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

Abstract

High-order epistasis has been observed in many genotype-phenotype maps. These multi-way interactions could have profound implications for evolution and may be useful for dissecting complex traits. Previous analyses have assumed a linear genotype-phenotype map, and then applied a linear high-order epistasis model to dissect epistasis. The assumption of linearity has not been tested in most of these data sets. Using simulations, we demonstrate that neglecting nonlinearity leads to spurious high-order epistasis. We find we can account for this nonlinearity in simulated maps using a power transform. We then measure and account for nonlinearity in experimental maps for which high-order epistasis has been previously reported. When applied to seven experimental genotype-phenotype maps, we find that five of the seven exhibited nonlinearity. Correcting for this nonlinearity had a large effect on the magnitudes and signs of the estimated high-order epistatic coefficients, but only a minor effect on additive and pairwise epistatic coefficients. Even after accounting for nonlinearity, we found statistically significant fourth-order epistasis in every map studied. One map even exhibited fifth-order epistasis. The contributions of high-order epistasis to the total variation in the map ranged from 2.2% to 31.0%, with an average across maps of 12.7%. Our work describes a simple method to account for nonlinearity in binary genotype-phenotype maps. Further, it provides strong evidence for extensive high-order epistasis, even after nonlinearity is taken into account.

Introduction

Epistasis is an important feature of genotype-phenotype maps (Wolf et al. 2000; Phillips 2008; Breen et al. 2012). It provides powerful insights for dissecting complex traits and regulatory pathways (Carlborg and Haley 2004; Shao et al. 2008; Hill et al. 2008; Wu and Lin 2006; Sackton and Hartl 2016). Further, it can play important roles in shaping evolutionary dynamics and outcomes (Poon and Chao 2005; Weinreich et al. 2006; Blount et al. 2008; Bridgham et al. 2009; Stern and Orgogozo 2009; Bloom et al. 2010; Østman et al. 2011; Pollock et al. 2012; Salverda et al. 2011; Breen et al. 2012; Soylemez and Kondrashov 2012; Dickinson et al. 2013; de Visser and Krug 2014; Harms and Thornton 2014; Kryazhimskiy et al. 2014; Shah et al. 2015).

Recent work has 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). High-order epistasis raises some intriguing possibilities. If it can be interpreted mechanistically, it may help dissect the complex architecture of biological systems (Lehár et al. 2008; Hu et al. 2011, 2013; Taylor and Ehrenreich 2015). Conversely, neglecting high-order epistasis could introduce bias into analyses of low-order epistasis (Otwinowski and Plotkin 2014). High-order epistasis also has profound implications for evolution (Weinreich et al. 2013). 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). High-order epistasis could, in principle, lead to long-range dependency across the map, such that a mutation has a different effect when introduced first, second, third, or even later in an evolutionary trajectory. This would amplify the importance of processes like contingency and entrenchment, which depend on mutations having different effects when introduced early or late in an evolutionary trajectory (Shah et al. 2015; Harms and Thornton 2014; Bridgham et al. 2009; Pollock et al. 2012).

Because the word *epistasis* is used in different, sometimes contradictory, ways in the literature (Phillips 2008), we will be explicit: we use epistasis to refer to the quantitative difference in the

phenotypic effect of mutations introduced together versus separately (sometimes called *statistical epistasis*) (Cordell et al. 2001; Phillips 1998, 2008). High-order epistasis is the difference in phenotype for a combination of mutations introduced together relative to the sum of their individual and low-order epistatic effects (Horovitz 1996; Cordell et al. 2001; Cordell 2002; Poelwijk et al. 2015).

High-order epistasis is thought provoking, but its biological and evolutionary interpretation is unclear. A major deficiency of previous studies is the assumption that phenotypes scale linearly (Anderson et al. 2015; Poelwijk et al. 2015; Weinreich et al. 2013; Yokoyama et al. 2015). If these maps are nonlinear, high-order epistasis may be an artifact arising from the assumption of linearity (Phillips 2008; Mani et al. 2008).

The difficulty presented by nonlinearity can be illustrated with an example. Imagine two mutations to an enzyme. When expressed in bacteria, these mutant enzymes exhibit negative epistasis on bacterial growth rate. This epistasis could have two origins. The first is at the level of the enzyme chemistry itself: maybe the mutations have a specific interaction that alters enzyme chemistry. This epistasis at the level of the enzyme directly translates to epistasis in growth rate (Fig 1A). Alternatively, epistasis could reflect a nonlinear relationship between enzyme activity and growth rate. When activity is low, small changes in activity lead to large changes in growth rate; when activity is high already, improving the activity further has little effect on growth rate. In this scenario, additive mutations at the level of enzyme chemistry will still exhibit negative epistasis at the level of bacterial growth rate (Fig 1B).

Epistasis arising at the level of the enzyme is *genotypic* epistasis: the genotype of the background determines non-additivity. Epistasis arising from growth-rate saturation is *phenotypic* epistasis: the phenotype of the background determines non-additivity. For clarity, we will refer to the former as *genotypic epistasis* and the latter as *phenotypic nonlinearity* throughout the text. Epistasis arising from phenotypic linearity has been referred to as prevailing magnitude epistasis (de Visser et al. 2009), global epistasis (Kryazhimskiy et al. 2014), and diminishing-returns epistasis (when effect-size decreases with increasing numbers of mutations) (Chou et al. 2011; MacLean et al. 2010; Otto and Feldman 1997; Tokuriki et al. 2012).

Linear models of epistasis assume genotypic epistasis—a scenario like Fig 1A—and attribute all variation in the effects of mutations to specific interactions between them. But this, potentially, conflates very different aspects of a biological system. If the map between mutations and observable is nonlinear, some fraction of the variation in the observable arises from nonlinearity. A linear model will naively partition this into the specific interactions. This will both overestimate the magnitude of genotypic epistasis and could even scramble the signs of specific epistatic coefficients. While the effects of nonlinearity can be understood intuitively for a two-site system, the effects on a high-order epistatic interaction are much more difficult to predict. Further, describing a nonlinear phenotype as specific interactions between mutations would miss the main "biology" of the system—in this case, saturation of growth rate.

These two origins of epistasis also have profoundly different evolutionary implications. Genotypic epistasis reveals specific collections of mutations that open or close evolutionary trajectories, potentially revealing highly-specific evolutionary contingencies. In contrast, epistasis arising from phenotypic nonlinearity reveals general limits on evolution, but does not imply radical dependence on a specific genetic background to follow a given evolutionary trajectory (Harms and Thornton 2014). For example, recent work has shown that pairwise genotypic epistasis leads to sequence-level unpredictability, while a nonlinear map leads to predictable phenotypes in evolution (Kryazhimskiy et al. 2014).

Given these considerations, we set out separate the effects of high-order genotypic epistasis and phenotypic nonlinearity in genotype-phenotype maps. We start with simulated maps with known genotypic epistasis and phenotypic nonlinearity, and then turn our attention to experimental maps in which high-order epistasis has been noted previously (Weinreich et al. 2013; Anderson et al. 2015; Poelwijk et al. 2015). Through this analysis, we find that both phenotypic nonlinearity and high-order genotypic epistasis make large contributions to experimental 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 L^{th}-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 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.

Genotypic epistasis model

We dissected genotypic epistasis using a linear epistasis model that decomposes binary genotype-phenotype maps into coefficients that capture contributions from individual mutations and interactions between them. These have been discussed extensively elsewhere (Heckendorn and Whitley 1999; Poelwijk et al. 2015; Weinreich et al. 2013); however, in the interest of clarity, we will briefly and informally review them here.

A linear, high-order epistasis model transforms a genotype-phenotype map into an orthogonal set of vectors (i.e. a change of basis) that account for all variation in the map (Fig 2). The lengths and signs of the vectors are epistatic coefficients that quantify the effect of mutations or 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.

$$\vec{P} = \mathbf{X}\vec{\beta}: \tag{1}$$

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

$$\vec{\beta} = \mathbf{X}^{-1}\vec{P}. \tag{2}$$

$\mathbf{X}$ can be formulated in a variety of ways (Poelwijk et al. 2015), but a common form in the genetics literature is derived from Walsh polynomials (Heckendorn and Whitley 1999; Weinreich et al. 2013; Poelwijk et al. 2015). 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 2). We encoded each mutation in each site in each genotype as -1 (wildtype) or +1 (mutant) (Heckendorn and Whitley 1999; Weinreich et al. 2013; Poelwijk et al. 2015). This leads to the following matrix for a three-mutation genotype-phenotype map:

$$\begin{bmatrix} P_{000} \\ P_{100} \\ P_{010} \\ P_{001} \\ P_{110} \\ P_{101} \\ P_{011} \\ P_{111} \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_i \\ \beta_j \\ \beta_k \\ \beta_{ij} \\ \beta_{ik} \\ \beta_{jk} \\ \beta_{ijk} \end{bmatrix} = \mathbf{X}_{walsh}\vec{\beta}. \tag{3}$$

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 ancestral state is given the bits $(1, 1, 1)$. The remaining states are encoded with bits that form corners of a tetrahedron. For example, the ancestral state 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 $(1, -1, -1)$.

Nonlinear scales

We accounted for nonlinearity in the genotype-phenotype map by a power transformation (see Results). The independent variable for the transformation was $\vec{P}_{add}$, the predicted phenotypes of all genotypes assuming purely additive affects for each mutation. The estimated additive phenotype of genotype i, is given by:

$$\hat{P}_{add,i} = \sum_{j=1}^{j \leq L} \langle \Delta P_j \rangle \, x_{i,j} \tag{4}$$

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 genotype-phenotype data.

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 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:

$$P_{obs,linear} = \{\hat{\lambda}(GM)^{\lambda-1}(P_{obs} - \hat{B}) - 1\}^{1/\hat{\lambda}} - \hat{A}. \quad (5)$$

Experimental uncertainty

We used a bootstrap approach to propagate uncertainty in measured phenotypes into uncertainty in genotypic 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, just like $\vec{P}$; 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 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

Phenotypic nonlinearity induces apparent high-order genotypic epistasis

Our first goal was to understand how phenotypic nonlinearity affects estimates of genotypic high-order epistasis. We constructed an additive five-site binary genotype-phenotype map, applied increasing amounts of nonlinearity, and then decomposed the map using a high-order genotypic epistasis model. To add nonlinearity, we transformed each phenotype using a simple saturation model:

$$P_{g,trans} = \frac{(1+K)P_g}{1+KP_g}, \tag{6}$$

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 3A. We then calculated $P_{g,trans}$ using the relatively shallow ($K = 2$) saturation curve shown in Fig 3B. 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 genotypic, high-order epistasis (Fig 3C). We observed epistasis up to the fourth order, despite building the map with additive coefficients. This result is unsurprising: the only mechanism by which a linear model can account for variation in phenotype is through epistatic coefficients. 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 3C, purple). But this four-way interaction is an artifact of the nonlinearity in phenotype of the map, rather than a specific interaction.

Genotypic epistasis and phenotypic nonlinearity induce different patterns of nonadditivity

Our next question was whether we could separate the effects of phenotypic nonlinearity and genotypic epistasis in binary maps. For a pair of mutations, it is impossible to distinguish these two origins of epistasis, as they give identical signals (Fig 1). As more mutations are characterized, however, it may become possible to disentangle these effects. In particular, by measuring the effect of a mutation across a large number of genetic backgrounds, one might be able to ask to what extent the genotype versus the phenotype of each genetic background predicts mutational effects.

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 additive mutational effects (P_{add}) (Rokyta et al. 2011; Schenk et al. 2013). In the absence of any epistasis, P_{obs} equals P_{add}, because each mutation would have the same, additive effect in all backgrounds. As a result, deviation from the $P_{obs} = P_{add}$ line reflects epistasis.

To disentangle the effects of genotypic epistasis from phenotypic nonlinearity, we simulated maps including both forms of epistasis and then constructed P_{obs} vs. P_{add} plots. We added genotypic epistasis by generating random epistatic coefficients then calculating linear phenotypes using Eq. 1. We introduced nonlinearity by transforming these phenotypes with Eq. 6. 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 genotypic epistasis and phenotypic nonlinearity.

Genotypic epistasis and a phenotypic nonlinearity gave qualitatively different P_{obs} vs. P_{add} plots. Fig 4A shows plots of P_{obs} vs. P_{add} for increasing phenotypic nonlinearity (left-to-right) and genotypic epistasis (top-to-bottom). As phenotypic nonlinearity increases, P_{obs} curves systematically relative to P_{add}. The smallest phenotypes are underestimated and the largest phenotypes overestimated, reflecting the saturation we added to the map. In contrast, genotypic epistasis induces random scatter away from the $P_{obs} = P_{add}$ line.

These patterns can be understood from the origins of epistasis. Genotypic epistasis is determined

solely by genotype, without reference to phenotype. This leads to scatter away from the $P_{add} = P_{obs}$ line, but no systematic structure in the curve with respect to P_{add}. For a phenotypic nonlinearity, the magnitude of the epistasis depends on the magnitude of the phenotype. This induces systematic structure in the relationship between P_{add} and P_{obs}—in this case, a saturating curve.

Nonlinearity can be separated from genotypic epistasis

The P_{obs} vs. P_{add} plots suggest an approach to disentangle genotypic epistasis from nonlinearity in phenotype. By fitting a function to the P_{obs} vs P_{add} curve, we can describe the observed nonlinearity (Schenk et al. 2013). Once the form of the nonlinearity is known, we can then linearize the phenotypes using this function. Any variation remaining after linearization (i.e. scatter) is due to genotypic epistasis.

In the absence of knowledge about the source of the nonlinearity, a natural choice for such an analysis is a power transform, 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 (Box and Cox 1964; Carroll and Ruppert 1981) 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 would not know P_{add}. We determined P_{add} above using the additive coefficients used to generate the space. These are not known in a real map. We therefore estimated P_{add} from P_{obs}. We determined $\hat{P}_{add}$ 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. 4).

We then fit the power transform to P_{obs} vs. $\hat{P}_{add}$ (solid red line, Fig 4B). The curve captures the nonlinearity added in the simulation. We linearized P_{obs} using the fit model (Eq. 5), and then extracted high-order genotypic epistatic coefficients. The extracted coefficients were highly correlated with the coefficients used to generate the map ($R^2 = 0.998$) (Fig 4C). In contrast, applying the linear epistasis model to this map without first accounting for nonlinearity gives much greater scatter between the input and output coefficients ($R^2 = 0.934$) (Fig 4D). This occurs

because phenotypic variation from nonlinearity is incorrectly partitioned into the linear epistatic coefficients.

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

Our next question was whether the high-order genotypic epistasis observed in experimental maps could be accounted for as an artifact of phenotypic nonlinearity. 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 5, 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 III is a free energy). Rather than asserting a scale by taking logarithms of 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 three were approximately linear (I, VII). All data sets gave random residuals after fitting the power transform (Fig 5, S1).

We then linearized the data with the power transform and re-measured genotypic epistasis. In an effort to avoid false positives, we took a conservative approach. We used bootstrap sampling of uncertainty in the measured phenotypes to determine the uncertainty of each epistatic coefficient (set 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 approach, we found high-order epistasis in every map studied (Fig 6A, S2). Every data set exhibited at least one statistically significant epistatic coefficient of fourth order or higher. We even detected statistically significant fifth-order genotypic epistasis (blue bar in Fig 6A, 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.

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. 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 genotypic epistatic coefficients extracted from each space. Fig 7 and S3 show correlation plots between genotypic epistatic coefficients extracted both with and without a correction for nonlinearity. The first-order coefficients were all highly correlated between the linear and nonlinear analyses for all data sets (Fig S4).

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 depending whether the phenotypic 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 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 7, III, bottom right quadrant).

Discussion

Nonlinearity is a common feature of genotype-phenotype maps

A key observation from our work is that the majority of the genotype-phenotype maps exhibit nonlinearity. This is, perhaps, expected given the nonlinearity intrinsic in biological systems. Because these maps cover relatively large stretches of sequence space—six mutations across—factors outside specific interactions between mutations come into play. While this complicates analyses of genotypic epistasis, it also provides insight into the architecture of these systems.

The less-than-additive maps were unsurprising. Many have previously observed saturating, 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 difficult as that phenotype improves.

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. 5), it appears this is an artifact of the original analysis used to generate the data set. 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. Our epistasis model

extracts this original exponential distribution (Fig S5). 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. 5). 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 statistical epistasis in binary genotype-phenotype maps. Unlike pairs of mutations, where nonlinearity is difficult to estimate because of the paucity of observations, the large number of genotypes characterized in these maps makes it possible to detect nonlinearity directly from the data. This provides information about the architecture of the system—in the form of its nonlinearity—and gives confidence in the assignment of genotypic epistatic coefficients. Our software 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 the choice of nonlinear model alters the observed high-order coefficients. A power transform captures the primary curvature within a data set. Use of a more complicated function would shift more of the variation in the data away from genotypic epistasis and into global structure in the map. Adding this complexity could be motivated by external biological knowledge about the map (Schenk et al. 2013). It could also be motived by examination of the P_{obs} vs $\hat{P}_{add}$ plot. Because genotypic epistasis is scatter off the scale line, standard nonlinear regression tools such as an F-test, Akaike Information criterion, and examination of fit residuals

could be used to identify a nonlinear function that captures the maximum amount of phenotypic variation in the data set without fitting stochastic (genotypic) variation in the data.

Practically, we believe the choice of a different nonlinear model would have little effect on our results for these data sets. Other possible models of nonlinearity—gamma functions, exponential functions, polynomials, etc.—would likely give similar curves when fit to the small amount of curvature in these data sets. Further, our bootstrap protocol integrates over uncertainty in the power transform coefficients, so higher and lower curvature fits consistent with the data are incorporated into the uncertainty in the epistatic coefficients.

Neglecting nonlinearity entirely has variable effects on different orders of epistatic coefficients. Overall, low-order coefficients were more robust to the linear assumption than high-order epistatic coefficients. Data set IV is a clear example of this behavior. The map exhibited noticeable nonlinearity (Fig 5). The first- and second-order terms were well correlated between the linear and nonlinear analyses (Fig 7, S3, S4). 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 these data sets, which assumed linear scales, are correct in their identification of the key mutations responsible for variation in the map, but that their analysis of higher-order epistatic coefficients was not reliable.

High-order epistasis

Finally, our work shows that high-order epistasis is 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 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 meaning of these genotypic epistatic coefficients is also an open question. 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? What can this epistasis tell us about the biological underpinning of these maps? The evolutionary implications are also unclear. How does this high-order epistasis shape evolutionary outcomes and 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.

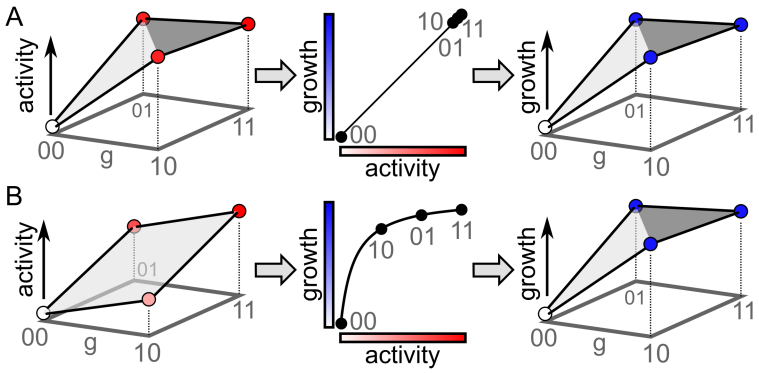

Fig 1. Epistasis can arise at the level of genotype or phenotype. A) Genotypic epistasis. The leftmost panel shows a two-site genotype/enzyme-activity map. Genotypes are given by numerical coordinates, with "0" denoting wildtype at a site and "1" denoting a mutation. Enzyme activity is encoded both on the z-axis and as a spectrum from white to red. The middle panel shows how enzyme activity (white-to-red, x-axis) maps to growth rate (white-to-blue, y-axis). In this case, the map is linear. The rightmost panel shows the observed epistasis in growth rate given the genotype/enzyme activity map and activity to growth rate map. B) Epistasis arising from phenotypic nonlinearity. The sub panels are colored and labeled as in panel B. For this map, enzyme activity behaves additively (left), but the relationship between activity and growth saturates (middle). This leads to epistasis in growth rate that is indistinguishable from genotypic epistasis (right).

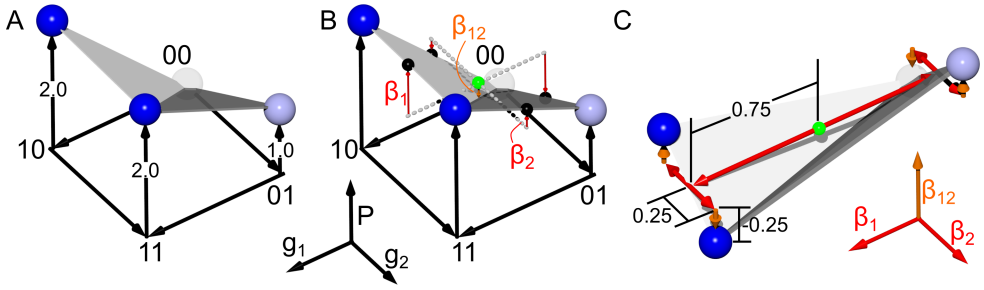

Fig 2: Genotypic epistasis can be quantified using Walsh polynomials. A) A genotype-phenotype map exhibiting negative epistasis. Axes are genotype at position 1 (g_1), genotype at position 2 (g_2), and phenotype (P). For genotypic axes, "0" denotes wildtype and "1" denotes a mutant. Phenotype is encoded 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. Geometrically, one finds the center of the genotype-phenotype map (green sphere). The first-order coefficients β_1 and β_2 (red arrows) are the average effect of each mutation relative to this center. The second-order coefficient β_{12} (orange arrow) is the magnitude and sign of the vector, along the phenotype axis, between the average phenotype and the line drawn between P_{00} and P_{11} (the "fold" in the map). C) The genotype-phenotype map transformed into the Walsh space. Axes are epistatic coefficients β_1, β_2 and β_{12}. Each phenotype, relative to the center of the space, is a linear sum of all epistatic coefficients (noted on the figure for P_{11} as dimensions).

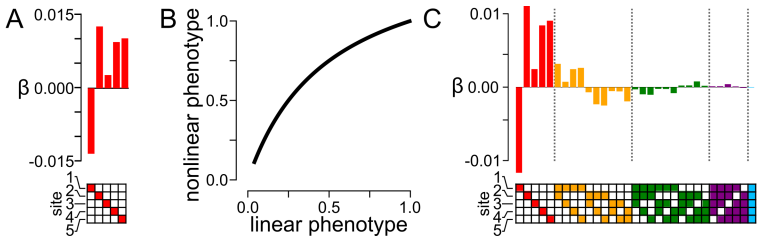

Fig 3: Nonlinearity in phenotype creates spurious high-order epistatic coefficients. A) Simulated, random, first-order epistatic coefficients. The mutated site is indicated by panel below the bar graph; bar indicates magnitude and sign of the epistatic coefficient. B) A nonlinear map between a linear 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.

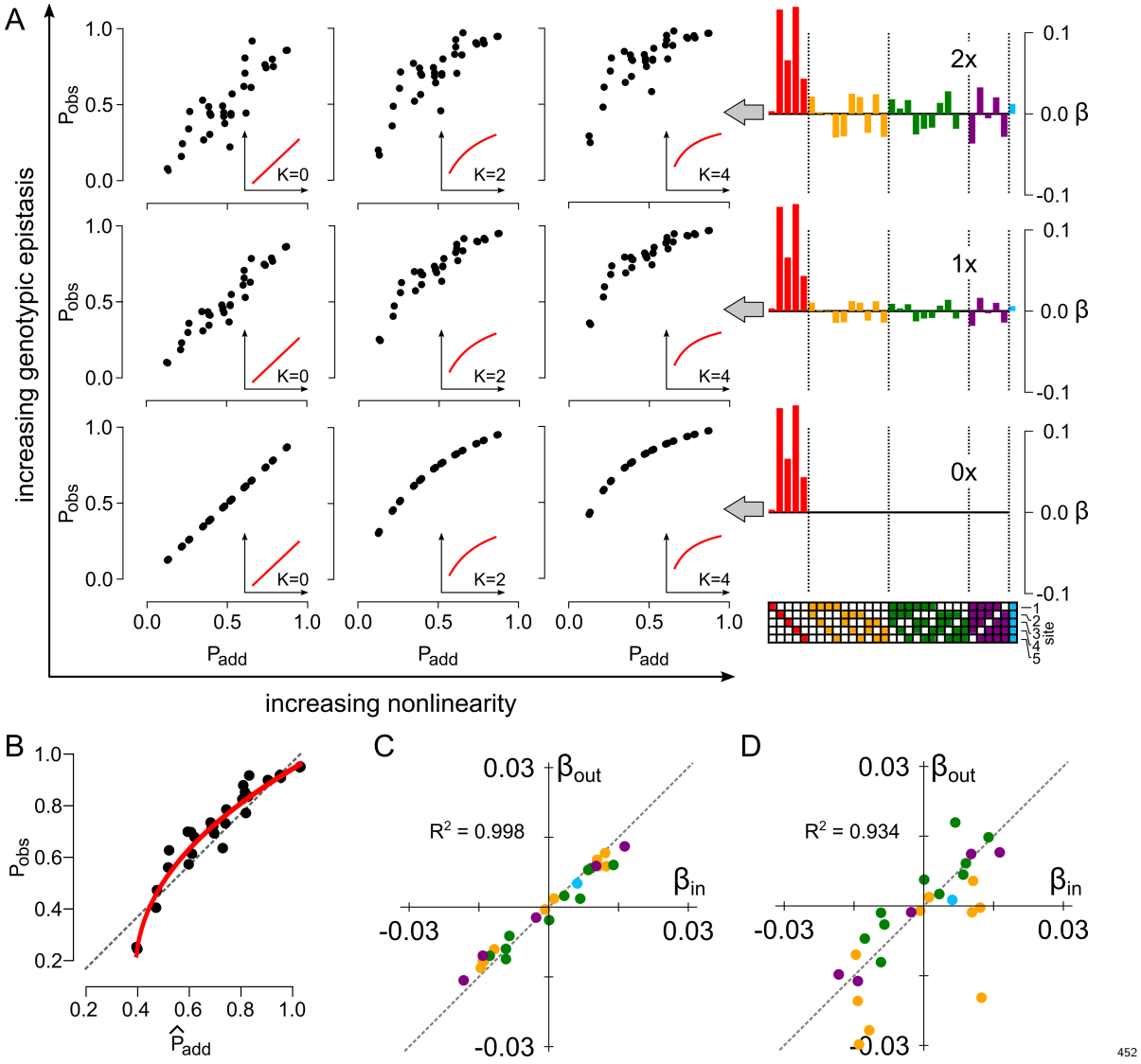

Fig 4: Genotypic epistasis and phenotypic nonlinearity induce different patterns of nonadditivity. A) Patterns of nonadditivity for increasing genotypic epistasis and phenotypic nonlinearity. Main panel shows grid ranging from no epistasis (bottom left) to high genotypic epistasis and nonlinearity (top 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 magnitude of input genotypic epistasis, with colors and annotation as in Fig 3C. B) Plot of P_{obs} against $\hat{P}_{add}$ for the

middle sub panel in 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 correction for nonlinearity with 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.

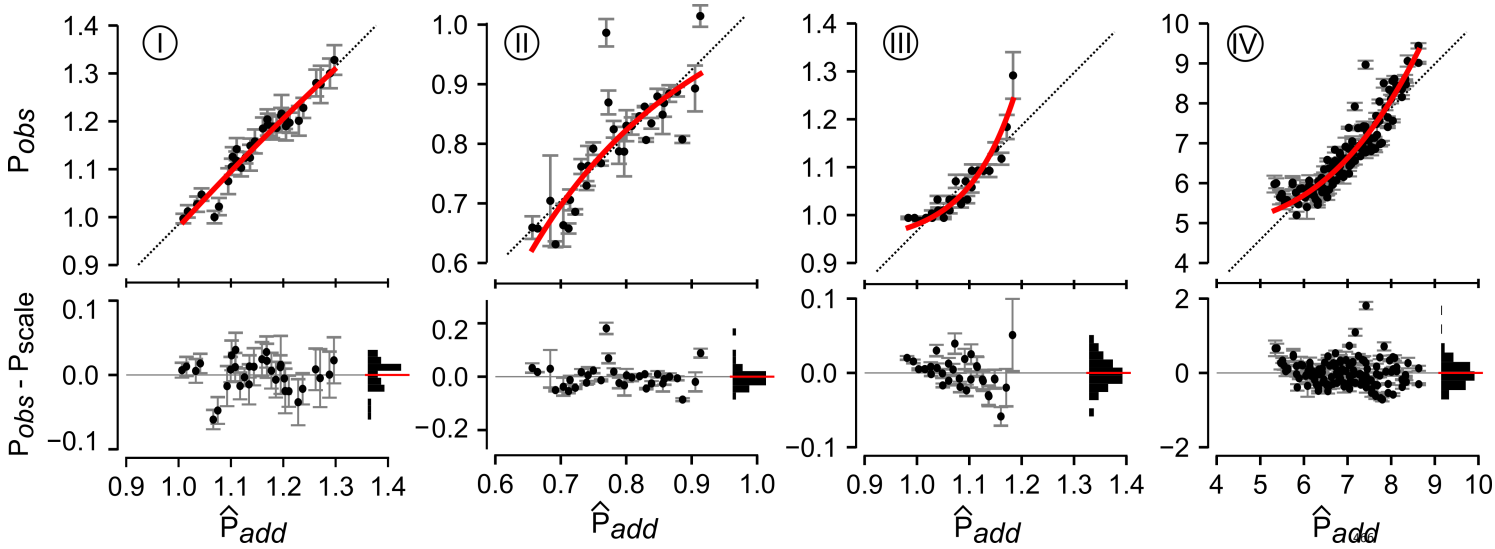

Fig 5: Experimental genotype-phenotype maps exhibit nonlinear phenotypes. Plots show observed phenotype P_{obs} plotted against $\hat{P}_{add}$ (Eq. 4) for data sets I through IV. Points are individual genotypes. Error bars are experimental standard deviations in phenotype. Red lines are the fit of the 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.

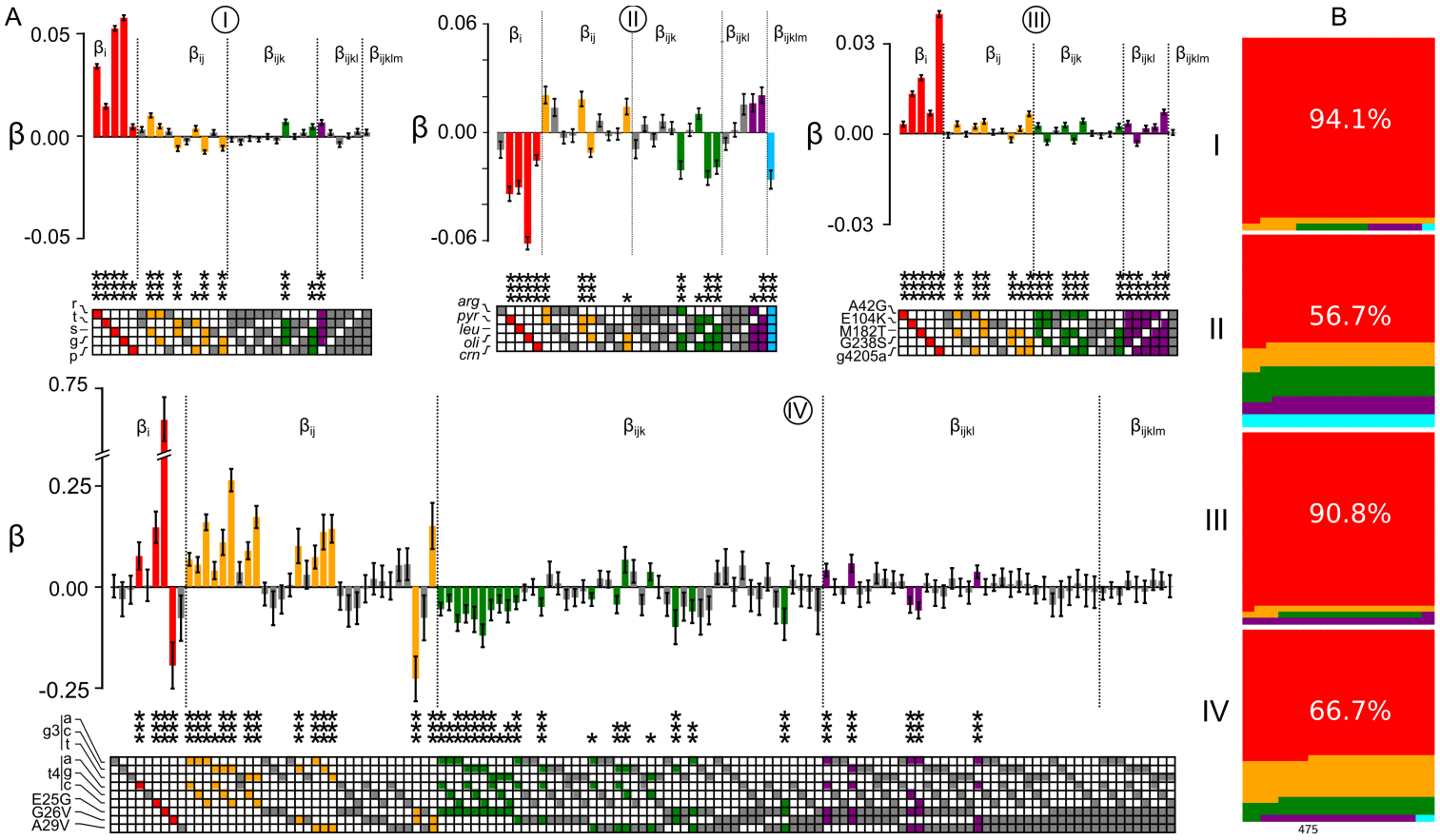

Fig 6: High-order epistasis is present in genotype-phenotype maps. A) Panels show epistatic coefficients extracted from data sets I-IV (Table 1, data set label circled above each graph). Bars denote coefficient magnitude and sign; error bars are propagated measurement uncertainty. Color denotes the order 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

to area.

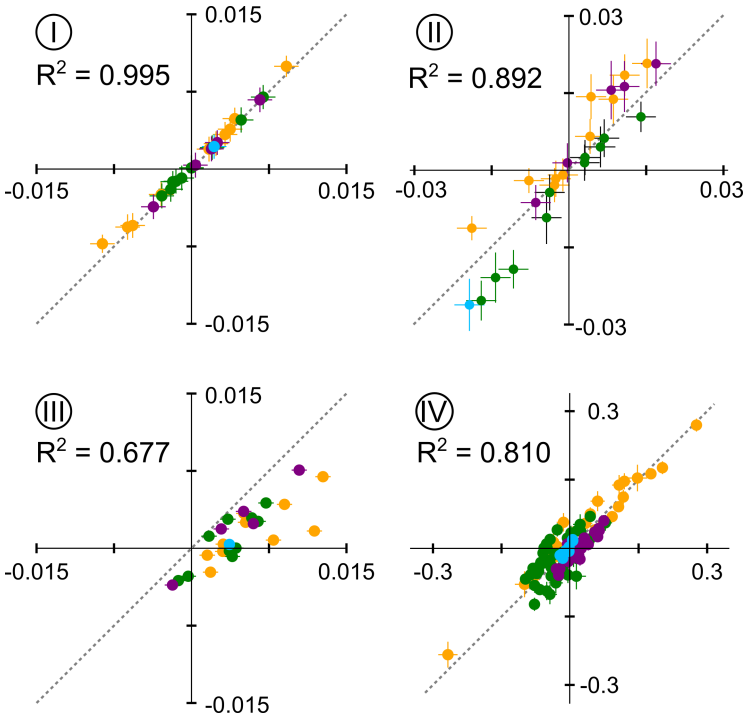

Fig 7: Nonlinear phenotypes distort measured epistatic coefficients. Sub-panels show correlation plots between epistatic coefficients extracted without accounting for nonlinearity (x-axis) and accounting for linearity (y-axis) for data sets I-IV. Each point is an epistatic coefficient, colored by order. Error bars are standard deviations from bootstrap replicates of each fitting approach.

Table 1

ID	genotype	phenotype	N	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).

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.

Jesse D. Bloom, Lizhi Ian Gong, and David Baltimore. Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance. *Science*, 328(5983):1272–1275, June 2010. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1187816.

Zachary D. Blount, Christina Z. Borland, and Richard E. Lenski. Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *PNAS*, 105(23): 7899–7906, October 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0803151105.

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.

Michael S. Breen, Carsten Kemena, Peter K. Vlasov, Cedric Notredame, and Fyodor A. Kondrashov. Epistasis as the primary factor in molecular evolution. *Nature*, 490(7421):535–538, October 2012. ISSN 0028-0836. doi: 10.1038/nature11510.

Jamie T. Bridgham, Eric A. Ortlund, and Joseph W. Thornton. An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature*, 461(7263):515–519, September 2009. ISSN 0028-0836. doi: 10.1038/nature08249.

Örjan Carlborg and Chris S. Haley. Epistasis: too often neglected in complex trait studies? *Nat Rev Genet*, 5(8):618–625, August 2004. ISSN 1471-0056. doi: 10.1038/nrg1407.

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.

Heather J. Cordell, John A. Todd, Natasha J. Hill, Christopher J. Lord, Paul A. Lyons, Laurence B. Peterson, Linda S. Wicker, and David G. Clayton. Statistical Modeling of Interlocus Interactions in a Complex Disease: Rejection of the Multiplicative Model of Epistasis in Type 1 Diabetes. *Genetics*, 158(1):357–367, May 2001. ISSN 0016-6731, 1943-2631.

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.

J. Arjan G. M. de Visser and Joachim Krug. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet*, 15(7):480–490, July 2014. ISSN 1471-0056. doi: 10.1038/nrg3744.

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.

Bryan C. Dickinson, Aaron M. Leconte, Benjamin Allen, Kevin M. Esvelt, and David R. Liu. Experimental interrogation of the path dependence and stochasticity of protein evolution using

phage-assisted continuous evolution. *PNAS*, 110(22):9007–9012, May 2013. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1220670110.

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.

William G. Hill, Michael E. Goddard, and Peter M. Visscher. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLOS Genet*, 4(2):e1000008, February 2008. ISSN 1553-7404. doi: 10.1371/journal.pgen.1000008.

Amnon Horovitz. Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Folding and Design*, 1(6):R121–R126, December 1996. ISSN 1359-0278. doi: 10.1016/S1359-0278(96)00056-9.

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.

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.

Ramamurthy Mani, Robert P. St Onge, John L. Hartman, Guri Giaever, and Frederick P. Roth. Defining genetic interaction. *Proc. Natl. Acad. Sci. U.S.A.*, 105(9):3461–3466, March 2008. ISSN 1091-6490. doi: 10.1073/pnas.0712255105.

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.

Bjørn Østman, Arend Hintze, and Christoph Adami. Impact of epistasis and pleiotropy on evolutionary adaptation. *Proceedings of the Royal Society of London B: Biological Sciences*, page rspb20110870, June 2011. ISSN 0962-8452, 1471-2954. doi: 10.1098/rspb.2011.0870.

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. The Language of Gene Interaction. *Genetics*, 149(3):1167–1171, January 1998. ISSN 0016-6731, 1943-2631.

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. *arXiv:1502.00726 [q-bio]*, February 2015.

David D. Pollock, Grant Thiltgen, and Richard A. Goldstein. Amino acid coevolution induces an evolutionary Stokes shift. *PNAS*, 109(21):E1352–E1359, May 2012. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1120084109.

Art Poon and Lin Chao. The Rate of Compensatory Mutation in the DNA Bacteriophage φX174. *Genetics*, 170(3):989–999, July 2005. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.104.039438.

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.

Timothy B. Sackton and Daniel L. Hartl. Genotypic Context and Epistasis in Individuals and Populations. *Cell*, 166(2):279–287, July 2016. ISSN 0092-8674. doi: 10.1016/j.cell.2016.06.047.

Merijn L. M. Salverda, Eynat Dellus, Florien A. Gorter, Alfons J. M. Debets, John van der Oost, Rolf F. Hoekstra, Dan S. Tawfik, and J. Arjan G. M. de Visser. Initial Mutations Direct Alternative Pathways of Protein Evolution. *PLOS Genet*, 7(3):e1001321, March 2011. ISSN 1553-7404. doi: 10.1371/journal.pgen.1001321.

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.

Haifeng Shao, Lindsay C. Burrage, David S. Sinasac, Annie E. Hill, Sheila R. Ernest, William O'Brien, Hayden-William Courtland, Karl J. Jepsen, Andrew Kirby, E. J. Kulbokas, Mark J. Daly, Karl W. Broman, Eric S. Lander, and Joseph H. Nadeau. Genetic architecture of complex traits: Large phenotypic effects and pervasive epistasis. *PNAS*, 105(50):19910–19914, December 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0810388105.

Onuralp Soylemez and Fyodor A. Kondrashov. Estimating the Rate of Irreversibility in Protein Evolution. *Genome Biol Evol*, 4(12):1213–1222, January 2012. ISSN , 1759-6653. doi: 10.1093/gbe/evs096.

David L. Stern and Virginie Orgogozo. Is Genetic Evolution Predictable? *Science*, 323(5915): 746–751, February 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1158997.

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.

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.

Jason B. Wolf, Edmund D. Brodie, and Michael John Wade. *Epistasis and the Evolutionary Process*. Oxford University Press, 2000. ISBN 978-0-19-512806-2.

Rongling Wu and Min Lin. Functional mapping — how to map and study the genetic architecture of dynamic complex traits. *Nat Rev Genet*, 7(3):229–237, March 2006. ISSN 1471-0056. doi: 10.1038/nrg1804.

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 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.

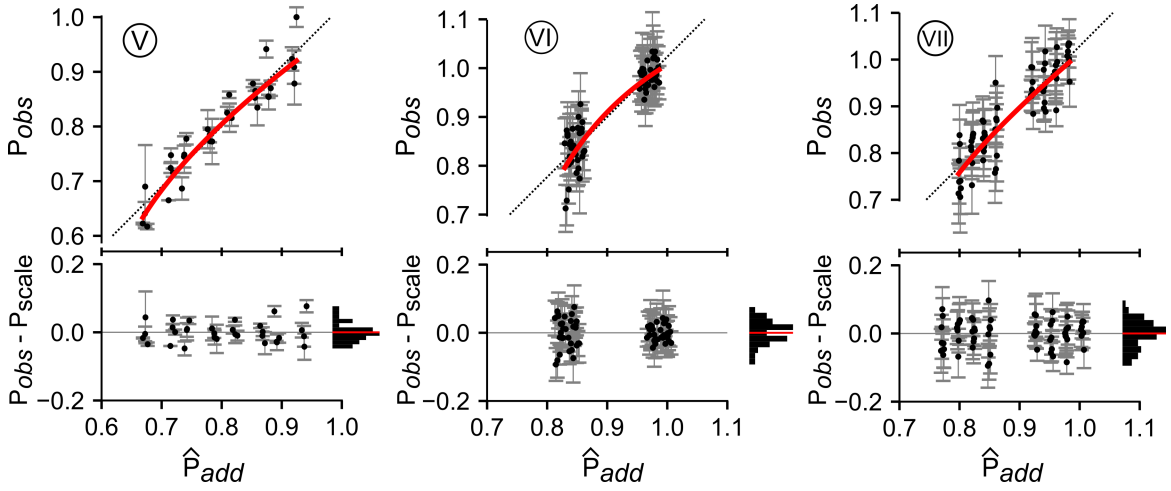

Fig S1: Experimental genotype-phenotype maps exhibit nonlinear phenotypes. Plots show observed phenotype P_{obs} plotted against $\hat{P}_{add}$ (Eq. 4) for data sets V through VII. Points are individual genotypes. Error bars are experimental standard deviations in phenotype. Red lines are the fit of 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.

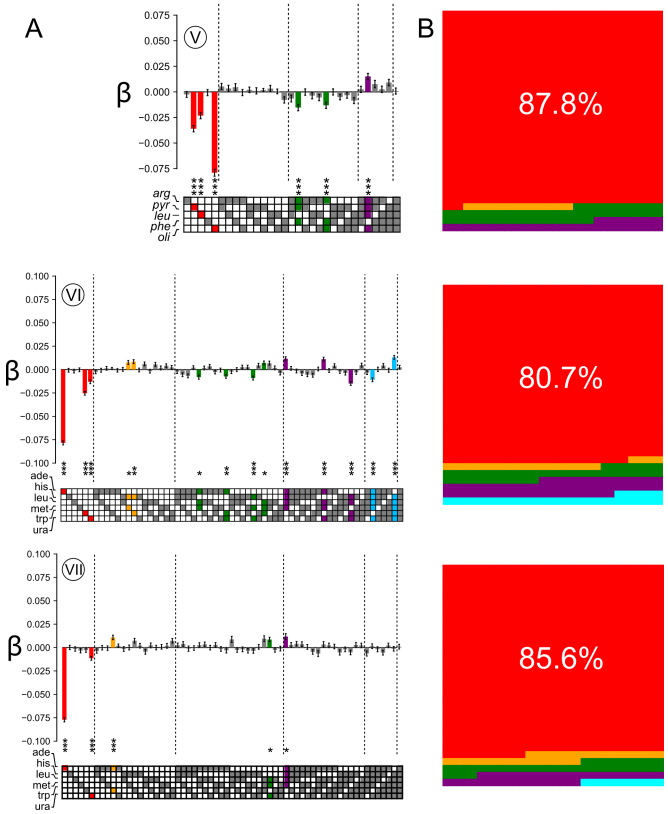

Fig S2: High-order epistasis is present in genotype-phenotype maps. A) Panels show epistatic coefficients extracted from data sets V-VII (Table 1, data set label circled above each graph). Bars denote coefficient magnitude and sign; error bars are propagated measurement uncertainty. Color denotes the 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.

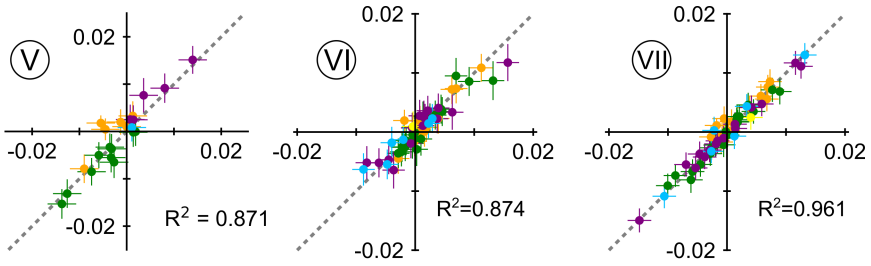

Fig S3: Nonlinear phenotypes distort measured epistatic coefficients. Sub-panels show correlation plots between epistatic coefficients extracted without accounting for nonlinearity (x-axis) and accounting for linearity (y -axis) for data sets V-VII. Each point is an epistatic coefficient, colored by order.

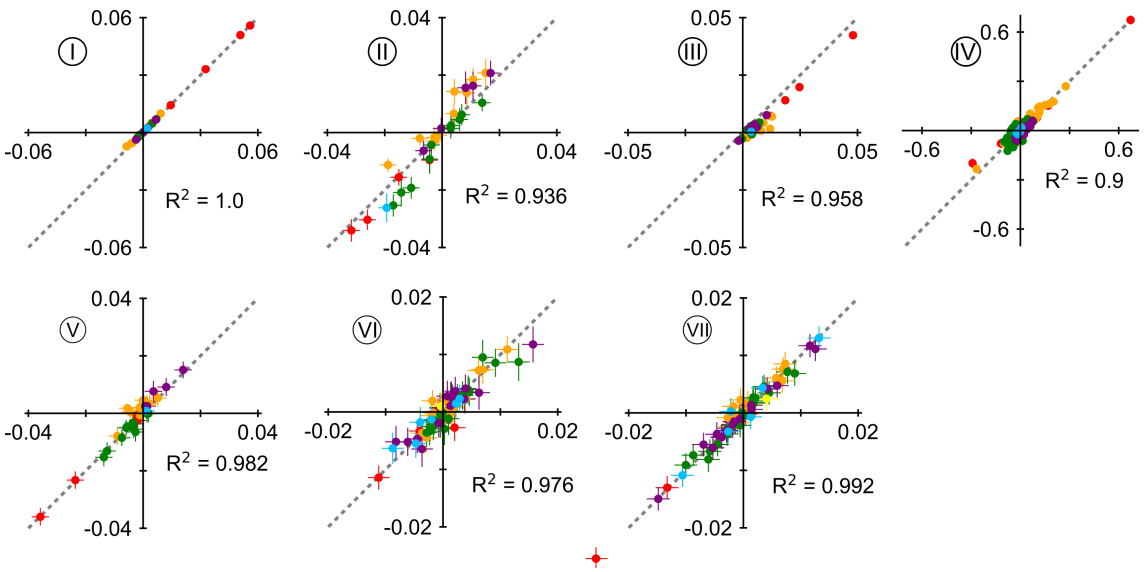

Fig S4: Additive coefficients are well estimated, even when nonlinearity is neglected. Sub-panels show correlation plots between both additive and epistatic coefficients extracted without accounting for nonlinearity (x-axis) and accounting for linearity (y-axis) for data sets I-VII. Each point is an epistatic coefficient, colored by order. Error bars are standard deviations from bootstrap replicates of each fitting approach.

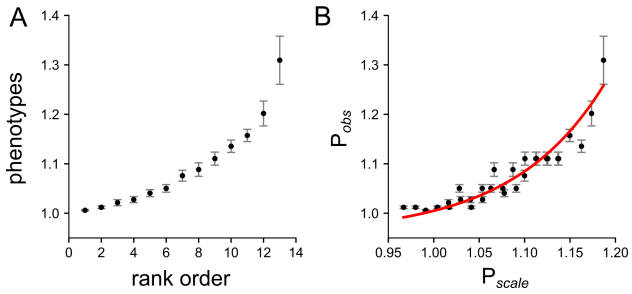

Fig S5: Exponential fitness model leads to global nonlinearity in the β-lactamase data set (III). A) A recapitulation of the map used in the original publication (Weinreich et al. 2006). We first rank-ordered the genotypes according to the measured property (the minimum inhibitory concentration of a 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.