-phenotype maps rather than specifying some as genotype-fitness maps). All data sets had a minimum of three independent measurements of the phenotype for each genotype. All data sets are available in a standardized ascii text format.

Nonlinear scale

We described nonlinearity in the genotype-phenotype map by a power transformation (see Results) (Box and Cox 1964; Carroll and Ruppert 1981). The independent variable for the transformation was $\vec{P}_{add}$, the predicted phenotypes of all genotypes assuming linear and additive affects for each mutation. The estimated additive phenotype of genotype i, is given by:

$$\hat{P}_{add,i} = \sum_{j=1}^{j \leq L} \langle \Delta P_j \rangle \, x_{i,j} \tag{1}$$

where $\langle \Delta P_j \rangle$ is the average effect of mutation j across all backgrounds, $x_{i,j}$ is an index that encodes whether or not mutation j is present in genotype i, and L is the number of sites. The dependent variables are the observed phenotypes $\vec{P}_{obs}$ taken from the experimental genotype-phenotype maps.

We use nonlinear least-squares regression to fit and estimate the power transformation from $\vec{P}_{add}$ to $\vec{P}_{obs}$:

$$\vec{P}_{obs} \sim \tau(\hat{\vec{P}}_{add}; \hat{\lambda}, \hat{A}, \hat{B}) + \hat{\varepsilon},$$

where ε is a residual and τ is a power transform function. This is given by:

$$\vec{P}_{obs} = \frac{(\hat{\vec{P}}_{add} + A)^{\lambda} - 1}{\lambda (GM)^{\lambda - 1}} + B,$$

where A and B are translation constants, GM is the geometric mean of $(\hat{\vec{P}}_{add} + A)$, and λ is a

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

scaling parameter. We used standard nonlinear regression techniques to minimize d:

$$d = (\vec{P}_{scale} - \vec{P}_{obs})^2 + \varepsilon.$$

We then reversed this transformation to linearize P_{obs} using the estimated parameters $\hat{A}$, $\hat{B}$, and $\hat{\lambda}$. We did so by the back-transform:

$$P_{obs,linear} = \{\hat{\lambda}(GM)^{\lambda-1}(P_{obs} - \hat{B}) + 1\}^{1/\hat{\lambda}} - \hat{A}. \tag{2}$$

High-order epistasis model

We dissected epistasis using a linear, high-order epistasis model. These have been discussed extensively elsewhere (Heckendorn and Whitley 1999; Poelwijk et al. 2016; Weinreich et al. 2013), so we will only briefly and informally review them here.

A high-order epistasis model is a linear decomposition of a genotype-phenotype map. It yields a set of coefficients that account for all variation in phenotype. The signs and magnitudes of the epistatic coefficients quantify the effect of mutations and interactions between them. A binary map with 2^L genotypes requires 2^L epistatic coefficients and captures all interactions, up to L^{th}-order, between them. This is conveniently described in matrix notation.

$$\vec{P} = \mathbf{X}\vec{\beta}: \tag{3}$$

a vector of phenotypes $\vec{P}$ can be transformed into a vector of epistatic coefficients $\vec{\beta}$ using a $2^L \times 2^L$ decomposition matrix that encodes which coefficients contribute to which phenotypes. If $\mathbf{X}$ is invertible, one can determine $\vec{\beta}$ from a collection of measured phenotypes by

$$\vec{\beta} = \mathbf{X}^{-1}\vec{P}. \tag{4}$$

$\mathbf{X}$ can be formulated in a variety of ways (Poelwijk et al. 2016). Following others in the genetics literature, we use the form derived from Walsh polynomials (Heckendorn and Whitley 1999; Wein-

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reich et al. 2013; Poelwijk et al. 2016). In this form, $\mathbf{X}$ is a Hadamard matrix. Conceptually, the transformation identifies the geometric center of the genotype-phenotype map and then measures the average effects of each mutation and combination of mutations in this "average" genetic background (Fig 1). To achieve this, we encoded each mutation at each site in each genotype as -1 (wildtype) or +1 (mutant) (Heckendorn and Whitley 1999; Weinreich et al. 2013; Poelwijk et al. 2016). This has been called a Fourier analysis,(Szendro et al. 2013; Neidhart et al. 2013), global epistasis (Poelwijk et al. 2016), or a Walsh space (Heckendorn and Whitley 1999; Weinreich et al. 2006). Another common approach is to use a single wildtype genotype as a reference and encode mutations as either 0 (wildtype) or 1 (mutant) (Poelwijk et al. 2016).

One data set (IV, Table I) has four possible states (A, G, C and T) at two of the sites. We encoded these using the WYK tetrahedral-encoding scheme(Zhang and Zhang 1991; Anderson et al. 2015). Each state is encoded by a three-bit state. The wildtype state is given the bits $(1, 1, 1)$. The remaining states are encoded with bits that form corners of a tetrahedron. For example, the wildtype of site 1 is G and encoded as the $(1, 1, 1)$ state. The remaining states are encoded as follows: A is $(1, -1, -1)$, C is $(-1, 1, -1)$ and T is $(1, -1, -1)$.

Experimental uncertainty

We used a bootstrap approach to propagate uncertainty in measured phenotypes into uncertainty in epistatic coefficients. To do so we: 1) calculated the mean and standard deviation for each phenotype from the published experimental replicates; 2) sampled the uncertainty distribution for each phenotype to generate a pseudoreplicate vector $\vec{P}_{pseudo}$ that had one phenotype per genotype; 3) rescaled $\vec{P}_{pseudo}$ using a power-transform; and 4) determined the epistatic coefficients for $\vec{P}_{pseudo,scaled}$. We then repeated steps 2-4 until convergence. We determined the mean and variance of each epistatic coefficient after every 50 pseudoreplicates. We defined convergence as the mean and variance of every epistatic coefficient changed by < 0.1 % after addition of 50 more pseduoreplicates. On average, convergence required $\approx 100,000$ replicates per genotype-phenotype map. Finally, we used a z-score to determine if each epistatic coefficient was significantly different than zero. To account for multiple testing, we applied a Bonferroni correction to all p-values (Abdi

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2007).

Computational methods

Our full epistasis software package—written in Python3 extended with Numpy and Scipy (van der Walt et al. 2011)—is available for download via github (https://harmslab.github.com/epistasis). We used the python package *scikit-learn* for all regression (Pedregosa et al. 2011). Plots were generated using *matplotlib* and *jupyter* notebooks (Hunter 2007; Perez and Granger 2007).

Results & Discussion

Nonlinear scale induces apparent high-order epistasis

Our first goal was to understand how a nonlinear scale, if present, would affect estimates of high-order epistasis. To probe this question, we constructed a five-site binary genotype-phenotype map on a nonlinear scale, and then extracted epistasis assuming a linear scale. The nonlinear scale we chose was saturating function:

$$P_{g,trans} = \frac{(1+K)P_g}{1+KP_g}, \tag{5}$$

where P_g is the linear phenotype of genotype g, $P_{g,trans}$ is the transformed phenotype of genotype g, and K is a scaling constant. As $K \to 0$, the map becomes linear. As K increases, mutations have systematically smaller effects when introduced into backgrounds with higher phenotypes.

We calculated P_g for all 2^L binary genotypes using the random, additive coefficients shown in Fig 2A. These coefficients included no epistasis. We then transformed P_g onto the nonlinear $P_{g,trans}$ scale using Equation 5 with the relatively shallow ($K = 2$) saturation curve shown in Fig 2B. Finally, we applied a linear epistasis model to $P_{g,trans}$ to extract epistatic coefficients.

We found that nonlinearity in the genotype-phenotype map induced extensive high-order epistasis when the nonlinearity was ignored (Fig 2C). We observed epistasis up to the fourth order, despite building the map with purely additive coefficients. This result is unsurprising: the only mechanism by which a linear model can account for variation in phenotype is through epistatic co-

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

efficients (Rothman et al. 1980; Frankel and Schork 1996; Cordell 2002). When given a nonlinear map, it partitions the variation arising from nonlinearity into specific interactions between mutations. This high-order epistasis is mathematically valid, but does not capture the major feature of the map—namely, saturation. Indeed, this epistasis is deceptive, as it is naturally interpreted as specific interactions between mutations. For example, this analysis identifies a specific interaction between mutations one, two, four, and five (Fig 2C, purple). But this four-way interaction is an artifact of the nonlinearity in phenotype of the map, rather than a specific interaction.

Nonlinear scale and specific epistatic interactions induce different patterns of non-additivity

Our next question was whether we could separate the effects of nonlinear scale and high-order epistasis in binary maps. One useful approach to develop intuition about epistasis is to plot the the observed phenotypes (P_{obs}) against the predicted phenotype of each genotype, assuming linear and additive mutational effects (P_{add}) (Rokyta et al. 2011; Schenk et al. 2013; Szendro et al. 2013). In a linear map without epistasis, P_{obs} equals P_{add}, because each mutation would have the same, additive effect in all backgrounds. If epistasis is present, phenotypes will diverge from the $P_{obs} = P_{add}$ line.

We simulated maps including varying amounts of linear, high-order epistasis, placed them onto increasingly nonlinear scales, and then constructed P_{obs} vs. P_{add} plots. We added high-order epistasis by generating random epistatic coefficients and then calculating phenotypes using Eq. 3. We introduced nonlinearity by transforming these phenotypes with Eq. 5. For each genotype in these simulations, we calculated P_{add} as the sum of the first-order coefficients used in the generating model. P_{obs} is the observable phenotype, including both high-order epistasis and nonlinear scale.

High-order epistasis and nonlinear scale had qualitatively different effects on P_{obs} vs. P_{add} plots. Fig 3A shows plots of P_{obs} vs. P_{add} for increasing nonlinearity (left-to-right) and high-order epistasis (bottom-to-top). As nonlinearity increases, P_{obs} curves systematically relative to P_{add}. This reflects the fact that P_{add} is on a linear scale and P_{obs} is on a saturating, nonlinear scale. The shape of the curve reflects the map between the linear and saturating scale: the smallest phenotypes are

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

underestimated and the largest phenotypes overestimated. In contrast, high-order epistasis induces random scatter away from the $P_{obs} = P_{add}$ line. This is because the epistatic coefficients used to generate the map are specific to each genotype, moving observations off the expected line, even if the scaling relationship is taken into account.

Nonlinearity can be separated from underlying high-order epistasis

The P_{obs} vs. P_{add} plots suggest an approach to disentangle high-order epistasis from nonlinear scale. By fitting a function to the P_{obs} vs P_{add} curve, we describe a transformation that relates the linear P_{add} scale to the (possibly nonlinear) P_{obs} scale (Schenk et al. 2013; Szendro et al. 2013). Once the form of the nonlinearity is known, we can then linearize the phenotypes so they are on an appropriate scale for epistatic analysis. Variation that remains (i.e. scatter) can then be confidently partitioned into epistatic coefficients.

In the absence of knowledge about the source of the nonlinearity, a natural choice is a power transform (Box and Cox 1964; Carroll and Ruppert 1981), which identifies a monotonic, continuous function through P_{obs} vs. P_{add}. A key feature of this approach is that power-transformed data are normally distributed around the fit curve and thus appropriately scaled for regression of a linear epistasis model.

We tested this approach using one of our simulated data sets. One complication is that, for an experimental map, we do not know P_{add}. In the analysis above, we determined P_{add} from the additive coefficients used to generate the space. In a real map, P_{add} is not known; therefore, we had to estimate P_{add}. We did so by measuring the average effect of each mutation across all backgrounds, and then calculating $\hat{P}_{add}$ for each genotype as the sum of these average effects (Eq. 1).

We fit the power transform to P_{obs} vs. $\hat{P}_{add}$ (solid red line, Fig 3B). The curve captures the nonlinearity added in the simulation. We linearized P_{obs} using the fit model (Eq. 2), and then extracted high-order epistatic coefficients. The extracted coefficients were highly correlated with the coefficients used to generate the map ($R^2 = 0.998$) (Fig 3C). In contrast, applying the linear epistasis model to this map without first accounting for nonlinearity gives much greater scatter

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

between the input and output coefficients ($R^2 = 0.934$) (Fig 3D). This occurs because phenotypic variation from nonlinearity is incorrectly partitioned into the linear epistatic coefficients.

Nonlinearity is a common feature of genotype-phenotype maps

Our next question was whether experimental maps exhibited nonlinear scales. We selected seven genotype-phenotype maps that had previously been reported to exhibit high-order epistasis (Table 1) and fit power transforms to each dataset (Fig 4, S1). We expected some phenotypes to be multiplicative (e.g. datasets I, II and IV were relative fitness), while we expected some to be additive (e.g. dataset IV is a free energy). Rather than rescaling the multiplicative datasets by taking logarithms of the phenotypes, we allowed our power transform to capture the appropriate scale. The power-transform identified nonlinearity in the majority of data sets. Of the seven data sets, three were less-than-additive (II, V, VI), two were greater-than-additive (III, IV), and two were approximately linear (I, VII). All data sets gave random residuals after fitting the power transform (Fig 4, S1).

High-order epistasis is a common feature of genotype-phenotype maps

With estimated scales in hand, we linearized the maps using Eq. 2 and re-measured epistasis (Fig S2). We used bootstrap sampling of uncertainty in the measured phenotypes to determine the uncertainty of each epistatic coefficient (see Methods), and then integrated these distributions to determine whether each coefficient was significantly different than zero. We then applied a Bonferroni correction to each p-value to account for multiple testing.

Despite our conservative statistical approach, we found high-order epistasis in every map studied (Fig 5A, S3). Every data set exhibited at least one statistically significant epistatic coefficient of fourth order or higher. We even detected statistically significant fifth-order epistasis (blue bar in Fig 5A, data set II). High-order coefficients were both positive and negative, often with magnitudes equal to or greater than the second-order terms. These results reveal that high-order epistasis is a robust feature of these maps, even when nonlinearity and measurement uncertainty in the genotype-phenotype map is taken into account.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We also dissected the relative contributions of each epistatic order to the remaining variation. To do so, we created truncated epistasis models: an additive model, a model containing additive and pairwise terms, a model containing additive through third-order terms, etc. We then measured how well each model accounted for variation in the phenotype using a Pearson's coefficient between the fit and the data. Finally, we asked how much the Pearson coefficient changed with addition of more epistatic coefficients. For example, to measure the contribution of pairwise epistasis, we took the difference in the correlation coefficient between the additive plus pairwise model and the purely additive model.

The contribution of epistasis to the maps was highly variable (Fig 5B, S3). For data set I, epistatic terms explained 5.9% of the variation in the data. The contributions of epistatic coefficients decayed with increasing order, with fifth-order epistasis only explaining 0.1% of the variation in the data. In contrast, for data set II, epistasis explains 43.3% of the variation in the map. Fifth-order epistasis accounts for 6.3% of the variation in the map. The other data sets had epistatic contributions somewhere between these extremes.

Accounting for nonlinear genotype-phenotype maps alters epistatic coefficients

Finally, we probed to what extent accounting for nonlinearity in phenotype altered the epistatic coefficients extracted from each space. Fig 6 and S4 show correlation plots between epistatic coefficients extracted both with and without linearization. The first-order coefficients were all highly correlated between the linear and nonlinear analyses for all data sets (Fig S5).

For the epistatic coefficients, the degree of correlation depended on the degree of nonlinearity in the dataset. Data set I—which was essentially linear—had identical epistatic coefficients whether the nonlinear scale was taken into account or not. In contrast, the other data sets exhibited scatter off of the line. Data set III was particularly noteworthy. The epistatic coefficients were systematically overestimated when the nonlinear scale was ignored. Two large and favorable pairwise epistatic terms in the linear analysis became essentially zero when nonlinearity was taken into account. These interactions—M182T/g4205a and G283S/g4205a—were both noted as determinants

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

of evolutionary trajectories in the original publication (Weinreich et al. 2006); however, our results suggest the interaction is an artifact of applying a linear model to a nonlinear data set. Further $\approx 20\%$ (six of 27) epistatic coefficients flipped sign when nonlinearity was taken into account (Fig 6, III, bottom right quadrant).

Overall, we found that low-order epistatic coefficients were more robust to the linear assumption than high-order coefficients. Data set IV is a clear example of this behavior. The map exhibited noticeable nonlinearity (Fig 4). The first- and second-order terms were well correlated between the linear and nonlinear analyses (Fig 6, S4, S5). Higher-order terms, however, exhibited much poorer overall correlation. While the R^2 for second-order coefficients was 0.95, the correlation was only 0.43 for third-order. This suggests that previous analyses of nonlinear genotype-phenotype maps correctly identified the key mutations responsible for variation in the map, but incorrectly estimated the high-order epistatic effects.

Discussion

Our results reveal that both nonlinear scales and high-order epistasis play important roles in shaping experimental genotype-phenotype maps. Five of the seven data sets we investigated exhibited nonlinear scales, and all of the data sets exhibited high-order epistasis, even after accounting for nonlinearity. This suggests that both should be taken into account in analyses of genotype-phenotype maps.

Origins of nonlinear scales

We observed two basic forms of nonlinearity in these maps: saturating, less-than-additive maps and exploding, greater-than-additive maps. Many have observed less-than-additive maps in which mutations have lower effects when introduced into more optimal backgrounds (MacLean et al. 2010; Chou et al. 2011). Such saturation has been proposed to be a key factor shaping evolutionary trajectories (MacLean et al. 2010; Chou et al. 2011; Kryazhimskiy et al. 2014; Tokuriki et al. 2012; Otto and Feldman 1997). Further, it is intuitive that optimizing a phenotype becomes more

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

difficult as that phenotype improves. Our nonlinear fits revealed this behavior in three different maps.

The greater-than-additive maps, in contrast, were more surprising: why would mutations have a larger effect when introduced into a more favorable background? For the β-lactamase genotype-phenotype map (III, Fig 4), it appears this is an artifact of the original analysis used to generate the data set (Weinreich et al. 2006). This data set describes the fitness of bacteria expressing variants of an enzyme with activity against β-lactam antibiotics. The original authors measured the minimum-inhibitory concentration (MIC) of the antibiotic against bacteria expressing each enzyme variant. They then converted their MIC values into apparent fitness by sampling from an exponential distribution of fitness values and assigning these fitness values to rank-ordered MIC values (Weinreich et al. 2006). Our epistasis model extracts this original exponential distribution (Fig S6). This result demonstrates the effectiveness of our approach in extracting nonlinearity in the genotype-phenotype map.

The origins of the growth in the transcription factor/DNA binding data set are less clear (IV, Fig 4). The data set measures the binding free energy of variants of a transcription factor binding to different DNA response elements. We are aware of no physical reason for mutations to have a larger effect on free energy when introduced into a background with better binding. One possibility is that the genotype-phenotype map reflects multiple features that are simultaneously altered by mutations, giving rise to this nonlinear shape. This is a distinct possibility in this data set, where mutations are known to alter both DNA binding affinity and DNA binding cooperativity (McKeown et al. 2014).

Best Practice

Because nonlinearity is a common feature of these maps, linearity should not be assumed in analyses of epistasis. Given a sufficient number of phenotypic observations, however, the appropriate scale can be estimated by construction of a P_{obs} vs. P_{add} plot and regression of a nonlinear scale model. With this scale in hand, one can then transform the genotype-phenotype map onto a linear scale appropriate for analysis using a high-order epistasis model. Our software

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

pipeline automates this process. It takes any genotype-phenotype map in a standard text format, fits for nonlinearity, and then estimates high-order epistasis. It is freely available for download (https://harmslab.github.com/epistasis).

One important question is how to select an appropriate function to describe the nonlinear scale. By visual inspection, all of the data sets we studied were monotonic in $\hat{P}_{add}$ and could be readily captured by a power transform. Other maps may be better captured with other functions. For example, inspection of a P_{obs} vs $\hat{P}_{add}$ plot could reveal a non-monotonic scale, leading to a better fit with a polynomial than a power-transform. Another possibility is that external biological knowledge motivates scale choice (Schenk et al. 2013).

The choice of model determines what fraction of the variation is assigned to "scale" versus "epistasis." The more complicated the function chosen, the more variation in the data is shifted from epistasis and into scale. One could, for example, fit a completely uninformative Lth-order polynomial, which would capture all of variation as scale and none as epistasis. Scale estimation should governed by the well-established principles of model regression: find the simplest function that captures the maximum amount of variation in the data set without fitting stochastic noise. Because epistasis is scatter off the scale line (noise), model-selection approaches like the F-test, Akaike Information Criterion, and inspection of fit residuals are a natural strategy for partitioning variation between scale and epistasis.

Interpretation

Another powerful aspect of this approach is that it allows explicit separation of two distinct origins of non-additivity in genotype-phenotype maps.

This can be illustrated with a simple, conceptual, example. Imagine mutations to an enzyme, expressed in bacteria, that have a less-than-additive effect on bacterial growth rate. To a first approximation, this epistasis could have two origins. The first is at the level of the enzyme: maybe the mutations have a specific, negative chemical interactions that alter enzyme rate. The second is at the level of the whole cell: maybe, above a certain activity, the enzyme is fast enough that some other part of the cell starts limiting growth. Mutations continue to improve enzyme activity, but

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

growth rate does not reflect this. These two origins of less-than-additive behavior will have different effects in a P_{add} vs. P_{obs} plot: saturation of growth rate will appear as nonlinearity, interactions between mutations at the enzyme level will appear as linear epistasis. Our analysis would reveal this pattern and set up further experiments to tease apart these possibilities.

This may also provide important evolutionary insights. One important question is to what extent evolutionary paths are shaped by global constraints versus specific interactions that lead to specific historical contingencies (Harms and Thornton 2014; Shah et al. 2015; Kryazhimskiy et al. 2014). For example, recent work has shown specific epistatic interactions lead to sequence-level unpredictability, while a globally less-than-additive scale leads to predictable phenotypes in evolution (Kryazhimskiy et al. 2014). Our analysis approach naturally distinguishes these origins of non-additivity, and thus these evolutionary possibilities. Prevailing magnitude epistasis (de Visser et al. 2009), global epistasis (Kryazhimskiy et al. 2014), and diminishing-returns epistasis (Chou et al. 2011; MacLean et al. 2010; Otto and Feldman 1997; Tokuriki et al. 2012) will all appear as nonlinear scales. In contrast, specific interactions will appear in specific coefficients in the linear epistasis model. Our detection of nonlinearity and high-order epistasis in most datasets suggests that both forms of non-additivity will be in play over evolutionary time.

High-order epistasis

Finally, our work reveals that high-order epistasis is, indeed, a common feature of genotype-phenotype maps. Our study could be viewed as an attempt to "explain away" previously observed high-order epistasis. To do so, we both accounted for nonlinearity in the map and propagated experimental uncertainty to the epistatic coefficients. Surprisingly—to the authors, at least—high-order epistasis was robust to these corrections.

High-order epistasis can make huge contributions to genotype-phenotype maps. In data set II, third-order and higher epistasis accounts for fully 31.0% of the variation in the map. The average contribution, across maps, is 12.7%. We also do not see a consistent decay in the contribution of epistasis with increasing order. In data sets II, V and VI, third-order epistasis contributes more variation to the map than second-order epistasis. This suggests that epistasis could go to even

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

higher orders in larger genotype-phenotype maps.

The generality of these results across all genotype-phenotype maps is unclear. The maps we analyzed were measured and published because they were "interesting," either from a mechanistic or evolutionary perspective. Further, most of the maps have a single, maximum phenotype peak. The nonlinearity and high-order epistasis we observed may be common for collections of mutations that, together, optimize a function, but less common in "flatter" or more random genotype-phenotype maps. This can only be determined by characterization of genotype-phenotype maps with different structural features.

The observation of this epistasis also raises important questions: What are the origins of third, fourth, and even fifth-order correlations in these data sets? What, mechanistically, leads to a five-way interaction between mutations? Does neglecting high-order epistasis bias estimates of low-order epistasis (Otwinowski and Plotkin 2014)? What can this epistasis tell us about the biological underpinning of these maps (Lehár et al. 2008; Hu et al. 2011, 2013; Taylor and Ehrenreich 2015)?

The evolutionary implications are also potentially fascinating. Epistasis creates temporal dependency between mutations: the effect of a mutation depends strongly on specific mutations that fixed earlier in time (Bedau and Packard 2003; Desai 2009; Harms and Thornton 2014; Shah et al. 2015) How does this play out for high-order epistasis, which introduces long-range correlations across genotype-phenotype maps? Do these low magnitude interactions matter for evolutionary outcomes or dynamics? These, and questions like them, are challenging and fascinating future avenues for further research.

Acknowledgments

We would like to thank Patrick Phillips, Jamie Bridgham and members of the Harms lab for helpful discussions and comments. We would also like to thank David Hall for providing the complete data for data sets VI and VII. Work was supported by start up funds from the University of Oregon (ZRS). MJH is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

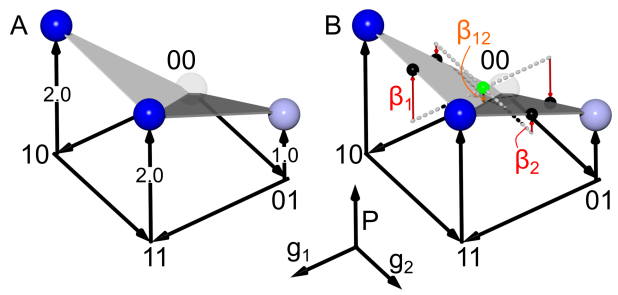

Fig 1: Epistasis can be quantified using Walsh polynomials. A) A genotype-phenotype map exhibiting negative epistasis. Axes are genotype at position 1 (g_1), genotype at position 2 (g_2), and phenotype (P). For genotypic axes, "0" denotes wildtype and "1" denotes a mutant. Phenotype is encoded both on the P-axis and as a spectrum from white to blue. The map exhibits negative epistasis: relative to wildtype, the effect of the mutations together ($P_{11} = 2$) is less than the sum of the individual effects of mutations ($P_{10} + P_{01} = 1 + 2 = 3$). B) The map can be decomposed into epistatic coefficients using a Walsh polynomial, which measures the effects of each mutation relative to the geometric center of the genotype-phenotype map (green sphere). The additive coefficients β_1 and β_2 (red arrows) are the average effect of each mutation in all backgrounds. The epistatic coefficient β_{12} (orange arrow) the variation not accounted for by β_1 and β_2. Geometrically, it is the distance between the center of the map and the "fold" given by vector connecting P_{00} and P_{11}.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

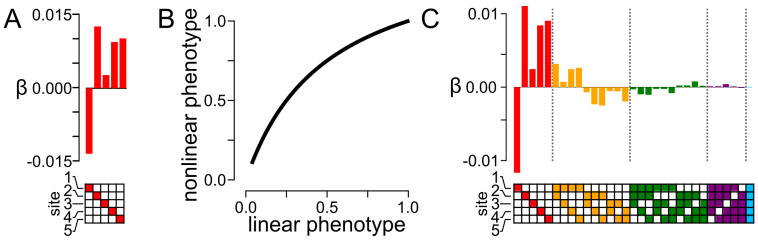

Fig 2: Nonlinearity in phenotype creates spurious high-order epistatic coefficients. A) Simulated, random, first-order epistatic coefficients. The mutated site is indicated by panel below the bar graph; bar indicates magnitude and sign of the additive coefficient. B) A nonlinear map between a linear phenotype and a saturating, nonlinear phenotype. The first-order coefficients in panel A are used to generate a linear phenotype, which is then transformed by the function shown in B. C) Epistatic coefficients extracted from the genotype-phenotype map generated in panels A and B. Bars denote coefficient magnitude and sign. Color denotes the order of the coefficient: first (β_i, red), second (β_{ij}, orange), third (β_{ijk}, green), fourth (β_{ijkl}, purple), and fifth (β_{ijklm}, blue). Filled squares in the grid below the bars indicate the identity of mutations that contribute to the coefficient.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

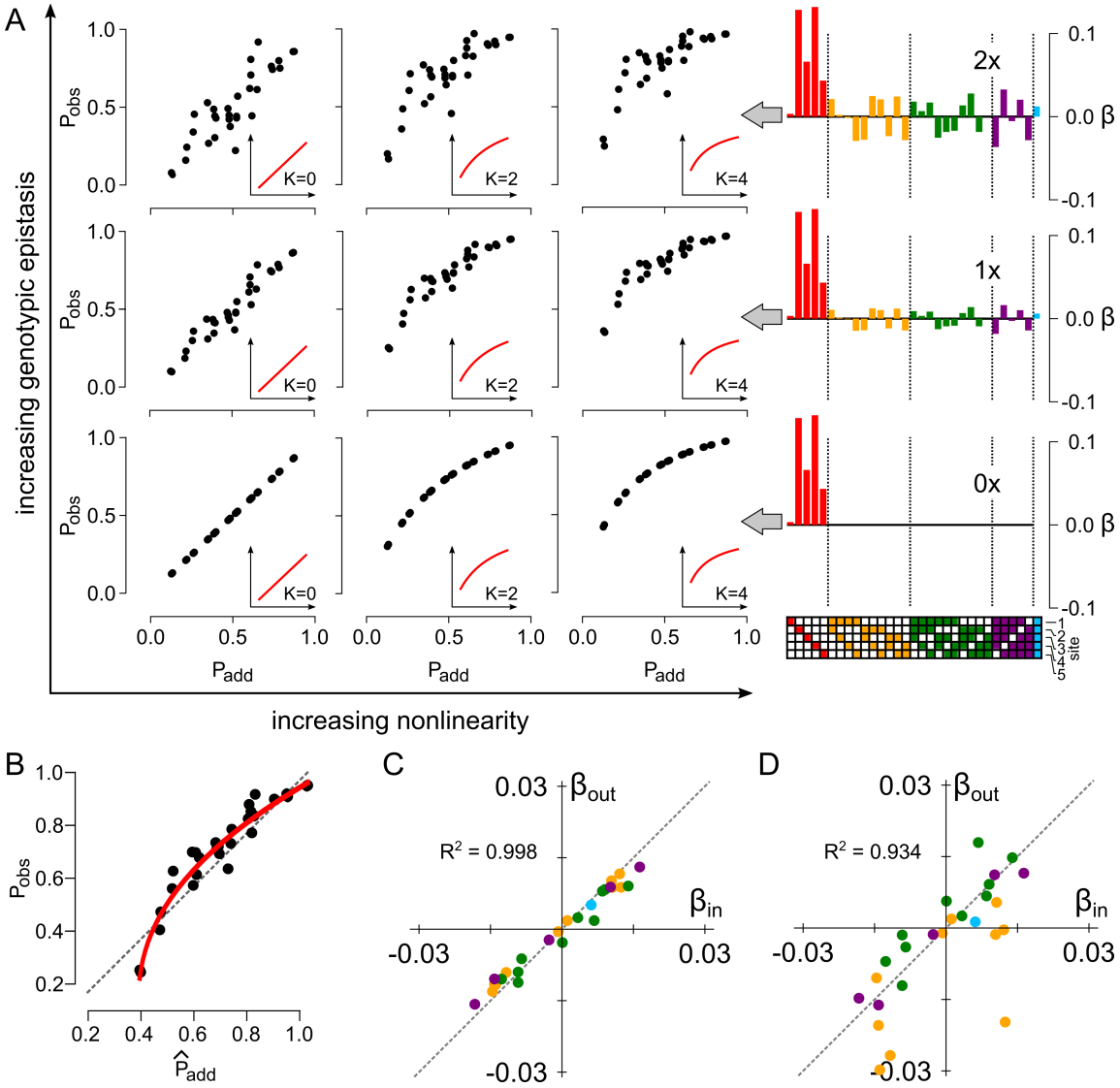

Fig 3: Epistasis and nonlinear scale induce different patterns of nonadditivity. A) Patterns of nonadditivity for increasing epistasis and nonlinear scale. Main panel shows a grid ranging from no epistasis, linear scale (bottom-left) to high epistasis, highly nonlinear scale (top-right). Insets in sub-panels show added nonlinearity. Going from left to right: $K = 0$, $K = 2$, $K = 4$. Epistatic coefficient plots to right show the magnitude of the input high-order epistasis, with colors and annotation as in Fig 2C. B) Plot of P_{obs} against $\hat{P}_{add}$ for the middle sub panel in

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

panel A. Red line is the fit of the power transform to these data. C) Correlation between epistatic coefficients input into the simulation and extracted from the simulation after linearization by the power transform. Each point is an epistatic coefficient, colored by order. The Pearson's correlation coefficient is shown in the upper-left quadrant. D) Correlation between epistatic coefficients input into the simulation and extracted from the simulation without application of the power transform.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

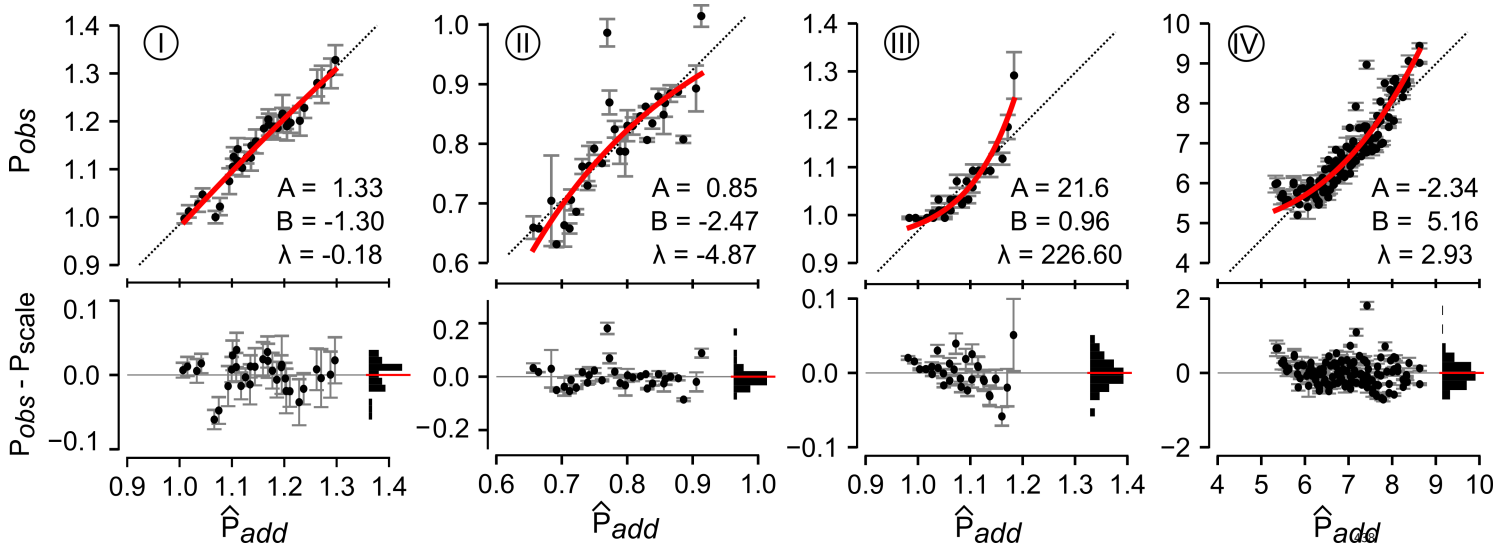

Fig 4: Experimental genotype-phenotype maps exhibit nonlinear phenotypes. Plots show observed phenotype P_{obs} plotted against $\hat{P}_{add}$ (Eq. 1) for data sets I through IV. Points are individual genotypes. Error bars are experimental standard deviations in phenotype. Red lines are the fit of the power transform to the data set. Pearson's coefficient for each fit are shown on each plot. Dashed lines are $P_{add} = P_{obs}$. Bottom panels in each plot show residuals between the observed phenotypes and the red fit line. Points are the individual residuals. Errorbars are the experimental standard deviation of the phenotype. The horizontal histograms show the distribution of residuals across 10 bins. The red lines are the mean of the residuals.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

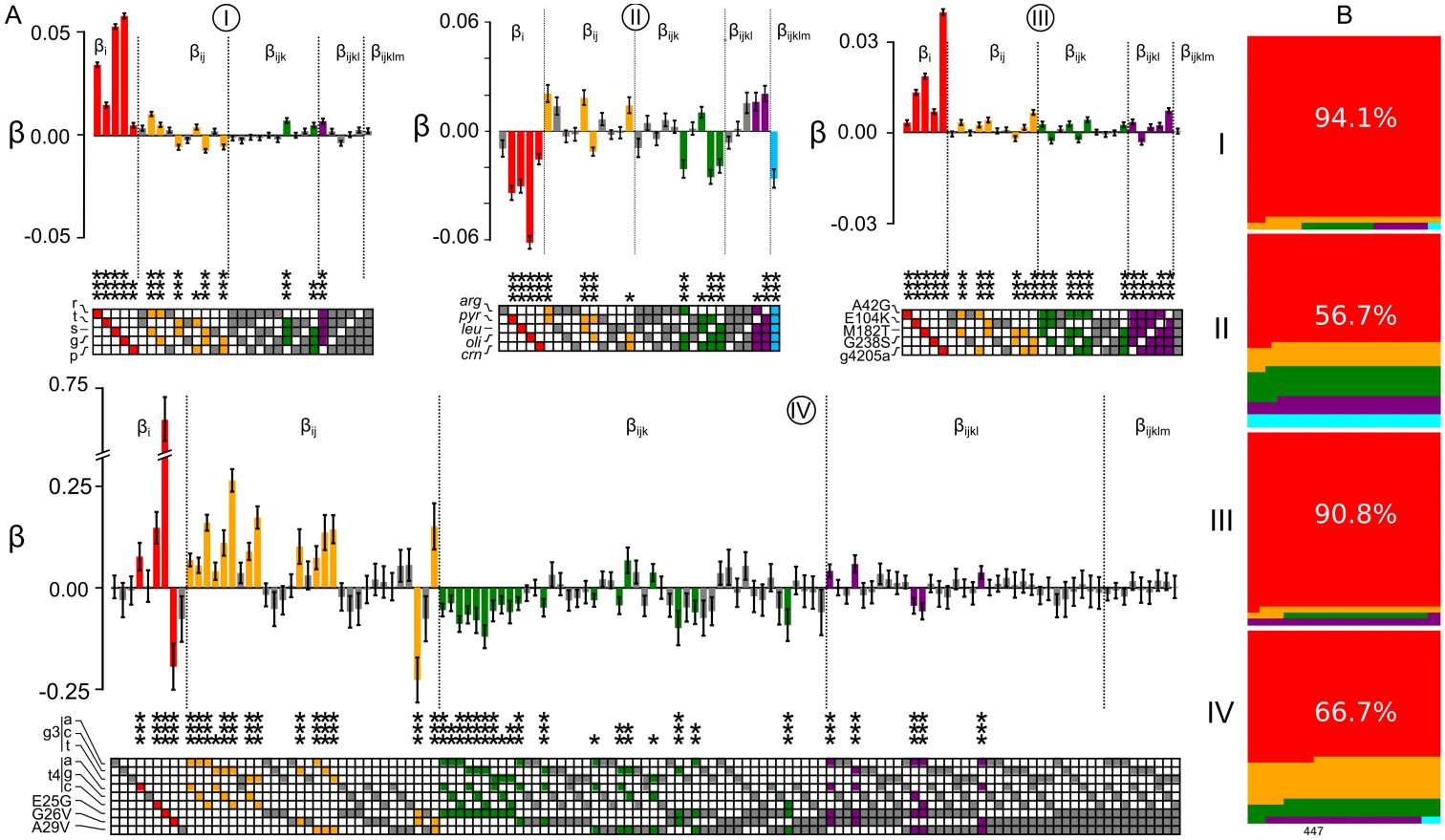

Fig 5: High-order epistasis is present in genotype-phenotype maps. A) Panels show epistatic coefficients extracted from data sets I-IV (Table 1, data set label circled above each graph). Bars denote coefficient magnitude and sign; error bars are propagated measurement uncertainty. Color denotes the order of the coefficient: first (β_i, red), second (β_{ij}, orange), third (β_{ijk}, green), fourth (β_{ijkl}, purple), and fifth (β_{ijklm}, blue). Bars are colored if the coefficient is significantly different than zero (Z-score with p-value < 0.05 after Bonferroni correction for multiple testing). Stars denote relative significance: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***). Filled squares in the grid below the bars indicate the identity of mutations that contribute to the coefficient. The names of the mutations, taken from the original publications, are indicated to the left of the grid squares. B) Sub-panels show fraction of variation accounted for by first through fifth order epistatic coefficients for data sets I-IV (colors as in panel A). Fraction described by each order is proportional

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to area.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

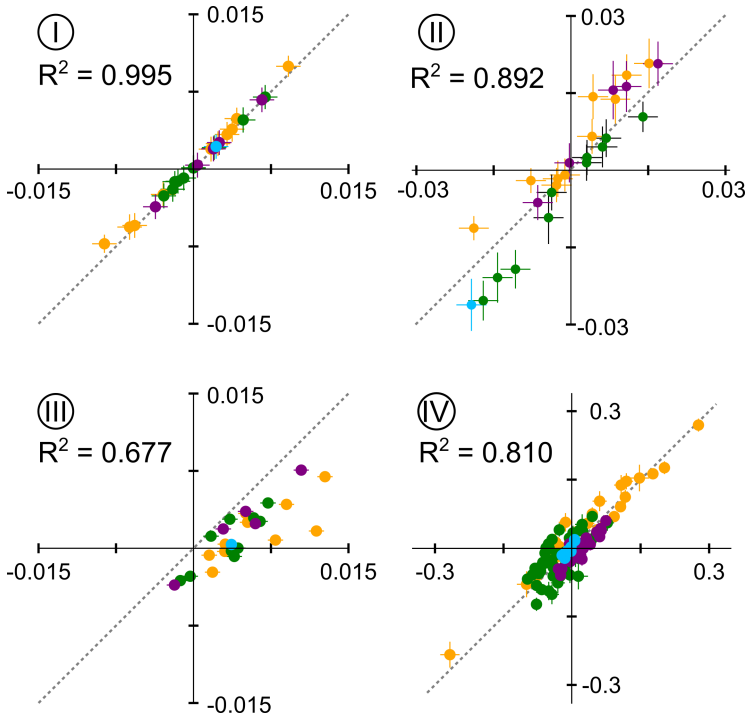

Fig 6: Nonlinear phenotypes distort measured epistatic coefficients. Sub-panels show correlation plots between epistatic coefficients extracted without accounting for nonlinearity (x-axis) and accounting for linearity (y-axis) for data sets I-IV. Each point is an epistatic coefficient, colored by order. Error bars are standard deviations from bootstrap replicates of each fitting approach.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1

ID	genotype	phenotype	L	reference
I	scattered genomic mutations	*E. coli* fitness	5	(Khan et al. 2011)
II	chromosomes in asexual fungi	*A. niger* fitness	5	(de Visser et al. 2009)
III	protein point mutants	bacterial fitness	5	(Weinreich et al. 2006)
IV	DNA/protein point mutants	*in vitro* DNA/protein binding affinity	5	(Anderson et al. 2015)
V	chromosomes in asexual fungi	*A. niger* fitness	5	(de Visser et al. 2009)
VI	alleles in biosynthetic network	*S. cerevisiae* haploid growth rate	6	(Hall et al. 2010)
VII	alleles in biosynthetic network	*S. cerevisiae* diploid growth rate	6	(Hall et al. 2010)

All data sets have 2^L genotypes except the DNA/protein interaction data set (IV), which has 128 genotypes. This occurs because the data set has 2 DNA sites (each of which have 4 possible bases) and 3 protein sites (each of which has two possible amino acids).

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Herve Abdi. The Bonferonni and sidak corrections for multiple comparisons. *Encyclopedia of measurement and statistics*, 3:103–107, 2007.

Dave W. Anderson, Alesia N. McKeown, and Joseph W. Thornton. Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. *eLife Sciences*, page e07864, June 2015. ISSN 2050-084X. doi: 10.7554/eLife.07864.

Mark A Bedau and Norman H Packard. Evolution of evolvability via adaptation of mutation rates. *Biosystems*, 69(2–3):143–162, May 2003. ISSN 0303-2647. doi: 10.1016/S0303-2647(02)00137-5.

François Blanquart. Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution*, 68(12):3537–54, 2014. ISSN 00143820. doi: 10.1111/evo.12545.

G. E. P. Box and D. R. Cox. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252, 1964. ISSN 0035-9246.

R. J. Carroll and David Ruppert. On prediction and the power transformation family. *Biometrika*, 68(3):609–615, January 1981. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/68.3.609.

Hsin-Hung Chou, Hsuan-Chao Chiu, Nigel F. Delaney, Daniel Segrè, and Christopher J. Marx. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science*, 332(6034):1190–1192, March 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1203799.

Heather J. Cordell. Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.*, 11(20):2463–2468, January 2002. ISSN 0964-6906, 1460-2083. doi: 10.1093/hmg/11.20.2463.

Jack da Silva, Mia Coetzer, Rebecca Nedellec, Cristina Pastore, and Donald E. Mosier. Fitness Epistasis and Constraints on Adaptation in a Human Immunodeficiency Virus Type 1 Protein Region. *Genetics*, 185(1):293–303, May 2010. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.109.112458.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

J. Arjan G. M. de Visser, Su-Chan Park, and Joachim Krug. Exploring the Effect of Sex on Empirical Fitness Landscapes. *The American Naturalist*, 174(s1):S15–S30, July 2009. ISSN 0003-0147, 1537-5323. doi: 10.1086/599081.

Michael M. Desai. Reverse evolution and evolutionary memory. *Nat Genet*, 41(2):142–143, February 2009. ISSN 1061-4036. doi: 10.1038/ng0209-142.

R.A. Fisher. The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Philosophical Transactions of the Royal Society of Edinburgh*, (52):399–433, 1918.

Wayne N. Frankel and Nicholas J. Schork. Who's afraid of epistasis? *Nat Genet*, 14(4):371–373, December 1996. doi: 10.1038/ng1296-371.

David W. Hall, Matthew Agan, and Sara C. Pope. Fitness Epistasis among 6 Biosynthetic Loci in the Budding Yeast Saccharomyces cerevisiae. *J Hered*, 101(suppl 1):S75–S84, January 2010. ISSN 0022-1503, 1465-7333. doi: 10.1093/jhered/esq007.

Michael J. Harms and Joseph W. Thornton. Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature*, 512(7513):203–207, August 2014. ISSN 0028-0836. doi: 10.1038/nature13410.

Robert B. Heckendorn and Darrell Whitley. Predicting Epistasis from Mathematical Models. *Evolutionary Computation*, 7(1):69–101, March 1999. ISSN 1063-6560. doi: 10.1162/evco.1999.7.1.69.

Ting Hu, Nicholas A. Sinnott-Armstrong, Jeff W. Kiralis, Angeline S. Andrew, Margaret R. Karagas, and Jason H. Moore. Characterizing genetic interactions in human disease association studies using statistical epistasis networks. *BMC Bioinformatics*, 12:364, 2011. ISSN 1471-2105. doi: 10.1186/1471-2105-12-364.

Ting Hu, Yuanzhu Chen, Jeff W. Kiralis, Ryan L. Collins, Christian Wejse, Giorgio Sirugo, Scott M. Williams, and Jason H. Moore. An information-gain approach to detecting three-way epistatic interactions in genetic association studies. *J Am Med Inform Assoc*, pages amiajnl–2012–001525, February 2013. ISSN , 1527-974X. doi: 10.1136/amiajnl-2012-001525.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

J. D. Hunter. Matplotlib: A 2D Graphics Environment. *Computing in Science Engineering*, 9(3): 90–95, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.55.

Marcin Imielinski and Calin Belta. Exploiting the pathway structure of metabolism to reveal high-order epistasis. *BMC Systems Biology*, 2:40, 2008. ISSN 1752-0509. doi: 10.1186/1752-0509-2-40.

Aisha I. Khan, Duy M. Dinh, Dominique Schneider, Richard E. Lenski, and Tim F. Cooper. Negative Epistasis Between Beneficial Mutations in an Evolving Bacterial Population. *Science*, 332 (6034):1193–1196, March 2011. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1203801.

Sergey Kryazhimskiy, Daniel P. Rice, Elizabeth R. Jerison, and Michael M. Desai. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science*, 344(6191):1519–1522, June 2014. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1250939.

Joseph Lehár, Andrew Krueger, Grant Zimmermann, and Alexis Borisy. High-order combination effects and biological robustness. *Molecular Systems Biology*, 4(1):215, January 2008. ISSN 1744-4292, 1744-4292. doi: 10.1038/msb.2008.51.

R. C. MacLean, G. G. Perron, and A. Gardner. Diminishing Returns From Beneficial Mutations and Pervasive Epistasis Shape the Fitness Landscape for Rifampicin Resistance in Pseudomonas aeruginosa. *Genetics*, 186(4):1345–1354, December 2010. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.110.123083.

Tomoaki Matsuura, Yasuaki Kazuta, Takuyo Aita, Jiro Adachi, and Tetsuya Yomo. Quantifying epistatic interactions among the components constituting the protein translation system. *Molecular Systems Biology*, 5(1):297, January 2009. ISSN 1744-4292, 1744-4292. doi: 10.1038/msb.2009.50.

Alesia N. McKeown, Jamie T. Bridgham, Dave W. Anderson, Michael N. Murphy, Eric A. Ortlund, and Joseph W. Thornton. Evolution of DNA Specificity in a Transcription Factor Family Produced a New Gene Regulatory Module. *Cell*, 159(1):58–68, September 2014. ISSN 0092-8674. doi: 10.1016/j.cell.2014.09.003.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Johannes Neidhart, Ivan G. Szendro, and Joachim Krug. Exact results for amplitude spectra of fitness landscapes. *Journal of Theoretical Biology*, 332:218–227, September 2013. ISSN 0022-5193. doi: 10.1016/j.jtbi.2013.05.002.

Sarah Perin Otto and Marcus W. Feldman. Deleterious Mutations, Variable Epistatic Interactions, and the Evolution of Recombination. *Theoretical Population Biology*, 51(2):134–147, April 1997. ISSN 0040-5809. doi: 10.1006/tpbi.1997.1301.

Jakub Otwinowski and Joshua B. Plotkin. Inferring fitness landscapes by regression produces biased estimates of epistasis. *PNAS*, 111(22):E2301–E2309, March 2014. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1400849111.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

F. Perez and B. E. Granger. IPython: A System for Interactive Scientific Computing. *Computing in Science Engineering*, 9(3):21–29, May 2007. ISSN 1521-9615. doi: 10.1109/MCSE.2007.53.

Mats Pettersson, Francois Besnier, Paul B. Siegel, and Örjan Carlborg. Replication and Explorations of High-Order Epistasis Using a Large Advanced Intercross Line Pedigree. *PLoS Genet*, 7(7):e1002180, July 2011. doi: 10.1371/journal.pgen.1002180.

Patrick C. Phillips. Epistasis —the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet*, 9(11):855–867, November 2008. ISSN 1471-0056. doi: 10.1038/nrg2452.

Frank J. Poelwijk, Vinod Krishna, and Rama Ranganathan. The Context-Dependence of Mutations: A Linkage of Formalisms. *PLOS Computational Biology*, 12(6):e1004771, June 2016. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1004771.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Marylyn D. Ritchie, Lance W. Hahn, Nady Roodi, L. Renee Bailey, William D. Dupont, Fritz F. Parl, and Jason H. Moore. Multifactor-Dimensionality Reduction Reveals High-Order Interactions among Estrogen-Metabolism Genes in Sporadic Breast Cancer. *The American Journal of Human Genetics*, 69(1):138–147, July 2001. ISSN 0002-9297. doi: 10.1086/321276.

Darin R. Rokyta, Paul Joyce, S. Brian Caudle, Craig Miller, Craig J. Beisel, and Holly A. Wichman. Epistasis between Beneficial Mutations and the Phenotype-to-Fitness Map for a ssDNA Virus. *PLOS Genet*, 7(6):e1002075, June 2011. ISSN 1553-7404. doi: 10.1371/journal.pgen.1002075.

Kenneth J. Rothman, Sander Greenland, and Alexander M. Walker. Concepts of Interaction. *Am. J. Epidemiol.*, 112(4):467–470, January 1980. ISSN 0002-9262, 1476-6256.

Martijn F. Schenk, Ivan G. Szendro, Merijn L. M. Salverda, Joachim Krug, and J. Arjan G. M. de Visser. Patterns of Epistasis between Beneficial Mutations in an Antibiotic Resistance Gene. *Mol Biol Evol*, 30(8):1779–1787, January 2013. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/mst096.

Daniel Segrè, Alexander DeLuna, George M. Church, and Roy Kishony. Modular epistasis in yeast metabolism. *Nat Genet*, 37(1):77–83, January 2005. ISSN 1061-4036. doi: 10.1038/ng1489.

Premal Shah, David M. McCandlish, and Joshua B. Plotkin. Contingency and entrenchment in protein evolution under purifying selection. *PNAS*, page 201412933, June 2015. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1412933112.

Jiya Sun, Fuhai Song, Jiajia Wang, Guangchun Han, Zhouxian Bai, Bin Xie, Xuemei Feng, Jianping Jia, Yong Duan, and Hongxing Lei. Hidden risk genes with high-order intragenic epistasis in Alzheimer's disease. *J. Alzheimers Dis.*, 41(4):1039–1056, 2014. ISSN 1875-8908. doi: 10.3233/JAD-140054.

Ivan G. Szendro, Martijn F. Schenk, Jasper Franke, Joachim Krug, and J. Arjan G. M. de Visser. Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech.*, 2013(01):P01005, January 2013. ISSN 1742-5468. doi: 10.1088/1742-5468/2013/01/P01005.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Matthew B. Taylor and Ian M. Ehrenreich. Higher-order genetic interactions and their contribution to complex traits. *Trends in Genetics*, 31(1):34–40, January 2015. ISSN 0168-9525. doi: 10.1016/j.tig.2014.09.001.

Nobuhiko Tokuriki, Colin J. Jackson, Livnat Afriat-Jurnou, Kirsten T. Wyganowski, Renmei Tang, and Dan S. Tawfik. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun*, 3:1257, December 2012. doi: 10.1038/ncomms2246.

Chia-Ti Tsai, Juey-Jen Hwang, Marylyn D. Ritchie, Jason H. Moore, Fu-Tien Chiang, Ling-Ping Lai, Kuan-Lih Hsu, Chuen-Den Tseng, Jiunn-Lee Lin, and Yung-Zu Tseng. Renin–angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort: Detection of high order gene–gene interaction. *Atherosclerosis*, 195(1):172–180, November 2007. ISSN 0021-9150. doi: 10.1016/j.atherosclerosis.2006.09.014.

S. van der Walt, S. C. Colbert, and G. Varoquaux. The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science Engineering*, 13(2):22–30, March 2011. ISSN 1521-9615. doi: 10.1109/MCSE.2011.37.

Yinhua Wang, Carolina Díaz Arenas, Daniel M. Stoebel, and Tim F. Cooper. Genetic background affects epistatic interactions between two beneficial mutations. *Biology Letters*, page rsbl20120328, August 2012. ISSN 1744-9561, 1744-957X. doi: 10.1098/rsbl.2012.0328.

Daniel M. Weinreich, Nigel F. Delaney, Mark A. DePristo, and Daniel L. Hartl. Darwinian Evolution Can Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, 312(5770):111–114, July 2006. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1123539.

Daniel M Weinreich, Yinghong Lan, C Scott Wylie, and Robert B. Heckendorn. Should evolutionary geneticists worry about higher-order epistasis? *Current Opinion in Genetics & Development*, 23 (6):700–707, December 2013. ISSN 0959-437X. doi: 10.1016/j.gde.2013.10.007.

Jianfeng Xu, James Lowey, Fredrik Wiklund, Jielin Sun, Fredrik Lindmark, Fang-Chi Hsu, Latchezar Dimitrov, Baoli Chang, Aubrey R. Turner, Wennan Liu, Hans-Olov Adami, Edward

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Suh, Jason H. Moore, S. Lilly Zheng, William B. Isaacs, Jeffrey M. Trent, and Henrik Grönberg. The Interaction of Four Genes in the Inflammation Pathway Significantly Predicts Prostate Cancer Risk. *Cancer Epidemiol Biomarkers Prev*, 14(11):2563–2568, January 2005. ISSN 1055-9965, 1538-7755. doi: 10.1158/1055-9965.EPI-05-0356.

Shozo Yokoyama, Ahmet Altun, Huiyong Jia, Hui Yang, Takashi Koyama, Davide Faggionato, Yang Liu, and William T. Starmer. Adaptive evolutionary paths from UV reception to sensing violet light by epistatic interactions. *Science Advances*, 1(8):e1500162, September 2015. ISSN 2375-2548. doi: 10.1126/sciadv.1500162.

Chun-Ting Zhang and Ren Zhang. Analysis of distribution of bases in the coding sequences by a digrammatic technique. *Nucl. Acids Res.*, 19(22):6313–6317, November 1991. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/19.22.6313.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

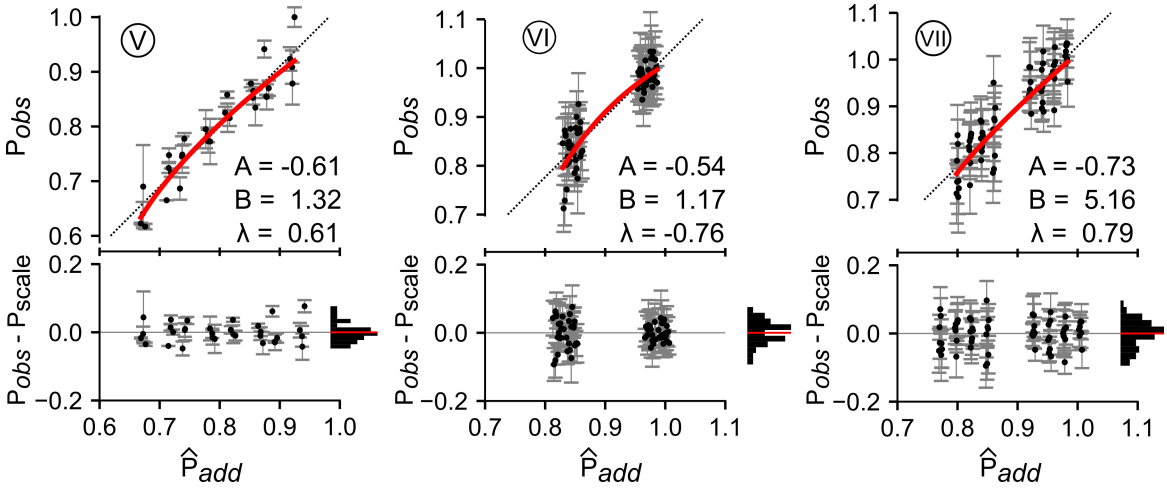

Fig S1: Experimental genotype-phenotype maps exhibit nonlinear phenotypes. Plots show observed phenotype P_{obs} plotted against $\hat{P}_{add}$ (Eq. 1) for data sets V through VII. Points are individual genotypes. Error bars are experimental standard deviations in phenotype. Red lines are the fit of the power transform to the data set. Pearson's coefficient for each fit are shown on each plot. Dashed lines are $P_{add} = P_{obs}$. Bottom panels in each plot show residuals between the observed phenotypes and the red fit line. Points are the individual residuals. Errorbars are the experimental standard deviation of the phenotype. The horizontal histograms show the distribution of residuals across 10 bins. The red lines are the mean of the residuals.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

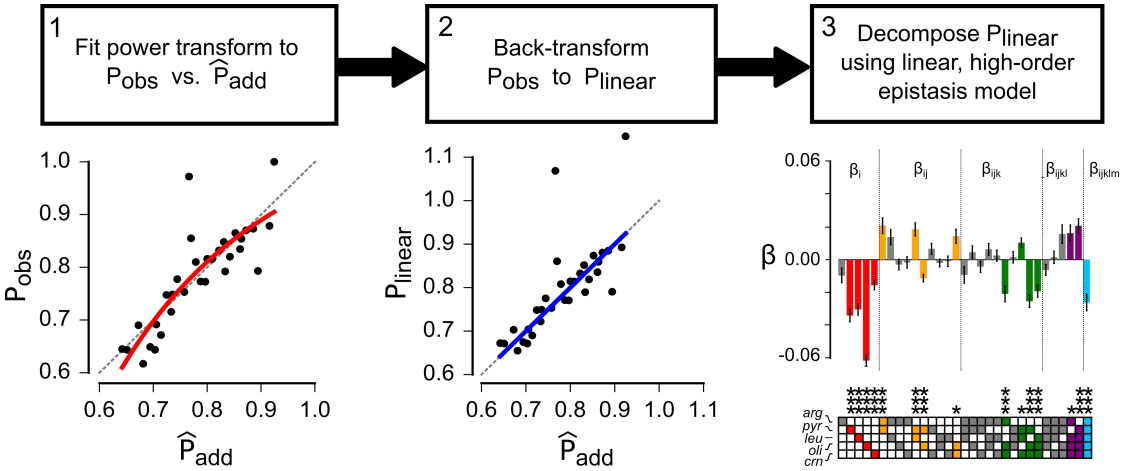

Fig S2: Nonlinear phenotypes can be transformed to linear scale to estimate high-order epistasis. Flowchart shows the steps for estimating high-order epistasis in nonlinear genotype-phenotype maps. The plots beneath the chart show this pipeline for data set II. In step 1, a power transform function is used to fit the P_{obs} versus $\hat{P}_{add}$ plot and estimate the map's scale. In step 2, the inverse of the fitted transform is used to back-transform P_{obs} to a linear scale, P_{linear}. In step 3, a linear, high-order epistasis model is used to fit the variation in P_{linear}. In the left plot, points are individual genotypes, red line is the resulting fit and dashed line is the $P_{add} = P_{obs}$. In the middle plot, the blue line is the new scale of P_{obs} after back transforming. In the right plot, bars represent additive and epistatic coefficients extracted from the linear phenotypes. Error bars are propagated measurement uncertainty. Color denotes the order of the coefficient: first (β_i, red), second (β_{ij}, orange), third (β_{ijk}, green), fourth (β_{ijkl}, purple), and fifth (β_{ijklm}, blue). Bars are colored if the coefficient is significantly different than zero (Z-score with p-value <0.05 after Bonferroni correction for multiple testing). Stars denote relative significance: p < 0.05 (*), p < 0.01 (**), p < 0.001 (***). Filled squares in the grid below the bars indicate the identity of mutations that contribute to the coefficient.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

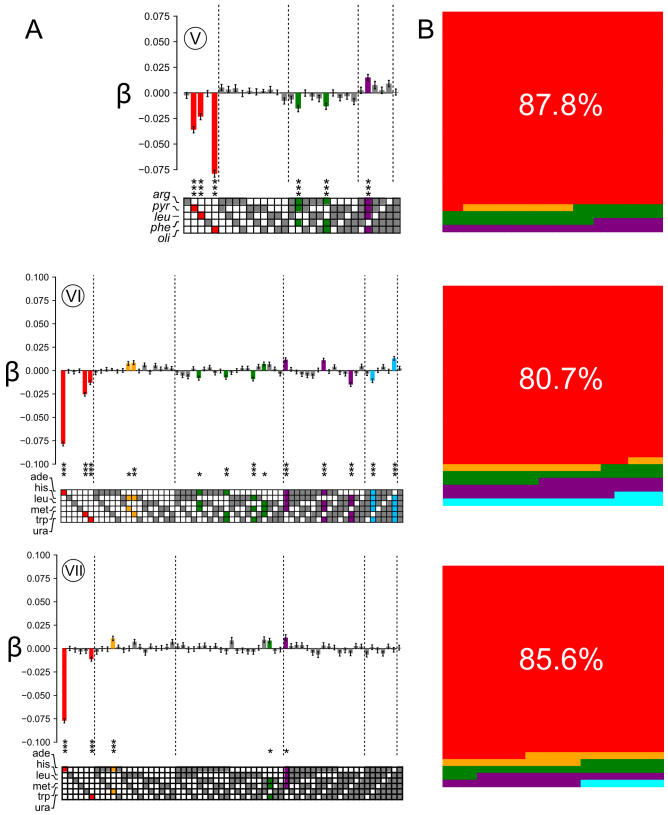

Fig S3: High-order epistasis is present in genotype-phenotype maps. A) Panels show epistatic coefficients extracted from data sets V-VII (Table 1, data set label circled above each graph). Bars denote coefficient magnitude and sign; error bars are propagated measurement uncertainty. Color denotes the order of the coefficient: first (β_i, red), second (β_{ij}, orange), third (β_{ijk}, green), fourth (β_{ijkl}, purple), and fifth (β_{ijklm}, blue). Bars are colored if the coefficient is significantly different than zero (Z-score with p-value <0.05 after Bonferroni correction for multiple testing). Stars denote relative significance: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***). Filled squares in the grid below the bars indicate the identity of mutations that contribute to the coefficient. The names of the mutations, taken from the original publications, are indicated to the left of the grid squares. B) Sub-panels show fraction of variation accounted for by first through fifth order epistatic coefficients for data sets I-IV (colors as in panel A). Fraction described by each order is proportional to area.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

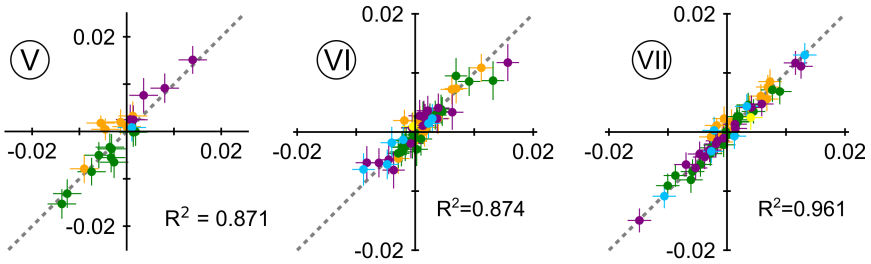

Fig S4: Nonlinear phenotypes distort measured epistatic coefficients. Sub-panels show correlation plots between epistatic coefficients extracted without accounting for nonlinearity (x-axis) and accounting for linearity (y -axis) for data sets V-VII. Each point is an epistatic coefficient, colored by order.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

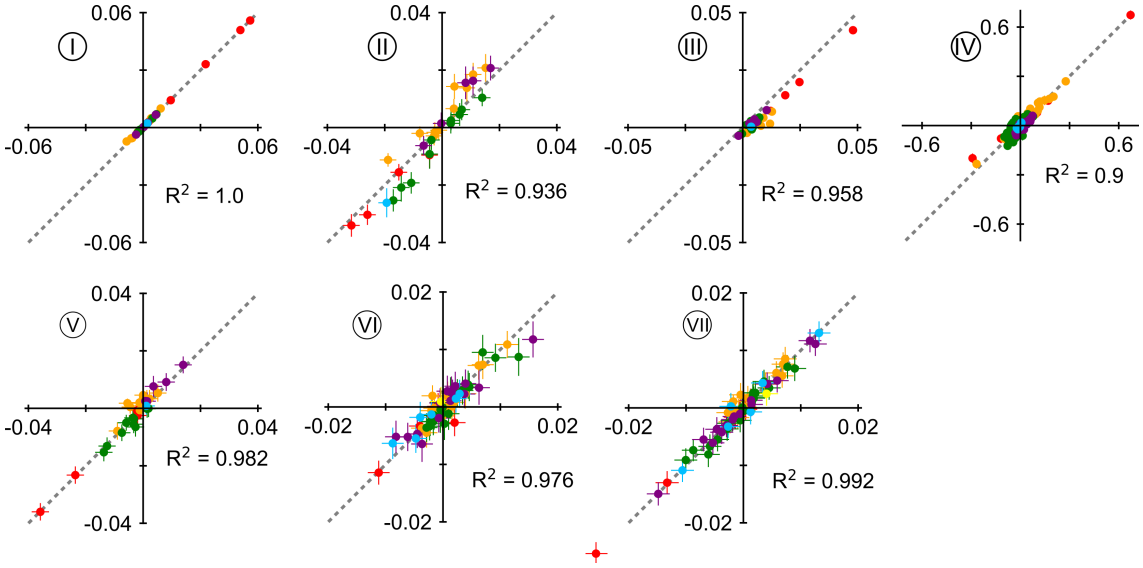

Fig S5: Additive coefficients are well estimated, even when nonlinearity is neglected. Sub-panels show correlation plots between both additive and epistatic coefficients extracted without accounting for nonlinearity (x-axis) and accounting for linearity (y-axis) for data sets I-VII. Each point is an epistatic coefficient, colored by order. Error bars are standard deviations from bootstrap replicates of each fitting approach.

bioRxiv preprint doi: https://doi.org/10.1101/072256; this version posted December 2, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

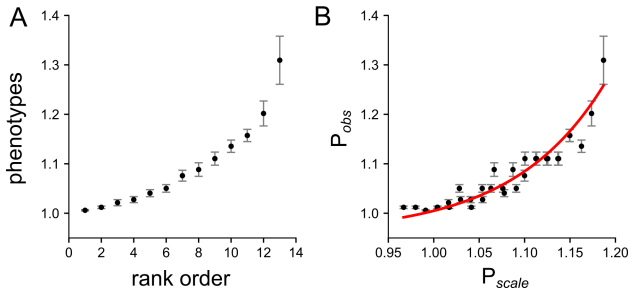

Fig S6: Exponential fitness model leads to global nonlinearity in the β-lactamase data set (III). A) A recapitulation of the map used in the original publication (Weinreich et al. 2006). We first rank-ordered the genotypes according to the measured property (the minimum inhibitory concentration of a β-lactam antibiotic against a clonal population of bacteria expressing that protein). This gave us 13 classes of genotypes, as some genotypes had equivalent MIC values. We then drew 3,000 random fitness values from the distribution $W = 1+x$, where x is an exponential distribution centered around $\bar{x} = 0.1$. We took the top 13 values from this distribution and assigned them, in value order, to each of the 32 β-lactamase genotypes. Panel A shows the average and standard deviation of the fitness values W assigned to each of these ranks if we repeat the protocol above 1,000 times. B) Best fit for the power-transform for data set III. Solid red line denotes the best fit (nonlinear). This fit successfully pulls out the original distribution of W.