

Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps

Zachary R. Sailer and Michael J. Harms¹

Institute of Molecular Biology, and Department of Chemistry and Biochemistry, University of Oregon, Eugene, Oregon 97403

ORCID ID: 0000-0002-0241-4122 (Z.R.S.)

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 article, 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.

KEYWORDS genotype-phenotype map; high-order epistasis; nonlinear; scale; statistical analysis

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; Hu *et al.* 2013; Weinreich *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 (Weinreich *et al.* 2013; Poelwijk *et al.* 2016). 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

Copyright © 2017 Sailer and Harms

doi: 10.1534/genetics.116.195214

Manuscript received August 29, 2016; accepted for publication January 9, 2017; published Early Online January 18, 2017.

Available freely online through the author-supported open access option.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material is available online at www.genetics.org/lookup/suppl/doi:10.1534/genetics.116.195214/-/DC1.

¹Corresponding author: 1370 Franklin Blvd., Room 269A Klamath Hall, Molecular Biology, University of Oregon, Eugene, OR 97403. E-mail: harms@uoregon.edu

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 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 (Heckendorn and Whitley 1999; Weinreich *et al.* 2013; Poelwijk *et al.* 2016).

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 article, 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.

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 (de Visser *et al.* 2009; Khan *et al.* 2011), and binary combinations of alleles within a biosynthetic network (Hall *et al.* 2010). Measured phenotypes included selection coefficients (Weinreich *et al.* 2006; de Visser *et al.* 2009; Khan *et al.* 2011), 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 American Standard Code for Information Interchange (ASCII) text format.

Nonlinear scale

We described nonlinearity in the genotype-phenotype map by a power transform (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 effects for each mutation. The estimated additive phenotype of genotype i is given by

$$\hat{P}_{\text{add},i} = \sum_{j=1}^{j \leq L} \langle \Delta P_j \rangle x_{ij}, \quad (1)$$

where $\langle \Delta P_j \rangle$ is the average effect of mutation j across all backgrounds, x_{ij} 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} :

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

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

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

where A and B are translation constants, GM is the geometric mean of $(\hat{\vec{P}}_{\text{add}} + A)$, and λ is a scaling parameter. We used standard nonlinear regression techniques to minimize d :

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

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:

$$P_{\text{obs,linear}} = \left[\hat{\lambda} (GM)^{\lambda-1} (P_{\text{obs}} - \hat{B}) + 1 \right]^{1/\hat{\lambda}} - \hat{A}. \quad (2)$$

High-order epistasis model

We dissected epistasis using a linear, high-order epistasis model. These have been discussed extensively elsewhere (Heckendorn and Whitley 1999; Weinreich *et al.* 2013; Poelwijk *et al.* 2016), 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.

$$\vec{P} = \mathbf{X}\hat{\beta} \quad (3)$$

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

$$\vec{\beta} = \mathbf{X}^{-1}\vec{P}. \quad (4)$$

\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; Weinreich *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 (Figure 1). To achieve this, we encoded each mutation at each site in each genotype as -1 (wild type) or $+1$ (mutant) (Heckendorn and Whitley 1999; Weinreich *et al.* 2013; Poelwijk *et al.* 2016). This has been called a Fourier analysis (Neidhart *et al.* 2013; Szendro *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 wild-type genotype as a reference and encode mutations as either 0 (wild type) or 1 (mutant) (Poelwijk *et al.* 2016).

One data set (IV, Table 1) 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 wild-type state is given the bits (1, 1, 1). The remaining states are encoded with bits that form corners of a tetrahedron. For example, the wild type 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 SD 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

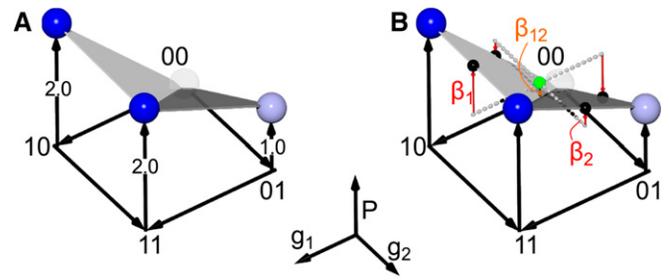


Figure 1 Epistasis can be quantified using Walsh polynomials. (A) A genotype-phenotype map exhibiting negative epistasis. Axes are genotype at position 1 (g_1), genotype at position 2 (g_2), and phenotype (P). For genotypic axes, “0” denotes wild type and “1” denotes a mutant. Phenotype is encoded both on the P -axis and as a spectrum from white to blue. The map exhibits negative epistasis: relative to wild type, 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) is 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} .

$\vec{P}_{\text{pseudo, scaled}}$. 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 pseudoreplicates. On average, convergence required $\sim 100,000$ 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://github.com/harmslab/epistasis>). We used the Python package scikit-learn for all regressions (Pedregosa *et al.* 2011). Plots were generated using matplotlib and Jupyter Notebooks (Hunter 2007; Perez and Granger 2007).

Data availability

The data sets and code used in this work are available at <https://github.com/harmslab/notebooks-nonlinear-high-order-epistasis>. The data sets are available in standard JSON format. The code is available as Jupyter Notebooks.

Results

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

Table 1 Data sets used in this study

ID	Genotype	Phenotype	L	Reference
I	Scattered genomic mutations	<i>Escherichia coli</i> fitness	5	Khan <i>et al.</i> (2011)
II	Chromosomes in asexual fungi	<i>Aspergillus niger</i> fitness	5	de Visser <i>et al.</i> (2009)
III	Protein point mutants	Bacterial fitness	5	Weinreich <i>et al.</i> (2006)
IV	DNA/protein point mutants	<i>In vitro</i> DNA/protein binding affinity	5	Anderson <i>et al.</i> (2015)
V	Chromosomes in asexual fungi	<i>A. niger</i> fitness	5	de Visser <i>et al.</i> (2009)
VI	Alleles in biosynthetic network	<i>Saccharomyces cerevisiae</i> haploid growth rate	6	Hall <i>et al.</i> (2010)
VII	Alleles in biosynthetic network	<i>S. cerevisiae</i> diploid growth rate	6	Hall <i>et al.</i> (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 two DNA sites (each of which have four possible bases) and three protein sites (each of which has two possible amino acids). ID, data set identifier.

genotype-phenotype map on a nonlinear scale, and then extracted epistasis assuming a linear scale. The nonlinear scale we chose was a saturating function:

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

where P_g is the linear phenotype of genotype g , $P_{g,\text{trans}}$ is the transformed phenotype of genotype g , and K is a scaling constant. As $K \rightarrow 0$, 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 Figure 2A. These coefficients included no epistasis. We then transformed P_g onto the nonlinear $P_{g,\text{trans}}$ scale using Equation 5 with the relatively shallow ($K = 2$) saturation curve shown in Figure 2B. Finally, we applied a linear epistasis model to $P_{g,\text{trans}}$ to extract epistatic coefficients.

We found that nonlinearity in the genotype-phenotype map induced extensive high-order epistasis when the nonlinearity was ignored (Figure 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 coefficients (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 inter-

preted as specific interactions between mutations. For example, this analysis identifies a specific interaction between mutations one, two, four, and five (Figure 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 nonadditivity

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 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 $P_{\text{obs}} = P_{\text{add}}$ 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 Equation 3. We introduced nonlinearity by transforming these phenotypes with Equation 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. Figure 3A shows plots of P_{obs} vs. P_{add} for increasing nonlinearity (left to right) and high-order epistasis (bottom to top). As nonlinearity

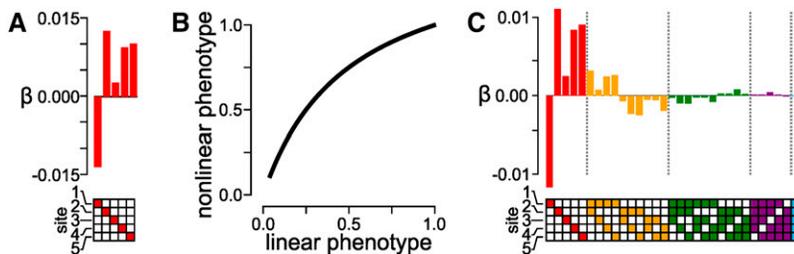


Figure 2 Nonlinearity in phenotype creates spurious high-order epistatic coefficients. (A) Simulated, random, first-order epistatic coefficients. The mutated site is indicated by panel below the bar graph; bar indicates magnitude and sign of the additive coefficient. (B) A nonlinear map between a linear phenotype and a saturating, nonlinear phenotype. The first-order coefficients in (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 (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 ■ in the grid below the bars indicate the identity of mutations that contribute to the coefficient.

Figure 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 (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 (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 ■ in the grid below the bars indicate the identity of mutations that contribute to the coefficient.

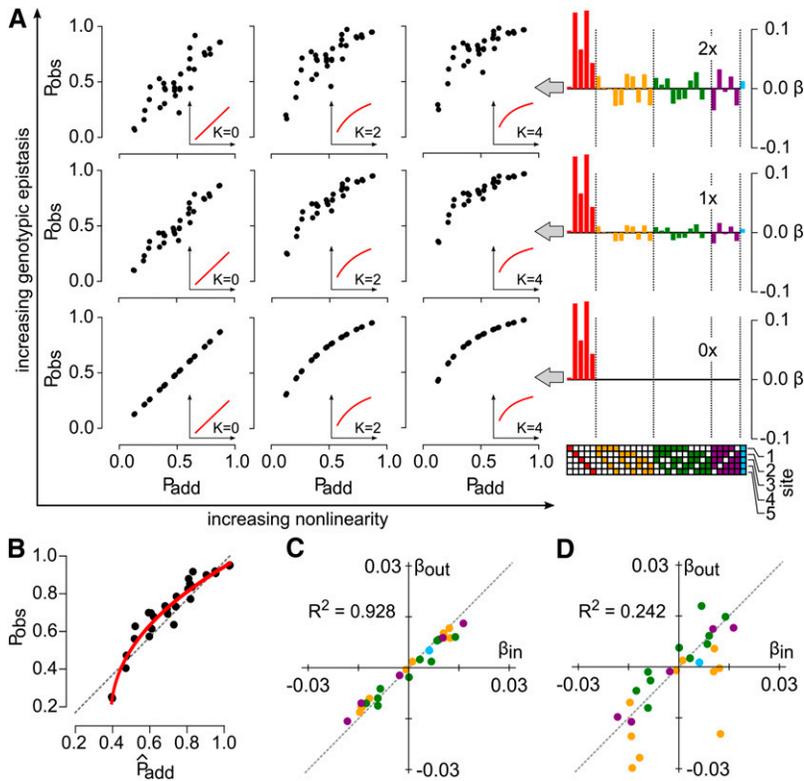


Figure 3 Epistasis and nonlinear scale induce different patterns of nonadditivity. (A) Patterns of nonadditivity for increasing epistasis and nonlinear scale. Main panel shows a grid ranging from no epistasis, linear scale (bottom left) to high epistasis, highly nonlinear scale (top right). Insets in subpanels 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 Figure 2C. (B) Plot of P_{obs} against \hat{P}_{add} for the middle subpanel in (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 top-left quadrant. (D) Correlation between epistatic coefficients input into the simulation and extracted from the simulation without application of the power transform.

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 underestimated and the largest phenotypes are overestimated. In contrast, high-order epistasis induces random scatter away from the $P_{\text{obs}} = P_{\text{add}}$ 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 (Equation 1).

We fit the power transform to P_{obs} vs. \hat{P}_{add} (solid red line, Figure 3B). The curve captures the nonlinearity added in the simulation. We linearized P_{obs} using the fit model (Equation 2), and then extracted epistatic coefficients. The extracted coefficients were highly correlated with the coefficients used to generate the map ($R^2 = 0.928$) (Figure 3C). 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 ($R^2 = 0.242$) (Figure 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 data set (Figure 4 and Supplemental Material, Figure S1 and File S1). We expected some phenotypes to be multiplicative (*e.g.*, data

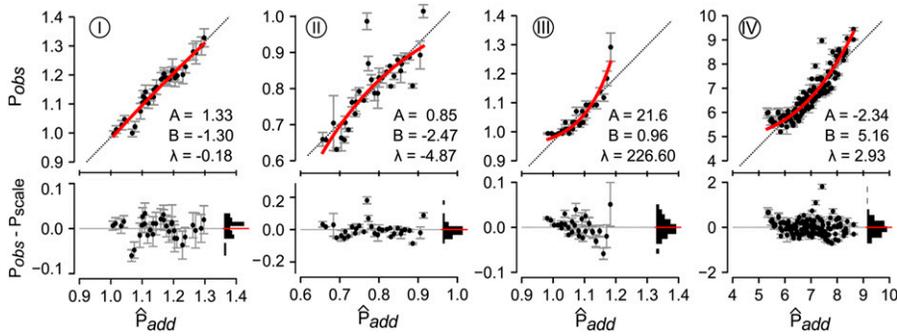


Figure 4 Experimental genotype-phenotype maps exhibit nonlinear phenotypes. Plots show observed phenotype P_{obs} plotted against \hat{P}_{add} (Equation 1) for data sets I–IV. Points are individual genotypes. Error bars are experimental SDs in phenotype. Red lines are the fit of the power transform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

sets I, II, and IV were relative fitness), while we expected some to be additive (e.g., data set IV is a free energy). Rather than rescaling the multiplicative data sets 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 the 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 (Figure 4 and Figure S1).

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

With estimated scales in hand, we linearized the maps using Equation 2 and remeasured epistasis (Figure 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 (Figure 5A and Figure 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 Figure 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.

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 ex-

ample, 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 (Figure 5B and Figure 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. Figure 6 and Figure 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 (Figure S5).

For the epistatic coefficients, the degree of correlation depended on the degree of nonlinearity in the data set. 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 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 $\sim 20\%$ (six of 27) of the epistatic coefficients flipped sign when nonlinearity was taken into account (Figure 6, III, bottom-right quadrant).

Overall, we found that low-order epistatic coefficients were more robust to the linear assumption than high-order

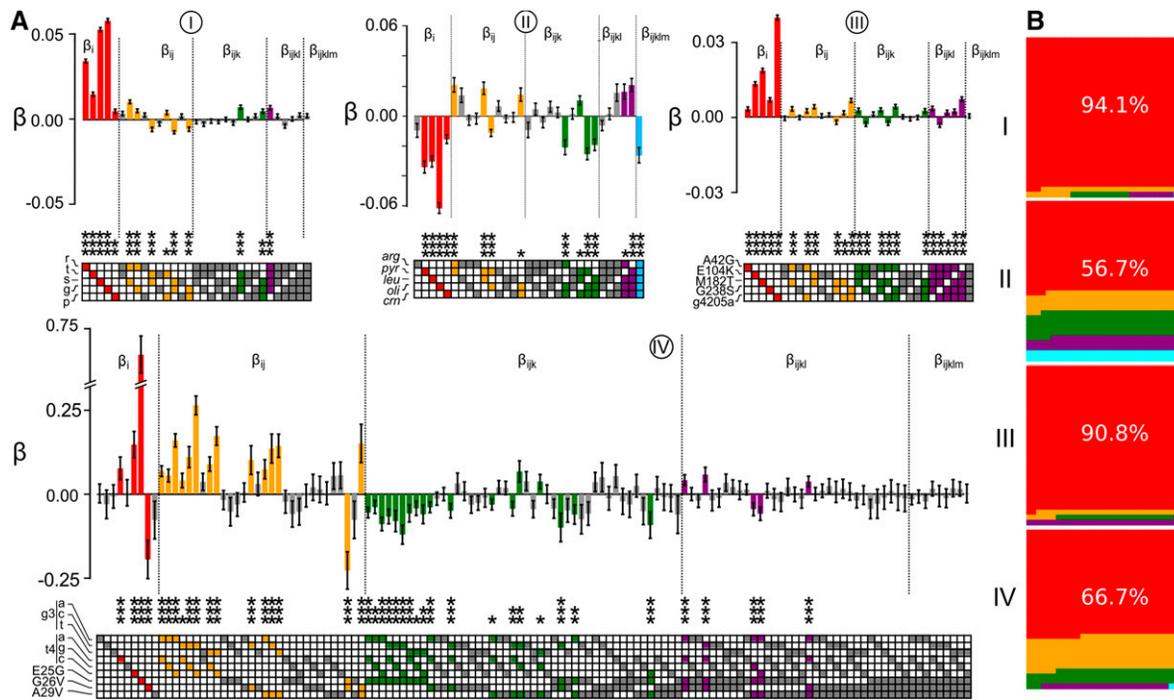


Figure 5 High-order epistasis is present in genotype-phenotype maps. (A) Shows epistatic coefficients extracted from data sets I–IV (Table 1; data set label circled above each graph). Bars denote coefficient magnitude and sign; error bars are propagated measurement uncertainty. Color denotes the order 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). * denote relative significance: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Filled ■ in the grid below the bars indicates 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) Subpanels show fraction of variation accounted for by first- through fifth-order epistatic coefficients for data sets I–IV [colors as in (A)]. Fraction described by each order is proportional to area.

coefficients. Data set IV is a clear example of this behavior. The map exhibited noticeable nonlinearity (Figure 4). The first- and second-order terms were well correlated between the linear and nonlinear analyses (Figure 6, Figure S4, and Figure 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 coefficients. 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 (Otto and Feldman 1997; MacLean *et al.* 2010; Chou *et al.* 2011; Tokuriki *et al.* 2012; Kryazhimskiy *et al.* 2014). Further, it is intuitive that optimizing a phenotype becomes more 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, Figure 4), this may be 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 (Figure S6). This result demonstrates the effectiveness of our approach in extracting nonlinearity in the genotype-phenotype map.

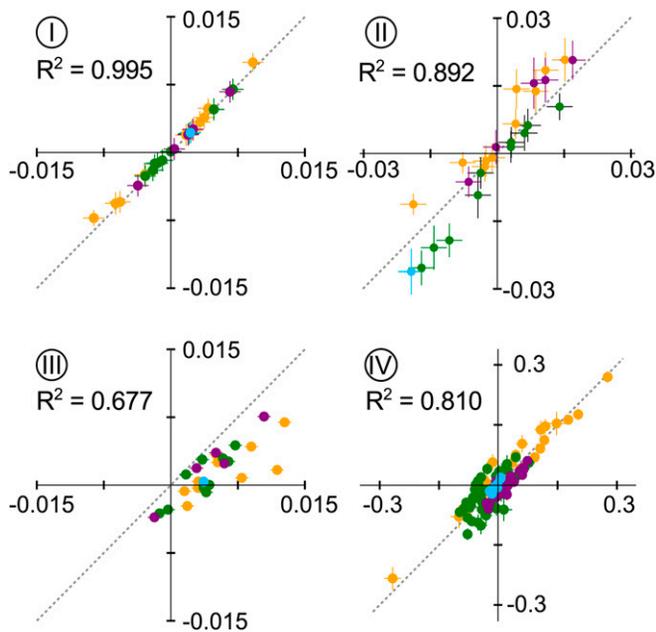


Figure 6 Nonlinear phenotypes distort measured epistatic coefficients. Subpanels show correlation plots between epistatic coefficients extracted without accounting for nonlinearity (x-axis) and accounting for linearity (y-axis) for data sets I–IV. Each point is an epistatic coefficient, colored by order.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

The origins of the growth in the transcription factor/DNA binding data set are less clear (IV, Figure 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 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://github.com/harmslab/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 nonmonotonic 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 vs. “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 L th-order polynomial, which would capture all of the variation as scale and none as epistasis. Scale estimation should be 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 nonadditivity 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 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 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, and 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 vs. specific interactions that lead to specific historical contingencies (Harms and Thornton 2014; Kryazhimskiy *et al.* 2014; Shah *et al.* 2015). 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 nonadditivity, and thus these evolutionary possibilities. Prevailing-magnitude epistasis (de Visser *et al.* 2009), global epistasis (Kryazhimskiy *et al.* 2014), and diminishing-returns epistasis (Otto and Feldman 1997; MacLean *et al.* 2010; Chou *et al.* 2011; 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 data sets 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 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 thank Patrick Phillips, Jamie Bridgham, and members of the Harms laboratory 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 (Z.R.S). M.J.H. is a Pew Scholar in the Biomedical Sciences, supported by The Pew Charitable Trusts.

Literature Cited

- Abdi, H., 2007 The Bonferonni and sidak corrections for multiple comparisons, pp. 103–107 in *Encyclopedia of Measurement and Statistics*, Vol. III. Sage, Thousands Oaks, CA.
- Anderson, D. W., A. N. McKeown, and J. W. Thornton, 2015 Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding site. *eLife*. 4: e07864.
- Bedau, M. A., and N. H. Packard, 2003 Evolution of evolvability via adaptation of mutation rates. *Biosystems* 69: 143–162.
- Blanquart, F., 2014 Properties of selected mutations and genotypic landscapes under Fisher's geometric model. *Evolution* 68: 3537–3554.
- Box, G. E. P., and D. R. Cox, 1964 An analysis of transformations. *J. R. Stat. Soc. Ser. B (Methodol.)* 26: 211–252.
- Carroll, R. J., and D. Ruppert, 1981 On prediction and the power transformation family. *Biometrika* 68: 609–615.
- Chou, H.-H., H.-C. Chiu, N. F. Delaney, D. Segrè, and C. J. Marx, 2011 Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332: 1190–1192.
- Cordell, H. J., 2002 Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* 11: 2463–2468.
- da Silva, J., M. Coetzer, R. Nedellec, C. Pastore, and D. E. Mosier, 2010 Fitness epistasis and constraints on adaptation in a human immunodeficiency virus type 1 protein region. *Genetics* 185: 293–303.
- Desai, M. M., 2009 Reverse evolution and evolutionary memory. *Nat. Genet.* 41: 142–143.
- de Visser, J. A. G. M., S.-C. Park, and J. Krug, 2009 Exploring the effect of sex on empirical fitness landscapes. *Am. Nat.* 174: S15–S30.
- Fisher, R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Philos. Trans. R. Soc. Edinburgh* 52: 399–433.
- Frankel, W. N., and N. J. Schork, 1996 Who's afraid of epistasis? *Nat. Genet.* 14: 371–373.
- Hall, D. W., M. Agan, and S. C. Pope, 2010 Fitness epistasis among 6 biosynthetic loci in the budding yeast *Saccharomyces cerevisiae*. *J. Hered.* 101: S75–S84.
- Harms, M. J., and J. W. Thornton, 2014 Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512: 203–207.
- Heckendorn, R. B., and D. Whitley, 1999 Predicting epistasis from mathematical models. *Evol. Comput.* 7: 69–101.
- Hu, T., N. A. Sinnott-Armstrong, J. W. Kiralis, A. S. Andrew, M. R. Karagas *et al.*, 2011 Characterizing genetic interactions in human disease association studies using statistical epistasis networks. *BMC Bioinformatics* 12: 364.
- Hu, T., Y. Chen, J. W. Kiralis, R. L. Collins, C. Wejse *et al.*, 2013 An information-gain approach to detecting three-way epistatic interactions in genetic association studies. *J. Am. Med. Inform. Assoc.* 20: 630–636.

- Hunter, J. D., 2007 Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9: 90–95.
- Imielinski, M., and C. Belta, 2008 Exploiting the pathway structure of metabolism to reveal high-order epistasis. *BMC Syst. Biol.* 2: 40.
- Khan, A. I., D. M. Dinh, D. Schneider, R. E. Lenski, and T. F. Cooper, 2011 Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332: 1193–1196.
- Kryazhimskiy, S., D. P. Rice, E. R. Jerison, and M. M. Desai, 2014 Microbial evolution. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* 344: 1519–1522.
- Lehár, J., A. Krueger, G. Zimmermann, and A. Borisy, 2008 High-order combination effects and biological robustness. *Mol. Syst. Biol.* 4: 215.
- MacLean, R. C., G. G. Perron, and A. Gardner, 2010 Diminishing returns from beneficial mutations and pervasive epistasis shape the fitness landscape for rifampicin resistance in *Pseudomonas aeruginosa*. *Genetics* 186: 1345–1354.
- Matsuura, T., Y. Kazuta, T. Aita, J. Adachi, and T. Yomo, 2009 Quantifying epistatic interactions among the components constituting the protein translation system. *Mol. Syst. Biol.* 5: 297.
- McKeown, A. N., J. T. Bridgham, D. W. Anderson, M. N. Murphy, E. A. Ortlund *et al.*, 2014 Evolution of DNA specificity in a transcription factor family produced a new gene regulatory module. *Cell* 159: 58–68.
- Neidhart, J., I. G. Szendro, and J. Krug, 2013 Exact results for amplitude spectra of fitness landscapes. *J. Theor. Biol.* 332: 218–227.
- Otto, S. P., and M. W. Feldman, 1997 Deleterious mutations, variable epistatic interactions, and the evolution of recombination. *Theor. Popul. Biol.* 51: 134–147.
- Otwinowski, J., and J. B. Plotkin, 2014 Inferring fitness landscapes by regression produces biased estimates of epistasis. *Proc. Natl. Acad. Sci. USA* 111: E2301–E2309.
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion *et al.*, 2011 Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12: 2825–2830.
- Perez, F., and B. E. Granger, 2007 IPython: a system for interactive scientific computing. *Comput. Sci. Eng.* 9: 21–29.
- Pettersson, M., F. Besnier, P. B. Siegel, and Ö. Carlborg, 2011 Replication and explorations of high-order epistasis using a large advanced intercross line pedigree. *PLoS Genet.* 7: e1002180.
- Phillips, P. C., 2008 Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9: 855–867.
- Poelwijk, F. J., V. Krishna, and R. Ranganathan, 2016 The context-dependence of mutations: a linkage of formalisms. *PLoS Comput. Biol.* 12: e1004771.
- Ritchie, M. D., L. W. Hahn, N. Roodi, L. R. Bailey, W. D. Dupont *et al.*, 2001 Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* 69: 138–147.
- Rokyta, D. R., P. Joyce, S. B. Caudle, C. Miller, C. J. Beisel *et al.*, 2011 Epistasis between beneficial mutations and the phenotype-to-fitness map for a ssDNA virus. *PLoS Genet.* 7: e1002075.
- Rothman, K. J., S. Greenland, and A. M. Walker, 1980 Concepts of interaction. *Am. J. Epidemiol.* 112: 467–470.
- Schenk, M. F., I. G. Szendro, M. L. M. Salverda, J. Krug, and J. A. G. M. de Visser, 2013 Patterns of epistasis between beneficial mutations in an antibiotic resistance gene. *Mol. Biol. Evol.* 30: 1779–1787.
- Segrè, D., A. DeLuna, G. M. Church, and R. Kishony, 2005 Modular epistasis in yeast metabolism. *Nat. Genet.* 37: 77–83.
- Shah, P., D. M. McCandlish, and J. B. Plotkin, 2015 Contingency and entrenchment in protein evolution under purifying selection. *Proc. Natl. Acad. Sci. USA* 112: E3226–E3235.
- Sun, J., F. Song, J. Wang, G. Han, Z. Bai *et al.*, 2014 Hidden risk genes with high-order intragenic epistasis in Alzheimer's disease. *J. Alzheimers Dis.* 41: 1039–1056.
- Szendro, I. G., M. F. Schenk, J. Franke, J. Krug, and J. A. G. M. de Visser, 2013 Quantitative analyses of empirical fitness landscapes. *J. Stat. Mech.* 2013: P01005.
- Taylor, M. B., and I. M. Ehrenreich, 2015 Higher-order genetic interactions and their contribution to complex traits. *Trends Genet.* 31: 34–40.
- Tokuriki, N., C. J. Jackson, L. Afriat-Jurnou, K. T. Wyganowski, R. Tang *et al.*, 2012 Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat. Commun.* 3: 1257.
- Tsai, C.-T., J.-J. Hwang, M. D. Ritchie, J. H. Moore, F.-T. Chiang *et al.*, 2007 Renin–angiotensin system gene polymorphisms and coronary artery disease in a large angiographic cohort: detection of high order gene–gene interaction. *Atherosclerosis* 195: 172–180.
- van der Walt, S., S. C. Colbert, and G. Varoquaux, 2011 The NumPy array: a structure for efficient numerical computation. *Comput. Sci. Eng.* 13: 22–30.
- Wang, Y., C. D. Arenas, D. M. Stoebel, and T. F. Cooper, 2012 Genetic background affects epistatic interactions between two beneficial mutations. *Biol. Lett.* 9: 20120328.
- Weinreich, D. M., N. F. Delaney, M. A. DePristo, and D. L. Hartl, 2006 Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312: 111–114.
- Weinreich, D. M., Y. Lan, C. S. Wylie, and R. B. Heckendorn, 2013 Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* 23: 700–707.
- Xu, J., J. Lowey, F. Wiklund, J. Sun, F. Lindmark *et al.*, 2005 The interaction of four genes in the inflammation pathway significantly predicts prostate cancer risk. *Cancer Epidemiol. Biomarkers Prev.* 14: 2563–2568.
- Yokoyama, S., A. Altun, H. Jia, H. Yang, T. Koyama *et al.*, 2015 Adaptive evolutionary paths from UV reception to sensing violet light by epistatic interactions. *Sci. Adv.* 1: e1500162.
- Zhang, C.-T., and R. Zhang, 1991 Analysis of distribution of bases in the coding sequences by a digrammatic technique. *Nucl. Acids Res.* 19: 6313–6317.

Communicating editor: C. Kendziorski