



Resolving discrepancies between chimeric and multiplicative measures of higher-order epistasis

Received: 23 May 2024

Accepted: 6 February 2025

Published online: 17 February 2025

 Check for updatesUthsav Chitra^{1,3}, Brian Arnold^{1,2,3} & Benjamin J. Raphael¹  

Epistasis - the interaction between alleles at different genetic loci - plays a fundamental role in biology. However, several recent approaches quantify epistasis using a chimeric formula that measures deviations from a multiplicative fitness model on an additive scale, thus mixing two scales. Here, we show that for pairwise interactions, the chimeric formula yields a different magnitude but the same sign of epistasis compared to the multiplicative formula that measures both fitness and deviations on a multiplicative scale. However, for higher-order interactions, we show that the chimeric formula can have both different magnitude and sign compared to the multiplicative formula. We resolve these inconsistencies by deriving mathematical relationships between the different epistasis formulae and different parametrizations of the multivariate Bernoulli distribution. We argue that the chimeric formula does not appropriately model interactions between the Bernoulli random variables. In simulations, we show that the chimeric formula is less accurate than the classical multiplicative/additive epistasis formulae and may falsely detect higher-order epistasis. Analyzing multi-gene knockouts in yeast, multi-way drug interactions in *E. coli*, and deep mutational scanning of several proteins, we find that approximately 10% to 60% of inferred higher-order interactions change sign using the multiplicative/additive formula compared to the chimeric formula.

A key problem in biology is deriving the map from genotype to fitness, or the average reproductive success of a genotype. This map is often referred to as the *fitness landscape* (first conceptualized in ref. 1). In its simplest form, the fitness effects of alleles at one locus are independent of those at other loci, such that multilocus fitness is either an additive or multiplicative function of the alleles across loci. However, the fitness landscape is complicated by the presence of *epistasis*, or genetic interactions where alleles at one locus modify the effects of alleles at other loci. Epistatic interactions reveal functional relationships between genes, with the sign of an epistatic interaction (positive or negative) often used to understand how genes are organized into genetic pathways², model protein function, and evolution³,

understand mechanisms of antibiotic resistance⁴, and interpret genome-wide association studies (GWAS)⁵.

While epistasis is a property of any quantitative trait, many studies have measured epistatic interactions using experimental fitness data from haploid genomes (reviewed in refs. 6–8). Most of these studies measure *pairwise epistasis*, or an interaction between a pair of genetic loci that is computed by comparing the observed fitness of the double-mutant to the expected fitness under a null model with no epistasis. The *sign* of the epistatic interaction is determined by whether the observed fitness is greater than or less than the expected fitness, resulting in a positive or negative interaction, respectively. The choice of the null model for the expected fitness depends on the quantitative

¹Department of Computer Science, Princeton University, Princeton, NJ, USA. ²Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA. ³These authors contributed equally: Uthsav Chitra, Brian Arnold. ✉ e-mail: braphael@princeton.edu

trait used as a proxy for fitness. Nearly all formulae for pairwise epistasis assume either an additive null model, where the expected fitness is the sum $f_{01} + f_{10}$ of the fitness values of the single-mutants, or a multiplicative null model, where the expected fitness is the product $f_{01}f_{10}$ of the fitness values of the single-mutants. For example, an additive null model is often used when fitness is measured using cellular growth rate (e.g., in fitness assays of microbes⁹) or fluorescence (e.g., in proteins¹⁰), while a multiplicative null model is typically used when fitness is measured by the total size of a clonal population¹¹. Under an additive null model, epistasis ϵ is computed as the difference $\epsilon = f_{11} - (f_{10} + f_{01})$ between observed and expected double-mutant fitness.

For the multiplicative null model, there is no agreement in the literature about how to quantify deviation from the null model. In the statistics literature, it is standard to compute multiplicative interaction effects using a *ratio* $\epsilon = \frac{f_{11}}{f_{01}f_{10}}$ between the observed and expected values (e.g., refs. 12,13). On the other hand, many studies in the genetics literature compute epistasis as the *difference* $\epsilon = f_{11} - f_{01}f_{10}$ between observed and expected fitness values of a double-mutant (e.g., refs. 14–23). We call the first formula the *multiplicative formula*, as it preserves the multiplicative measurement scale, while we call the second formula the *chimeric formula*, as it measures deviations from a multiplicative model on an additive scale and thus is a “chimera” of additive and multiplicative scales.

Here, we show that the chimeric and multiplicative formula result in different quantitative measures of pairwise epistasis, which may affect findings on the strength of an epistatic interaction. Nevertheless, we also show that the two formulae always yield the same *sign* (or *direction*) of a pairwise interaction. The sign is often the quantity of interest in genetics studies, e.g., negative epistatic interactions are used to quantify functional redundancy^{24,25}. Thus, the focus of existing literature on the sign of interactions, as well as the focus on pairwise epistasis, may explain why the differences between the multiplicative and chimeric formula are not broadly recognized.

The discrepancies between the multiplicative and chimeric formula are more consequential for *higher-order* interactions between three or more loci, which are becoming more widely studied with larger genetic datasets and high-throughput measurements of fitness^{24,26–28}. Recent studies in yeast genetics^{24,26} and antibiotic resistance²⁷ independently derived analogous chimeric formula to quantify epistasis between three or more loci and higher-order interactions between components, respectively, under a multiplicative fitness model. These chimeric formulae were derived *de novo* and without consideration of the two distinct formula – chimeric and multiplicative – for pairwise epistasis, nor the consequences of conflating multiplicative and additive scales. However, unlike in the pairwise setting, we show that for three or more loci, the chimeric formula is *not* guaranteed to produce the same *sign* of an interaction as the multiplicative formula. Thus, the chimeric formula may indicate a *positive* epistatic interaction while the multiplicative formula shows a *negative* epistatic interaction, and vice-versa. Such inconsistencies raise questions about the validity of reported higher-order epistasis in biological applications.

We resolve the mathematical and biological inconsistencies between the different epistasis formulae by deriving connections between epistasis and the parameters of the multivariate Bernoulli distribution (MVB), a probability distribution on binary random variables²⁹. In particular, we show that a wide array of approaches for quantifying epistasis – including the additive, multiplicative, and chimeric formulae, as well as the regression models commonly used in GWAS and QTL analyses^{2,5} and the Walsh coefficients for measuring background-averaged epistasis^{30–32} – are equivalent to computing different parameterizations of the MVB, showing that the MVB provides a unifying statistical framework for the different epistasis measures.

We use the connections to the multivariate Bernoulli distribution to analyze the higher-order (i.e., ≥ 3 -way interactions) chimeric epistasis formulae derived by Kuzmin et al.^{24,26} and Tekin et al.²⁷. We show that the chimeric formulae for pairwise epistasis and the chimeric formulae for higher-order epistasis correspond to the joint cumulants of the MVB, a concept from probability theory for measuring interactions between continuous variables³³. However, the joint cumulant is known to *not* be an appropriate measure of higher-order interactions for binary random variables^{34,35}. Accordingly, we argue that the chimeric epistasis formula are not appropriate for measuring higher-order epistasis between biallelic mutations. In this way, just like how the hero Bellerophon in the *Iliad* slayed the monstrous chimera, the multivariate Bernoulli distribution allows us to “slay” the chimeric epistasis formula.

We demonstrate that the mathematical issues with the chimeric epistasis formula lead to markedly different biological interpretations of perturbation experiments using haploid genomes. Analyzing multi-gene knockout data in yeast using the more appropriate multiplicative formula changes the sign of 12% of the 7957 trigenic interactions that Kuzmin et al.^{24,26} reported using the chimeric formula. Many of these sign changes are concentrated on negative interactions, which are more functionally informative than positive interactions and are commonly used to measure functional redundancy between genes²⁵. In particular, the multiplicative epistasis formula identifies nearly 500 negative interactions not reported by Kuzmin et al.^{24,26} that are significantly enriched for several measures of functional redundancy, thus extending the trigenic interaction network by 25%.

We further demonstrate that the multiplicative and additive formulae yield markedly different interactions compared to the chimeric formula in two other applications: the identification of higher-order synergistic and antagonistic drug interactions in *Escherichia coli* and the identification of epistatic interactions between protein mutations in deep mutational scanning (DMS) experiments which is important for 3-D protein structure prediction^{36,37}, protein engineering⁷, genome editing optimization³⁸, variant effect prediction³⁹, and other applications. Notably, we show that the discordance between the different formulae increases with interaction order: the additive formula shows significantly less antagonism between five-way interactions compared to the chimeric formula used in ref. 40, while for some proteins there is substantial (up to 60%) disagreement in the sign of interaction between the multiplicative and chimeric formulae.

Results

Pairwise epistasis: additive, multiplicative, and chimeric

Pairwise epistasis describes interactions between two genetic loci. We consider haploid genomes and assume that each locus is *biallelic*, i.e., each locus has two alleles labeled 0 and 1. Thus for a pair of loci there are four possible genotypes: the wild-type 00, the single mutants 01 and 10, and the double mutant 11. Accordingly, for a pair (i, j) of loci there are four corresponding fitness values: the wild-type fitness f_{\emptyset} , corresponding to the wild-type genotype 00 with no mutations; the single-mutant fitnesses f_i, f_j , corresponding to the genotypes 01 and 10 with either locus i or locus j mutated, respectively; and the double-mutant fitness f_{ij} , corresponding to the genotype 11 with both loci mutated. Pairwise epistasis is measured by comparing the observed double-mutant fitness f_{ij} to the expected fitness under a null model with no epistasis.

In practice, the fitness f of a genotype cannot be directly measured. Instead, experiments typically measure traits that are expected to be highly correlated with fitness. For example, in populations of microbes or proteins, it is standard to estimate the fitness of a genotype with either its population growth rate or relative frequency, which are typically assumed to follow an additive or multiplicative scale, respectively (“Methods”). Accordingly, the two standard null models of

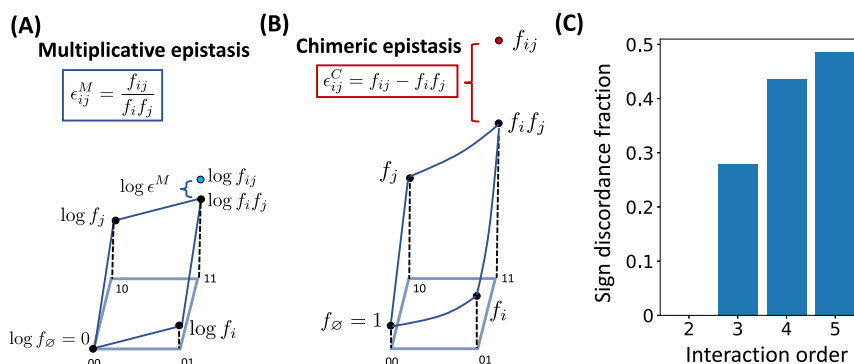


Fig. 1 | Comparison of multiplicative and chimeric epistasis measures. **A** For a pair (i, j) of loci, the multiplicative epistasis measure $e_{ij}^M = \frac{f_{ij}}{f_i f_j}$ is the ratio between the observed fitness f_{ij} of the double mutant and the expected fitness $f_i f_j$ under a multiplicative null model. Equivalently, the logarithm $\log e_{ij}^M$ of the epistasis measure is given by $\log e_{ij}^M = \log f_{ij} - \log f_i f_j$, or the difference between the observed and expected values in log-fitness space. **B** The chimeric epistasis measure $e_{ij}^C = f_{ij} - f_i f_j$ is the difference between the observed and expected fitness values of the double-mutant under a multiplicative fitness model. The chimeric measure e_{ij}^C thus

mixes scales by measuring deviations from multiplicativity on an additive scale. **C** The fraction of instances where the signs $\text{sgn}(\log e^M)$ and $\text{sgn}(e^C)$ of the multiplicative and chimeric fitness formula, respectively, disagree ("sign discordance fraction") for interaction orders $L = 2, \dots, 5$, where fitness values f_i, f_j, \dots are sampled uniformly at random from the interval $[0, 1]$. For two loci, the sign of the two measures always agree (see Proposition 1), but for more than two loci, there is substantial disagreement.

fitness for measuring epistasis are the *additive* model and the *multiplicative* model.

In the *additive* model, mutations are assumed to have an additive effect on fitness, and the pairwise epistasis measure e_{ij}^A is equal to the difference between the observed and expected double-mutant fitness values:

$$e_{ij}^A = f_{ij} - (f_i + f_j), \quad (1)$$

under the assumption that fitness values are normalized such that the wild-type fitness $f_\emptyset = 0$. The sign of the interaction (i.e., positive vs. negative) is given by the sign $\text{sgn}(e_{ij}^A)$ of the epistasis measure e_{ij}^A . The additive model was first posed by Fisher⁴¹, who used the term "epistacy" to refer to any statistical deviation from additivity⁶.

In the *multiplicative* fitness model, mutations have a multiplicative effect on fitness, and the multiplicative pairwise epistasis measure e_{ij}^M is given by the ratio between the observed and expected double-mutant fitness values:

$$e_{ij}^M = \frac{f_{ij}}{f_i f_j}, \quad (2)$$

under the typical assumption that the wild-type fitness f_\emptyset is equal to 1. The sign of the interaction is determined by whether the multiplicative measure e_{ij}^M is greater than or less than 1. Moreover, if fitness values f are multiplicative, then the log-fitness values $\log f$ are additive; thus, the sign of interaction under the multiplicative model is also given by the sign $\text{sgn}(\log e_{ij}^M)$ of the log-epistasis measure $\log e_{ij}^M$. The additive and multiplicative epistasis measures are closely related to the linear/log-linear regression frameworks^{5,12} and the Walsh coefficients^{30–32,42,43} used in the genetics literature; see Methods for details.

Curiously, there is a third epistasis formula that is widely used for the multiplicative fitness model. Here, deviations from the multiplicative model are measured on an additive scale, resulting in the following *chimeric* formula for pairwise epistasis:

$$e_{ij}^C = f_{ij} - f_i f_j. \quad (3)$$

We refer to e_{ij}^C as the chimeric epistasis measure because it measures deviations from a multiplicative null model on an additive scale and is thus a *chimera* of both the multiplicative and additive measurement scales. As in the additive model, the sign of the interaction is given by the sign $\text{sgn}(e_{ij}^C)$ of the chimeric measure e_{ij}^C .

The chimeric epistasis measure e_{ij}^C appears in the genetics literature (e.g., refs. 14–23) and in the drug interaction literature (e.g., refs. 27,40,44,45) because of its interpretation as a *residual*, i.e., the difference between the observed and expected values of a measurement. However, despite the simplicity of this explanation, residuals are typically only appropriate for additive models. For multiplicative models, it is standard to compute statistical interactions using the *ratio* between observed and expected measurements (as in equation (2)), rather than the difference¹². Moreover, Wagner^{13,46} notes that preserving the multiplicative measurement scale (by using the ratio) is required in order to guarantee meaningful notions of statistical and functional interactions.

While both the chimeric measure e_{ij}^C and the multiplicative measure e_{ij}^M are described as measuring deviations from a multiplicative fitness model, the two measures are not equal. In particular, the (log-) multiplicative epistasis measure $\log e_{ij}^M = \log f_{ij} - \log f_i f_j$ computes the difference between the observed and expected double-mutant fitness values on a logarithmic scale (Fig. 1A) while the chimeric epistasis measure $e_{ij}^C = f_{ij} - f_i f_j$ computes the difference directly (Fig. 1B). When the double-mutant fitness f_{ij} and single-mutant fitness values f_i, f_j are close to 1, we show that the chimeric measure e_{ij}^C is approximately equal to the log-multiplicative measure $\log e_{ij}^M$ (Supplementary Note 1). However, if the fitness values are substantially different from 1, then the chimeric epistasis measure e_{ij}^C may over- or under-state the strength of a pairwise interaction in a multiplicative fitness model as we demonstrate numerically (Supplementary Note 2).

Nevertheless, we prove ("Methods") that the chimeric measure e_{ij}^C has the same *sign* of an interaction as the multiplicative measure e_{ij}^M but not the same *magnitude*. Thus, using either the chimeric or multiplicative measures will not affect findings that depend on the *sign* of an epistatic interaction, and the sign is often the quantity of interest (e.g., negative epistasis is used to quantify functional redundancy²¹). However, the agreement between the multiplicative and chimeric measures on the sign of interaction is true only for pairwise epistasis and not higher-order epistasis, as we will show below.

Higher-order epistasis

For higher-order epistasis, or interactions between three or more genetic loci, we find that the difference between the multiplicative and chimeric epistasis measures are more consequential. Under the multiplicative fitness model, the three-way epistasis measure e_{ijk}^M between loci i, j, k is given by the ratio between observed and expected triple-

Table 1 | Correspondence between epistasis measures and the parametrizations of the multivariate Bernoulli distribution when the fitness values are proportional to the indicated quantities of the distribution

	Parameters of multivariate Bernoulli distribution	Fitness values f proportional to
Additive epistasis measure ϵ^A	Natural parameters β	Log-probabilities $\log \mathbf{p}$
Multiplicative epistasis measure ϵ^M	Natural parameters β	Probabilities \mathbf{p}
Chimeric epistasis measure ϵ^C	Joint cumulants κ	Moments μ
Walsh coefficients	Moments of $(1 - 2X_1, \dots, 1 - 2X_L)$	Probabilities \mathbf{p}

mutant fitness:

$$\epsilon_{ijk}^M = \frac{f_{ijk}}{f_i f_j f_k \epsilon_{ij}^M \epsilon_{ik}^M \epsilon_{jk}^M} = \frac{f_{ijk} f_i f_j f_k}{f_{ij} f_{ik} f_{jk}}. \quad (4)$$

Recent work in the yeast genetics^{24,26} and antibiotic resistance²⁷ literature claim to use a multiplicative fitness model, but derive a different epistasis formula:

$$\epsilon_{ijk}^C = f_{ijk} - (f_i f_j f_k + \epsilon_{ij}^C f_k + \epsilon_{ik}^C f_j + \epsilon_{jk}^C f_i), \quad (5)$$

where $\epsilon_{ij}^C, \epsilon_{ik}^C, \epsilon_{jk}^C$ are the pairwise chimeric epistasis measures in (3). Note that as in the pairwise case, formula (5) mixes the additive and multiplicative scales in a complex manner. Thus, we refer to ϵ_{ijk}^C as the *chimeric* three-way epistasis measure.

As in the pairwise setting, the three-way chimeric measure ϵ_{ijk}^C in (5) is clearly different from the three-way multiplicative measure ϵ_{ijk}^M in (4). However, we show that these formula often differ in both the magnitude of epistasis (as in the pairwise setting) and in the sign of epistasis. Thus, one formula may indicate positive epistasis between three loci while another formula may indicate negative epistasis, and vice-versa. In simulations, we find that approximately 28% of triples have different signs between the two formulae (Fig. 1C).

Tekin et al.²⁷ extended the three-way chimeric epistasis formula (5) to compute a 4-way chimeric epistasis measure ϵ_{ijkl}^C and a 5-way chimeric epistasis measure ϵ_{ijklm}^C . We find even more substantial differences in the sign of epistasis between these 4-way and 5-way chimeric epistasis measures and the 4-way and 5-way multiplicative epistasis measures (Equation (18) in “Methods”). In simulations, only approximately 57% and 52% of 4-way and 5-way interactions, respectively, have the same sign using the chimeric and multiplicative epistasis formulae (Fig. 1C).

This substantial disagreement between the chimeric and multiplicative epistasis measure motivates a deeper mathematical understanding of the various epistasis formulae, which we undertake in the next section.

Unifying epistasis measurements with the multivariate Bernoulli distribution

A genotype of biallelic mutations on L loci can be represented as a binary string of length L , where 0 corresponds to the wild-type allele, and 1 corresponds to the mutant, or derived, allele. For example, the string 01100 represents the genotype of $L = 5$ loci with mutations in the second and third loci. The fitness values of all genotypes, often referred to as the fitness landscape, correspond to a function f that maps a binary string $\mathbf{x} \in \{0, 1\}^L$ to its fitness $f_{\mathbf{x}}$.

A natural approach for studying a fitness landscape function f is to view it as a *distribution* on the set $\{0, 1\}^L$ of binary strings, where the probability $p_{\mathbf{x}}$ of a binary string \mathbf{x} is derived from its fitness $f_{\mathbf{x}}$. Such distributions are often used by protein structure models³⁹. Moreover, many real-world fitness datasets – including the yeast fitness data and many of the protein datasets analyzed in this manuscript – measure

the fitness of a genotype \mathbf{x} in terms of its relative frequency in a large population of genotypes, i.e., its probability $p_{\mathbf{x}}$.

Here, we model the fitness landscape using the *multivariate Bernoulli* (MVB) distribution^{29,47} which describes *any* distribution on the set $\{0, 1\}^L$ of binary strings. Formally, a multivariate random variable (X_1, \dots, X_L) distributed according to a MVB is parametrized by the probabilities $p_{\mathbf{x}} = P((X_1, \dots, X_L) = \mathbf{x})$ for each binary string $\mathbf{x} = (x_1, \dots, x_L) \in \{0, 1\}^L$. We model the genotype (X_1, \dots, X_L) of an organism as a random variable distributed according to a MVB parametrized by the probabilities $\mathbf{p} = (p_{\mathbf{x}})_{\mathbf{x} \in \{0, 1\}^L}$.

We prove that the additive, multiplicative, and chimeric measures of epistasis – as well as the Walsh coefficients described in refs. 30–32,42,43 – correspond to different parametrizations of the MVB distribution (Table 1, “Methods”). We briefly describe these results below.

Multiplicative and additive epistasis. Suppose the fitness values $f_{\mathbf{x}} \in \mathbb{R}$ of each genotype $\mathbf{x} = (x_1, \dots, x_L) \in \{0, 1\}^L$ are proportional to the corresponding probability $p_{\mathbf{x}}$ of a multivariate Bernoulli random variable (X_1, \dots, X_L) , i.e., $f_{\mathbf{x}} = c \cdot p_{\mathbf{x}}$ for some $c > 0$. We prove that the (log-) multiplicative epistasis measures are equal to the *natural parameters* of the MVB. The natural parameters $\beta = \{\beta_S\}_{S \subseteq \{1, \dots, L\}}$ are another parameterization of the MVB that encodes conditional independence relations between the random variables X_1, \dots, X_L ; see refs. 29,48. We prove a similar result for the additive epistasis measure under the assumption that the fitness $f_{\mathbf{x}}$ is proportional to the log probability $\log p_{\mathbf{x}}$. See Methods and Supplementary Note 3 for theorem statements and proofs.

Our theoretical results provide a connection between the multiplicative epistasis measure and interaction coefficients in a log-linear regression model. This is because for each subset S of loci, the natural parameter β_S corresponds to the interaction term for the subset S in a log-linear regression model^{29,47,48}. Such interaction terms are a standard approach for measuring epistasis in genetics, e.g., GWAS or QTL analyses for quantitative traits^{2,5}.

We also prove that the natural parameters β of the MVB are closely related to the two standard approaches for measuring pairwise SNP-SNP interactions in a case-control GWAS: logistic regression and conditional independence testing⁴⁹. Specifically, we prove that the interaction term in a logistic regression is equal to a 3-way interaction term β_{ijk} in an MVB, while the conditional independence test is equivalent to testing whether a 2-way interaction term β_{ij} and a 3-way interaction term β_{ijk} are both equal to zero. These interaction terms are equal to the corresponding log-multiplicative epistasis measures $\log \epsilon^M$.

Thus, our results show that the additive and multiplicative epistasis measures are equivalent to computing interaction terms in regression models commonly used in genetics.

Chimeric epistasis. The connection between the epistasis formulae and the MVB distribution allows us to derive a mathematically rigorous definition of the chimeric epistasis formula. Specifically, suppose the fitness value $f_{\mathbf{x}}$ of each genotype $\mathbf{x} = (x_1, \dots, x_L)$ is equal to a corresponding *moment* $E[X_1^{x_1} \dots X_L^{x_L}]$ of the random variable (X_1, \dots, X_L) . Then, we define the chimeric epistatic measure $\epsilon_{i_1 \dots i_K}^C$ as the K -th order

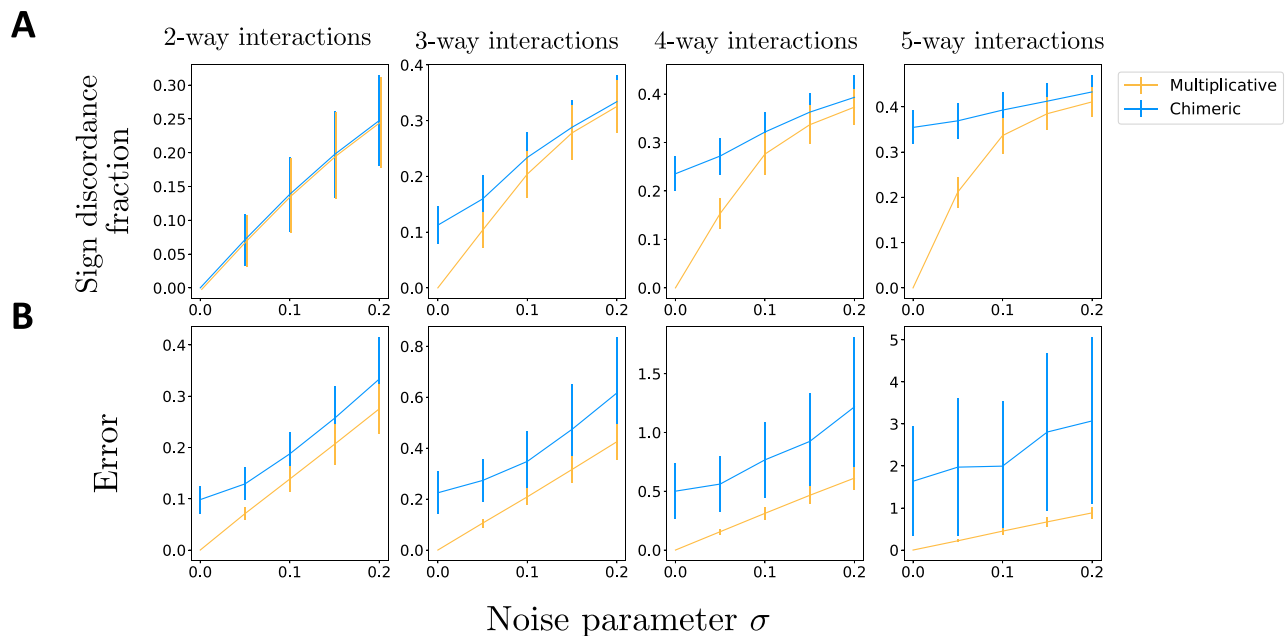


Fig. 2 | Comparison of epistasis measures using simulated data from a multiplicative fitness model. Fitness values \mathbf{f} are simulated following a multiplicative fitness model with interaction parameters β , for different choices of the maximum interaction order K , and multiplicative Gaussian noise with standard deviation σ . **A** The fraction of K -way interactions where the sign of the log-multiplicative epistasis measure $\log \epsilon^M$ (orange) and the chimeric epistasis measure ϵ^C (blue) do not match the sign of the true interaction parameter β . **B** The average absolute

difference ("error") $|\beta - \log \epsilon^M|$ and $|\beta - \epsilon^C|$ between the true interaction parameter β and (orange) the log-multiplicative measure $\log \epsilon^M$ and (blue) the chimeric measure ϵ^C , respectively. These quantities are computed for different values of the maximum interaction order K and noise parameter σ and are averaged across 200 simulated fitness values. Error bars indicate standard deviation across simulated instances.

joint cumulant $\kappa(X_{i_1}, \dots, X_{i_k})$ of the random variables X_{i_1}, \dots, X_{i_k} (Table 1). Joint cumulants are a concept from probability theory that are used to quantify higher-order interactions between random variables³³. See Methods for a formal definition.

We emphasize that prior literature on higher-order interactions do not provide a rigorous statistical interpretation of the chimeric epistasis measure. For example, Kuzmin et al.^{24,26} does not explicitly state the connection between the joint cumulant and their three-way chimeric formula, while Tekin et al.²⁷ heuristically uses the joint cumulant formulae without specifying random variables or a probability distribution – thus obscuring any assumptions made by using joint cumulants to measure higher-order interactions.

Our explicit definition of the K -th order chimeric epistasis measure $\epsilon_{i_1 \dots i_K}^C$ as the K -th order joint cumulant reveals two critical issues with the chimeric formula. First, the assumption that the fitness values f are equivalent to the moments of an MVB random variable is not biologically reasonable for higher-order interactions between three or more loci. This is because the moments assumption implies that the fitness of a particular genotype depends on the probability of many other genotypes. For example, if we assume that the fitness values for $L = 4$ loci are moments of the MVB, then the fitness f_{1100} of a double mutant is equal to the moment $E[X_1 X_2]$, which is equal to

$$E[X_1 X_2] = P(X_1 = 1, X_2 = 1) = p_{1100} + p_{1101} + p_{1110} + p_{1111}. \quad (6)$$

However, it is not clear why the fitness f_{1100} of a *single* genotype, 1100, should equal the sum of the probabilities of *four different* genotypes, 1100, 1101, 1110, and 1111.

The second issue is that joint cumulants are not an appropriate measure of higher-order interactions between *binary* random variables. The differences between the joint cumulants κ and natural parameters β have been previously investigated in the neuroscience literature, as both quantities have been used to quantify higher-order interactions in neuronal data. For example, Staude et al.³⁴ write that the

joint cumulants κ and natural parameters β "do not measure the same kind of dependence. While higher-order cumulant correlations [κ] indicate additive common components ... the [natural parameters β] directly change the probabilities of certain patterns multiplicatively". In particular, the natural parameters β measure "to what extent the probability of certain binary patterns can be explained by the probabilities of its sub-patterns"³⁴. Thus, for biallelic genotype data, the natural parameters β correspond exactly with the epistasis we aim to measure, i.e., how the fitness of a binary pattern can be explained by the fitness of its "sub-patterns", while the joint cumulants κ do not.

Simulations using a multiplicative fitness model

We performed simulations to demonstrate the discrepancy between the multiplicative epistasis measure and the chimeric epistasis measure. Since both the multiplicative and chimeric measures use a multiplicative fitness model, we simulated fitness values \mathbf{f} for $L = 10$ loci following a multiplicative fitness model with K -way interactions β for different choices of interaction order K , and with multiplicative Gaussian noise with standard deviation σ (Methods). We computed the K -way multiplicative measure ϵ_S^M and chimeric measure ϵ_S^C for each set $S \subseteq \{1, \dots, L\}$ of loci of size $|S| = K$, and we compared these two measures to the true interaction measure β_S .

We first assessed whether the *sign* of the epistasis measures, i.e., $\text{sgn}(\log \epsilon_S^M)$ and $\text{sgn}(\epsilon_S^C)$, match the sign $\text{sgn}(\beta_S)$ of the corresponding interaction term β_S , since the sign of a measure indicates whether there is a positive or negative interaction between mutations in the loci S . We observed (Fig. 2A) that for pairwise interactions ($K = 2$), both the multiplicative measure ϵ^M and chimeric measure ϵ^C have the same sign as the true interaction measure β for the same fraction of instances, which matches our theoretical result (Proposition 1, Methods). However, for higher-order interactions ($K > 2$), the chimeric measure ϵ^C has an incorrect sign more often than the multiplicative measure ϵ^M (Fig. 2A). In particular, for $K = 5$ -way interactions, even with no noise (i.e., $\sigma = 0$), the chimeric measure has a different sign than the true

interaction parameter σ for more than 30% of simulated instances. We also highlight that when there is no noise, i.e., $\sigma = 0$, the multiplicative measure always has the same sign as the true interaction parameter β , i.e., $\text{sgn}(\log \epsilon^M) = \text{sgn}(\beta)$, which agrees with Theorem 1.

We next compared how well the *magnitudes* of the multiplicative and chimeric epistasis measures agree with the magnitude of the true interaction parameters. We computed the average absolute difference (“error”) $|\log \epsilon^M - \beta|$ and $|\epsilon^C - \beta|$ between the true interaction measure β and the estimated multiplicative and chimeric epistasis measures, respectively, for all subsets S of loci of size $|S|=K$. We found (Fig. 2B) that the multiplicative measure has a smaller error for all interaction orders K and noise parameters σ . In particular, we observe that the multiplicative measure has a smaller error than the chimeric measure even for pairwise interactions ($K=2$) – i.e., when both the multiplicative and chimeric measures have the same sign – and that the error of the chimeric measure ϵ^C increases with the interaction order K . The reason that the chimeric measure has much larger error than the multiplicative measure for pairwise interactions is that the chimeric measure ϵ^C_{ij} is approximately equal to the (log-)multiplicative measure *only when* $f_{ij} \approx 1$ and $f_i f_j \approx 1$, with the two measures being noticeably different otherwise (Supplementary Figs. 1 and 2 and Supplementary Note 1). We also emphasize that when there is no noise, i.e., $\sigma = 0$, the multiplicative measure has zero error, i.e., $\log \epsilon^M = \beta$, matching our theoretical results (Theorem 1, Methods). (Note that Theorem 1 does not apply when there is multiplicative Gaussian noise, i.e., $\sigma > 0$, as this noise will cause the fitness values to not follow a log-linear model.)

Thus, our results demonstrate that the multiplicative measure ϵ^M yields a more accurate measurement of pairwise and higher-order epistasis in a multiplicative fitness model compared to the chimeric measure ϵ^C which conflates additive and multiplicative factors.

Simulations using the NK fitness model

We next compared the multiplicative and chimeric epistasis measures using the NK model, a classical model for simulating random fitness landscapes \mathbf{f} with varying degrees of “ruggedness”⁵⁰. The NK model has two parameters: the number N of loci, which we call L below; and K , a measure of the ruggedness of the fitness landscape \mathbf{f} , where the fitness landscape is smoothest at $K=0$ and most rugged for $K=L-1$. Each locus $\ell=1, \dots, L$ interacts with K random other loci, meaning that the fitness landscape contains at most $(K+1)$ -way interactions. Since the NK model simulates fitness values under an additive model, we exponentiated the NK fitness values.

Each simulated fitness landscape \mathbf{f} has an associated graph $G=(V, E)$ which describes a (simulated) *genetic interaction network*, where the vertices $V=\{1, \dots, L\}$ are the L loci and the edges E connect pairs of interacting loci⁵¹. For example, for $K=0$, the graph G has no

edges, indicating that there are no interactions between loci, while for $K=1$ the graph G has edges connecting loci with pairwise interactions. (For $K \geq 2$, one may also describe the interaction relationships with a *hypergraph* where hyperedges connect sets of interacting loci, e.g., ref. 51.)

We find that the chimeric measure falsely indicates the presence of higher-order interactions that are not present in the simulated fitness landscape \mathbf{f} while the multiplicative measure does not. For example, when the fitness landscape \mathbf{f} contains only *pairwise* interactions (i.e., $K=1$), then the 3-way multiplicative epistasis measure $\epsilon^M_{ijk} = 0$ is equal to zero for all triples (i, j, k) of loci. However, if the NK model graph G contains a triangle (i, j, k) , then the 3-way chimeric measure $\epsilon^C_{ijk} \neq 0$ will be non-zero with high probability (Fig. 3A). Thus, the chimeric measure ϵ^C falsely indicates the presence of three-way interactions that do not exist in the simulated fitness landscape. (As this is sometimes a point of confusion: we note that triangles in a graph are sometimes referred to as higher-order structures⁵². However, as our simulation demonstrates, it is quite possible to have a triangle in a graph, i.e., three pairwise interactions, without having a genuine higher-order (3-way) interaction.) More generally, for any value $K > 0$ of the ruggedness parameter, the fitness landscape \mathbf{f} only contains at most $(K+1)$ -way interactions. The $(K+2)$ -way multiplicative measure ϵ^M is always equal to zero, reflecting that there are no $(K+2)$ -way interactions. However, we empirically observe that the $(K+2)$ -way chimeric measure ϵ^C is often non-zero (Fig. 3B).

Thus, our analyses demonstrate how the chimeric measure ϵ^C will often erroneously identify higher-order interactions that are not present in the underlying fitness landscape.

Three-way epistasis in budding yeast

We investigate the biological implications of using the chimeric epistasis measure instead of the multiplicative epistasis measure by re-analyzing two triple-gene-deletion studies in budding yeast by Kuzmin et al.^{24,26}. These studies used triple-mutant synthetic genetic arrays (SGA)^{53,54} to measure the fitness of single-, double-, and triple-mutant strains. The authors use a multiplicative fitness model since the SGA protocol models yeast colony sizes as a product of fitness, time, and experimental factors²³. The Kuzmin et al. studies, refs. 24 and 26, measure fitness values for 195,666 and 256,861 gene triplets, respectively. They calculate the three-way chimeric epistasis measure ϵ^C_{ijk} and report 3196²⁴ and 2466²⁶ negative three-way epistatic interactions, respectively.

We calculated the multiplicative epistasis measure ϵ^M_{ijk} (formula (4)) and the chimeric epistasis measure ϵ^C_{ijk} (formula (5)) used by Kuzmin et al.^{24,26} for the 189,340 gene triplets (i, j, k) whose single-, double- and triple-mutant fitness values were available in the publicly

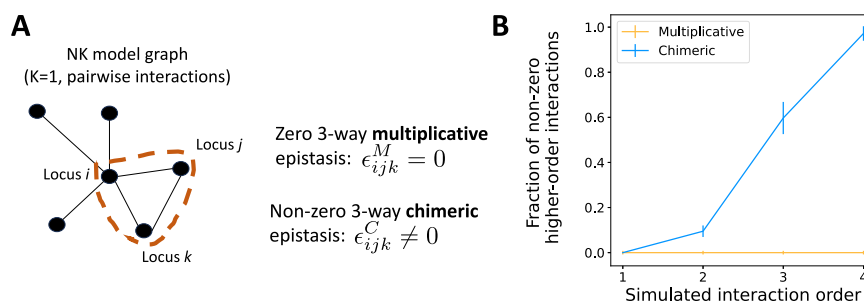


Fig. 3 | Comparison of epistasis measures using simulated data from the NK fitness model. **A** A fitness landscape \mathbf{f} simulated following the NK fitness model with “ruggedness” parameter $K=1$ contains only pairwise interactions. These interactions are represented with an interaction graph G . The 3-way multiplicative measure $\epsilon^M_{ijk} = 0$ equals zero for all loci triples (i, j, k) . However, if the triple (i, j, k) forms a triangle in the graph G (shown as a dashed red line), then the 3-way chimeric epistasis measure ϵ^C_{ijk} is non-zero and incorrectly indicates the presence of a higher-

order interaction. **B** The fraction of non-zero $(K+2)$ -way interactions (“fraction of non-zero higher-order interactions”) identified by the multiplicative measure ϵ^M (orange) and the chimeric measure ϵ^C (blue) across 100 fitness landscapes \mathbf{f} simulated according to the NK fitness model with ruggedness parameter K , with error bars indicating standard deviation across simulated instances. The fitness landscape \mathbf{f} contains at most $(K+1)$ -way interactions, but the chimeric measure ϵ^C spuriously detects many non-zero $(K+2)$ -way interactions.

Table 2 | Comparison of signs of trigenic interactions in budding yeast calculated using the multiplicative epistasis measure and the chimeric epistasis measure on fitness data from Kuzmin et al.²⁶

		Chimeric measure ϵ_{ijk}^C		
		Positive	Ambiguous	Negative
Multiplicative measure ϵ_{ijk}^M	Positive	1197	259	0
	Ambiguous	116	4291	91
	Negative	10	466	1527

Values not on the diagonal correspond to gene triplets having a different sign of epistasis using the multiplicative measure versus the chimeric measure (approximately 12% of triplets).

available data from refs. 24,26 and with a reported p -value of $p_{ijk} < 0.05$. Following Kuzmin et al.^{24,26} we say a gene triplet (i, j, k) has a *positive chimeric interaction* if $\epsilon_{ijk}^C > 0.08$; a *negative chimeric interaction* if $\epsilon_{ijk}^C < -0.08$; and an *ambiguous chimeric interaction* if $-0.08 < \epsilon_{ijk}^C < 0.08$. Accordingly, using the same quantile as the chimeric threshold of 0.08, we say that a gene triplet (i, j, k) has a *positive (resp. ambiguous, negative) multiplicative interaction* if $\epsilon_{ijk}^M > 1.105$ (resp. $0.905 < \epsilon_{ijk}^M < 1.105$, $\epsilon_{ijk}^M < 0.905$). See Supplementary Note 4 and Supplementary Figs. 3 and 4 for specific details on data processing and reproducing the Kuzmin et al. results.

We observed considerable differences between the signs of the multiplicative epistatic measure versus the chimeric epistatic measure (Table 2). In particular, approximately 12% of gene triplets have a *different* interaction sign with the multiplicative measure compared to the chimeric measure. The difference between the two measures is especially pronounced for *negative* interactions, which are typically more functionally informative than positive interactions^{23,24,26}. In particular, there were 476 gene triplets (i, j, k) with a negative *multiplicative-only* interaction, or triplets with a negative multiplicative interaction but not a negative chimeric interaction (Fig. 4A). On the other hand, there were only 91 gene triplets with a negative *chimeric-only* interaction, or triplets with a negative chimeric interaction but not a negative multiplicative interaction (Fig. 4A); in fact, some of these 91 triplets even had *positive* multiplicative interaction (Fig. 4A). We also observe a qualitatively similar discrepancy between the two formula using the earlier fitness data from Kuzmin et al. (2018)²⁴; on this data, we find that there were 746 gene triplets with a negative multiplicative-only interaction versus 177 triplets with a negative chimeric-only interaction (Supplementary Fig. 5). Our results were also qualitatively similar when we did not restrict to triplets with reported p -value $p_{ijk} < 0.05$ (Supplementary Fig. 6).

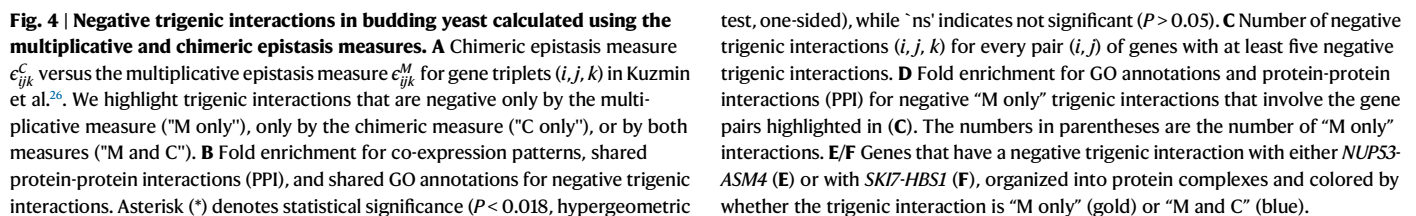
Negative trigenic interactions often contain genes whose proteins are partially redundant in their functions²⁵ and are enriched for other features that arise from biological models of functional redundancy, including shared expression patterns^{55,56}, shared protein-protein interactions⁵⁷, GO annotation, and amino acid divergence^{56,57}. We observed (Fig. 4B) that gene triplets with negative multiplicative-only interactions – that is, gene triplets not identified by the chimeric formula used in Kuzmin et al. (2020)²⁶ – are significantly enriched for co-expression ($P = 0.017$, hypergeometric test), shared protein-protein interactions ($P < 1.5 \times 10^{-4}$, hypergeometric test), and similar GO annotations ($P < 2.1 \times 10^{-5}$, hypergeometric test). In contrast, gene triplets with a negative chimeric-only interaction are not significantly enriched for any of these features (Fig. 4B). In this way using the multiplicative measure extends the network of functionally redundant genes by almost 25% compared to the chimeric measure. We obtain a similar result when analyzing the fitness data from the earlier Kuzmin et al. (2018) study²⁴ (Supplementary Fig. 5) and also when we do not remove gene triplets with large reported p -values p_{ijk} as computed by Kuzmin et al.^{24,26} (Supplementary Fig. 6). These results demonstrate that using the appropriate three-way multiplicative formula for a

multiplicative fitness model leads to more biologically meaningful higher-order genetic interactions compared to using the chimeric epistasis formula that mixes additive and multiplicative scales in an statistically unsound manner.

In particular, trigenic interactions also reveal the functional redundancy of *paralogs*, or pairs of duplicated genes with overlapping functions, since two functionally similar genes tend to have a negative trigenic interaction with a third gene more often compared to gene pairs with non-overlapping functions²⁶. Thus, we evaluated whether the gene triplets with negative multiplicative-only interactions involve functionally redundant gene pairs. We quantified the functional redundancy between two genes by calculating the number of negative trigenic interactions to which both genes belong, where we restricted our calculation to gene pairs involved in at least two negative multiplicative interactions. We found that many pairs of genes had additional multiplicative-only interactions (Fig. 4C). Thus the multiplicative measure identified additional functional redundancies not found using the chimeric measure. As additional validation, we note that Kuzmin et al.²⁶ quantify functional redundancy between two genes using a related quantity that they call the *trigenic interaction fraction* (see Supplementary Note 4 for more details). We observed that for most gene pairs, the trigenic interaction fraction is larger when computed using the multiplicative formula versus using the chimeric formula (Supplementary Fig. 7). This observation further supports the conclusion that the multiplicative formula uncovers additional functional redundancies between these paralogs that was not detected by the chimeric measure.

We expect paralogs with large increases in the number of multiplicative-only interactions to be functionally redundant. Of the 130 paralogs we analyzed, there are fifteen paralogs with at least 10 negative multiplicative-only interactions (highlighted in Fig. 4C). The three paralogs with the largest number of negative multiplicative-only interactions were *RPS24A-RPS25B*, *MSN2-MSN4*, and *ARE1-ARE2*. For these three paralogs, the multiplicative formula quadrupled the number of total trigenic interactions compared to the number of such interactions reported by Kuzmin et al.²⁶ using the chimeric formula. These three paralogs also appear to have redundant functions according to other patterns of sequence evolution: all three have highly correlated position-specific evolutionary rates (Table S12 in ref. 26) and two of them (*RPS24A-RPS25B* and *ARE1-ARE2*) have low sequence divergence rates (Supplementary Fig. 8). Moreover, negative genetic interactions have been previously documented for *MSN2-MSN4*⁵⁸, *ARE1-ARE2*⁵⁹, and *RPS25A-RPS25B*^{21,60}.

The paralogs with many multiplicative-only interactions are also enriched for shared PPIs or GO annotations with the genes they interact with (Fig. 4D). In particular, the paralogs *NUP53-ASM4*, which are components of the large nuclear pore complex⁶¹, had 36 additional negative multiplicative-only interactions. These epistatic interactions are highly enriched for shared PPIs and GO annotations (Fig. 4D) and also involve members of the same protein complexes (Fig. 4E). One of the 36 additional genes that interact with *NUP53-ASM4* is *NUP145*, which also forms part of the nuclear pore⁶². Interestingly, while the gene triplet *NUP53-ASM4-NUP145* has a negative multiplicative interaction ($\epsilon^M = 0.684 < 1$), the same gene triplet was reported to have a *positive* chimeric interaction ($\epsilon^C = 0.25 > 0$; Kuzmin et al.²⁶). Another example of one of the 36 additional interactions is *SAC3*, which encodes a nuclear pore-associated protein that functions in mRNA transport⁶³. The gene triplet *NUP53-ASM4-SAC3* has a very negative multiplicative interaction ($\epsilon^M = 0.046 < 1$), but in the original study²⁶ was reported to have a slightly positive chimeric interaction ($\epsilon^C = 0.014 > 0$). Moreover, both *NUP145* and *SAC3* share at least one protein-protein interaction and GO category with *NUP53* and *ASM4*. These findings provide additional support to the hypothesis by Kuzmin et al.²⁶ that *NUP53* and *ASM4* have overlapping functions.



In addition to the negative interactions just described, we also highlight an example of a biologically relevant *positive* trigenic interaction that is missed by the chimeric epistasis measure but detected by the multiplicative measure. The gene triplet *CIK1-VIK1-SUP35*-*td*, which consists of two paralogs, *CIK1* and *VIK1*, involved in mitosis⁶⁹, and the essential gene *SUP35*-*td*⁷⁰, has an ambiguous, negative chimeric interaction ($\epsilon_{ijk}^C = -0.03$) but has a very large, positive multiplicative interaction ($\epsilon_{ijk}^M = 75.337$). Examining the fitness values (Supplementary

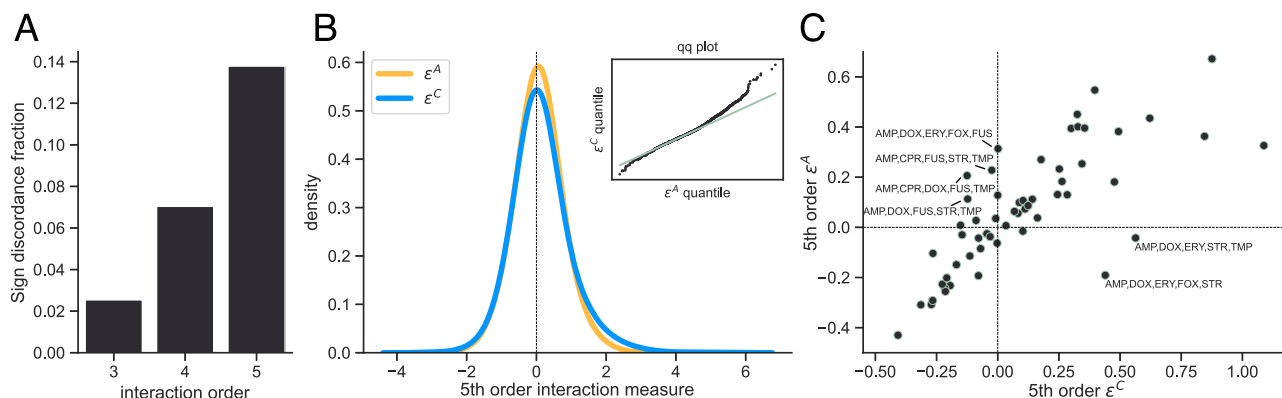


Fig. 5 | Higher-order interactions between antibiotics in *E. coli* using drug response data from Tekin et al.⁴⁰ **A** Proportion of *E. coli* cultures where the sign (positive vs. negative) of the chimeric and additive measures disagree. The sign discordance fraction, or proportion of interactions where the sign of the two measures disagree, increases with the interaction order, consistent with the

simulations shown in Fig. 1. **B** Distributions and Q-Q plots (insets) for the additive (orange) and chimeric (blue) measures for 5th-order interactions. **C** Scatter plot of median relative growth rates for each 5-way combination of antibiotics across concentration levels and replicates. Dashed horizontal and vertical lines indicate zero additive and chimeric epistasis measures, respectively.

Fig. 9) shows that the fitness of the *CIK1-VIK1-SUP35-td* triple mutant is more than 100 times larger than the fitness of the *CIK1-SUP35-td* double mutant. Moreover, positive interactions have been previously documented between pairs of these genes: *VIK1* deletion mutants suppress several phenotypes of *CIK1* deletion mutants, including a mitotic delay phenotype and a temperature-dependent fitness defect⁶⁹; and a phenotypic suppression interaction exists between *CIK1* and *SUP35*, where deletion of *CIK1* reduces the ability of *SUP35-td* to form prions⁶⁹. These previously identified positive pairwise interactions, together with the large triple-mutant fitness value, demonstrate that the gene triplet *CIK1-VIK1-SUP35-td* is more likely to have a positive interaction as indicated by the multiplicative measure, rather than a neutral interaction as indicated by the chimeric measure.

Overall, our results demonstrate not only the degree to which the multiplicative and chimeric formula may lead to distinct interpretations of fitness data, but also that genetic interactions measured using the multiplicative formula appear to be more consistent with other biological features compared to interactions measured using the chimeric formula.

Higher-order interactions in drug responses

We next reanalyzed a drug response dataset⁴⁰ in which three-way, four-way, and five-way interactions between drug combinations were quantified using the chimeric formula. For these data, the authors exposed *Escherichia coli* cultures to between one and five antibiotics (out of eight total) at one of three different concentrations. They measured fitness as the difference in exponential growth rates between the culture exposed to antibiotics and a negative control with no antibiotics. The authors then used the chimeric epistasis measure ϵ^C to identify third-, fourth-, and fifth-order interactions between different combinations of antibiotics. We compared their results with the additive epistasis measure ϵ^A . We used the additive measure ϵ^A because, under the standard assumption that antibiotic exposure multiplicatively affects the survival probability of individual cells⁴⁶, then antibiotic exposure will have an additive effect on the exponential growth rates of the population of cells^{11,71}.

The signs of the chimeric interaction measure ϵ^C and the additive interaction measure ϵ^A disagree for three-way, four-way, and five-way interactions, with the discrepancy between the two measures increasing with the interaction order (Fig. 5A), which is consistent with our earlier simulations (Fig. 1C). The discrepancy is largest for fifth-order interactions, with approximately 14% of fifth-order interactions having a different sign using the additive measure versus the chimeric measure (Fig. 5A).

The discrepancy between the additive and chimeric measures may lead to different conclusions on the type of interactions between antibiotics, i.e., whether a given combination of antibiotics is *synergistic* (more effective at killing bacteria when taken together versus taken individually, i.e., a *negative* interaction) or *antagonistic* (less effective together versus individually, i.e., a *positive* interaction). For fifth-order interactions, the chimeric measure ϵ^C was more positively skewed than the additive measure ϵ^A (Fig. 5B), with a Pearson skewness coefficient of 0.87 for the chimeric measure versus 0.17 for the additive measure. Thus, the chimeric measure is significantly more likely to identify antagonistic interactions than the additive measure ($P < 7 \times 10^{-43}$, paired t-test).

We then examined specific five-way combinations of antibiotics with different interaction signs following the procedure of ref. 27 and ref. 45. For each five-way combination of antibiotics we first calculated the median relative growth rate of *E. coli* across replicates and concentrations, and then used these median relative growth values to compute both the additive and chimeric measures (Fig. 5C). The interaction between the antibiotic combination Ampicillin (AMP), Doxycycline hyclate (DOX), Erythromycin (ERY), Streptomycin (STR), Trimethoprim (TMP) is highly antagonistic using the chimeric measure (i.e., $\epsilon^C = 0.56 > 0$) but synergistic using the additive measure (i.e., $\epsilon^A = -0.04 < 0$). A similar pattern also holds for the antibiotic combination consisting of AMP, DOX, ERY, STR, and Cefoxitin sodium salt (FOX). We emphasize that because we use the same fitness values as reported in ref. 40, the differences between the additive and chimeric measures arise *solely* from the use of the additive versus chimeric measures as opposed to variability arising from biological or technical replicates.

Epistasis between protein mutations

We further demonstrate the difference between the multiplicative and chimeric epistasis measures using experimental fitness data of eleven different proteins^{10,38,72–79}. These fitness values were measured using deep mutational scanning (DMS), a recent class of technologies which use high-throughput sequencing to measure the fitness of many variants of a protein. The published analyses of each of these datasets quantified epistasis using either an additive or multiplicative epistasis measure, depending on the fitness measurement being made. We reanalyzed each dataset using the chimeric measure to demonstrate the differences between the chimeric measure and the additive/multiplicative measures.

We found that the multiplicative and chimeric measures have substantial disagreement in quantifying higher-order epistasis within

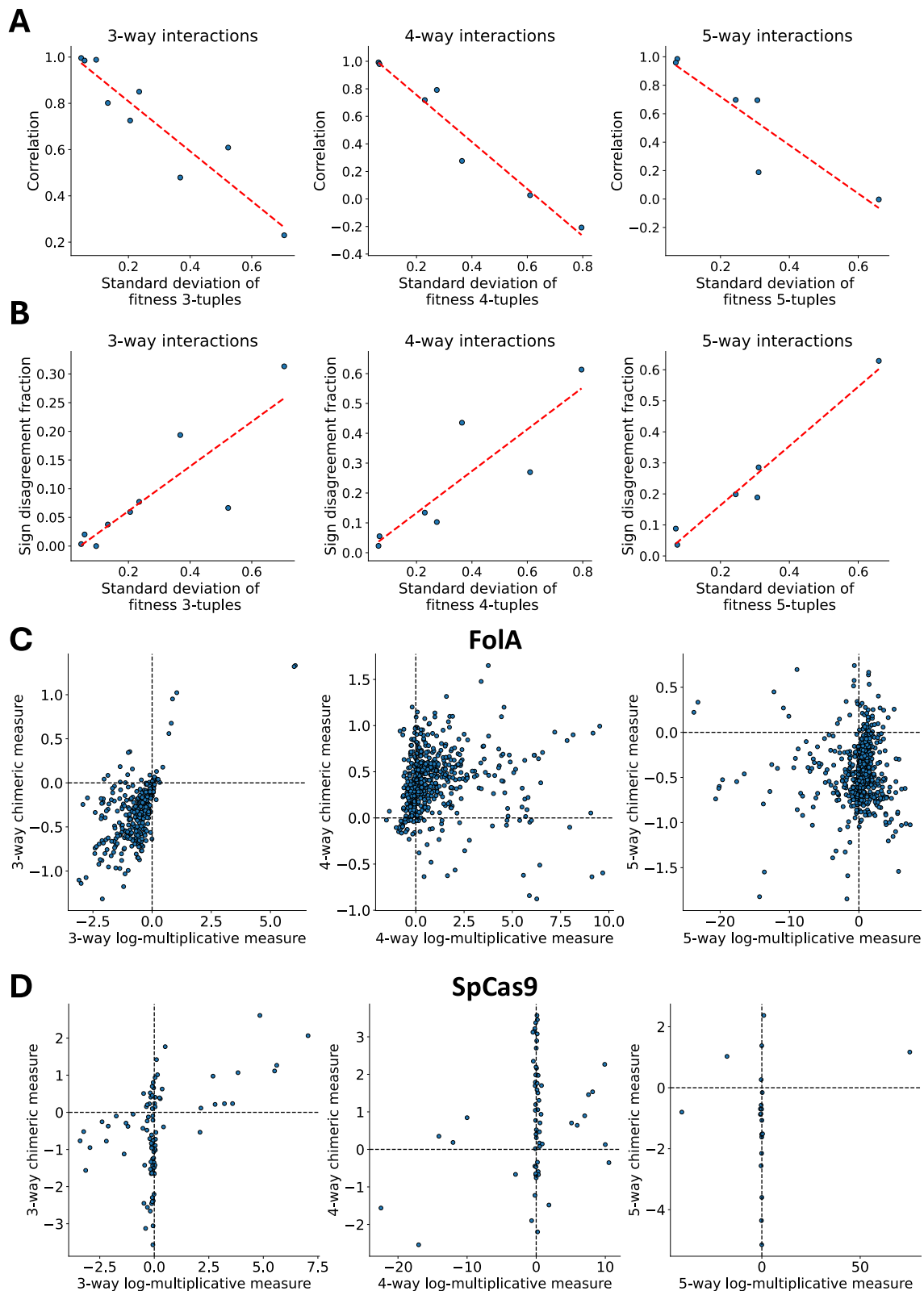


Fig. 6 | Epistasis between protein mutations in eleven different proteins.

A, B Standard deviation of fitness values across all (left) three-, (middle) four-, and (right) five-way tuples of mutations versus the average (A) correlation and (B) sign disagreement fraction of the log-multiplicative measure $\log \epsilon^M$ versus the chimeric

measure ϵ^C . Line of best fit is shown as a dashed red line. **C, D** Log-multiplicative measure $\log \epsilon^M$ versus chimeric measure ϵ^C for the (C) FolsA⁷² and (D) *Streptococcus pyogenes* Cas9 (SpCas9) nuclease³⁸ proteins. Dashed vertical and horizontal lines indicate zero log-multiplicative and chimeric epistasis measures, respectively.

several of the proteins. Furthermore, we observe that both the correlation between the measures and the sign disagreement fraction vary as a function of the standard deviation s of the 2^K fitness values $\{f_{0\dots 0}, \dots, f_{1\dots 1}\}$ across all K -tuples of mutation. Specifically, the correlation between the chimeric measure and the multiplicative measure *decreases* as a function of the fitness standard deviation s (Fig. 6A), while the sign disagreement function *increases* as a function of the fitness standard deviation s . These results show the large difference between the multiplicative and chimeric measures for proteins with a large standard deviation in fitness values.

The FoaA metabolic protein from *E. coli* has the second largest standard deviation s in fitness of the eleven proteins that we analyzed. This data from ref. 72 includes the fitness of approximately 260,000 mutations at nine single-nucleotide loci. The three-way multiplicative and chimeric measures for the FoaA protein have correlation 0.6086, while the four- and five-way measures have correlation < 0.05 – i.e., the two measures are almost uncorrelated for four- and five-way interactions (Fig. 6C). There is also substantial sign disagreement between the multiplicative and chimeric measures, with over 60% sign disagreement for five-way interactions.

Another protein with large fitness standard deviation s is the *Streptococcus pyogenes* Cas9 (SpCas9) nuclease, a widely used protein for genome editing across biology. The fitness landscape of SpCas9 was profiled in ref. 38, where fitness was measured by the editing efficiency of the SpCas9 protein. For the SpCas9 protein, the sign disagreement between the two epistasis measures is over 20% for three-, four-, and five-way interactions (Fig. 6D). The large sign disagreement between the two epistasis measures is likely because for many protein variants, the log-multiplicative measure $\log e^M$ is close to 0 while the chimeric measure e^C varies substantially between -4 and 2 .

Overall, our results demonstrate the extent to which one may infer substantially different higher-order epistasis between protein mutations – including different *signs* of epistasis – if the chimeric measure is used in place of the additive/multiplicative measures.

Discussion

Higher-order interactions between genetic variants, drugs, and other perturbations play a large role in shaping the fitness landscape of an organism^{1,80}. Yet despite the importance of these interactions, there are multiple different – and sometimes inconsistent – formulae used in the literature for measuring higher-order interactions, most notably for measuring higher-order epistasis between mutations. In particular, many researchers use a *chimeric* formula that quantifies epistasis as an additive deviation from a multiplicative null model and is thus a “chimera” of additive and multiplicative measurement scales.

In this work, we show that there is considerable disagreement between the chimeric epistasis measure and the additive and multiplicative measures. For higher-order interactions, the chimeric measure often has a different *sign* compared to the multiplicative measure (Fig. 1C). We demonstrate that this inconsistency is not purely a mathematical curiosity but also leads to markedly different biological conclusions in yeast genetics^{24,26} (Fig. 4), antibiotic resistance^{40,45} (Fig. 5), and protein epistasis (Fig. 6), raising potential questions about some reported higher-order epistatic interactions in the literature. Furthermore, we show that the different epistasis measures are equal to different parametrizations of the multivariate Bernoulli distribution (MVB)²⁹ (Table 1) and demonstrate that the chimeric epistasis measure is less statistically sound than the additive and multiplicative measures. Our connection between epistasis measures and parameters of the multivariate Bernoulli measure is general and unifies many different epistasis measures: the additive, multiplicative, and chimeric measures; and the Walsh coefficients^{30–32,42,43}. Overall, our results demonstrate that the more appropriate multiplicative and additive formulae for higher-order epistasis yield more mathematically sound and

biologically meaningful results compared to the chimeric formula which improperly conflates measurement scales.

Historically, most work in epistasis has focused on pairwise interactions, where the chimeric and multiplicative measures agree on the interaction sign, and thus the differences between these two measures are not widely reported. However, even in the pairwise setting, the two measures have different magnitudes, which may still affect biological findings. For example, Costanzo et al.²¹ recently built a large-scale pairwise interaction network for yeast using the chimeric epistasis measure, where they included an edge between two genes if the absolute value of the chimeric measure was greater than a certain threshold. From our results with the trigenic yeast network (Section 2.6), it is possible that the edges in the network would change if one used the more appropriate multiplicative measure instead, which may lead to the inference of different genetic interactions and thus the functional relationships and regulatory mechanisms identified by Costanzo et al.²¹.

There are several future directions for our work. First, it would be useful to further investigate the relationship between the MVB and regression-based approaches for quantifying epistasis in GWAS^{81,82}. For example, regression-based approaches often do not require that one has measured the fitness of all 2^L genotypes, which may make the estimation of the interaction parameters of the MVB more challenging. Moreover, these regression approaches may sometimes produce biased estimates of epistasis⁸³, and we imagine that the MVB would provide a useful statistical framework for characterizing such statistical biases. A second direction is to incorporate uncertainty of fitness measurements in the MVB, e.g., by using a Bayesian framework. Thirdly, one could generalize our theoretical results by relaxing the assumption that fitness is proportional to genotype probability, e.g., by incorporating genetic drift or other more evolutionarily realistic factors. Fourth, our statistical framework could be extended to model how higher-order interactions contribute to evolutionary trajectories in a fitness landscape⁸⁴. Fifth, it would be quite interesting to investigate the connections between the MVB and the circuit formulae used to quantify the shape of a fitness landscape^{19,42,43,85}. Finally, we note that our MVB framework provides an approach for doing formal model comparisons. Thus, an interesting and important future direction would be to derive a statistical test using the MVB to test whether an additive or multiplicative fitness model better fits data from different experimental technologies.

Ultimately, future studies on interactions in genetics, drug response, protein fitness landscapes, and other domains should take care to use the mathematically appropriate additive or multiplicative formula for measuring higher-order interactions, and not fall victim to the chimera.

Methods

Pairwise epistasis

We start with the simplest setting where the genotype consists of two loci, each with two alleles labeled 0 and 1, i.e., a haploid genome with a single biallelic mutation. Thus, there are four possible genotypes – the wild-type 00, the single mutants 01 and 10, and the double mutant 11 – with corresponding fitness values f_{00}, f_{01}, f_{10} , and f_{11} (Fig. 1). There are two standard null models that relate genotype to fitness: the additive model and the multiplicative model.

Additive fitness model. In the first model, mutations are assumed to have an *additive* effect on fitness^{2,19}, e.g., in drug resistance^{42,85} and protein binding^{30,32}. The effect of a mutation is quantified by the difference in fitness when one locus is mutated; for example, $f_{11} - f_{10}$ measures the effect of a mutation in the second locus, where the genetic background is a mutation in the first locus. Interaction between mutations in the two loci, i.e., *pairwise epistasis*, is measured by the difference in the effect of a mutation in one locus across the two

possible genetic backgrounds (Supplementary Fig. 10A). The pairwise interaction measure ϵ^A is given by

$$\epsilon^A = (f_{11} - f_{10}) - (f_{01} - f_{00}). \quad (7)$$

Note that the definition (7) of the pairwise epistasis measure is invariant to the choice of which locus is mutated, i.e., $\epsilon^A = (f_{11} - f_{10}) - (f_{01} - f_{00}) = (f_{11} - f_{01}) - (f_{10} - f_{00})$. In practice, the fitness values are often normalized so that $f_{00} = 0$, i.e., the fitness f_{00} of the wild-type is zero, resulting in the following commonly-used equation for pairwise epistasis under an additive fitness model:

$$\epsilon^A = f_{11} - (f_{01} + f_{10}). \quad (8)$$

Equivalently, the pairwise epistasis measure ϵ^A is the difference between the observed double-mutant fitness f_{11} and the expected double-mutant fitness $f_{01} + f_{10}$ under a null model with no epistasis. As ref. 2 notes, this definition of pairwise epistasis is similar to Fisher's original definition of epistasis⁴¹.

The sign $\text{sgn}(\epsilon^A)$ of the pairwise epistasis measure ϵ^A determines the type of epistatic interaction. If $\epsilon^A = 0$, then there is no interaction between the two loci and so the fitness f_{11} of a double mutant is completely determined by the sum $f_{11} = f_{01} + f_{10}$ of the single mutant fitnesses f_{01}, f_{10} . If $\epsilon^A > 0$ then there is a *positive* interaction between the two loci, in the sense that the fitness f_{11} of the double mutant is *larger* than the fitness if there was no pairwise interaction. Similarly, if $\epsilon^A < 0$ then there is a *negative* interaction between the two loci, in the sense that the fitness f_{11} of the double mutant is *smaller* than the fitness if there was no pairwise interaction.

The pairwise epistasis measure ϵ^A is equivalent to two other notions of epistasis used in the genetics literature. First, the pairwise epistasis measure ϵ^A is equal to the pairwise interaction term in the standard linear *regression* framework for quantifying epistasis¹². Specifically, if the fitness values $f_{00}, f_{01}, f_{10}, f_{11}$ follow a linear model of the form

$$f_{x_1x_2} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2, \quad (9)$$

then the coefficient β_{12} of the interaction term x_1x_2 is equal to the pairwise epistasis measure ϵ^A in (7). Second, the epistasis measure ϵ^A is equal (up to a constant factor) to the 2nd-order Walsh coefficient that is often used to measure “background-averaged” epistasis^{30–32,42,43}.

Multiplicative fitness model. In this model, mutations are assumed to have a *multiplicative* effect on fitness, e.g., modeling cellular growth rates^{2,14,21,23,24}. The multiplicative pairwise epistasis measure (Supplementary Fig. 10B) is given by

$$\epsilon^M = \frac{f_{11}}{f_{10}} / \frac{f_{01}}{f_{00}} = \frac{f_{11}f_{00}}{f_{10}f_{01}}. \quad (10)$$

As in the additive model, in practice the fitness values are typically normalized such that $f_{00} = 1$, resulting in the following equation for pairwise epistasis:

$$\epsilon^M = \frac{f_{11}}{f_{01}f_{10}}. \quad (11)$$

That is, the pairwise epistasis measure ϵ^M is the *ratio* between the double-mutant fitness f_{11} and the product $f_{01}f_{10}$ of the single-mutant fitness values.

The multiplicative fitness model is closely related to the additive fitness model: if fitnesses f are multiplicative, then the *log*-fitnesses $\log f$ are additive. Thus, the sign of the interaction is determined by the difference between the epistasis measure ϵ^M and 1, or equivalently the sign $\text{sgn}(\log \epsilon^M)$ of the $\log \epsilon^M$ of the epistasis measure ϵ^M (Fig. 1A). If

$\epsilon^M > 1$, i.e., $\log \epsilon^M > 0$, there is a *positive* interaction between the two loci; if $\epsilon^M = 1$, i.e., $\log \epsilon^M = 0$, then there is no interaction between the two loci; and if $\epsilon^M < 1$, i.e., $\log \epsilon^M < 0$, then there is a *negative* interaction between the two loci.

The multiplicative pairwise epistasis measure is closely related to the pairwise interaction term in the standard *log-linear* regression framework for epistasis¹². Specifically, if the fitness values $f_{00}, f_{01}, f_{10}, f_{11}$ follow a log-linear regression model of the form

$$\log f_{x_1x_2} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_{12}x_1x_2, \quad (12)$$

then $\beta_{12} = \epsilon^M$.

Chimeric formula. Many studies in genetics use a multiplicative fitness model but do not measure pairwise epistasis with the multiplicative epistasis measure ϵ^M . Instead, these papers use a multiplicative null model but measure deviations with an *additive* scale, yielding the following epistasis measurement:

$$\epsilon^C = f_{11} - f_{01}f_{10}. \quad (13)$$

We call ϵ^C_{ij} a “chimeric” measure as it measures deviations from a multiplicative null model on an additive scale, and is thus a *chimera* of both the multiplicative and additive measurement scales. The chimeric measure has been widely used in the genetics literature (e.g., refs. 2,14,23,86) and in the drug interaction literature (e.g., refs. 27,40,44–48). In these applications, similar to the additive measure, the sign of an interaction between two loci is determined by the $\text{sgn}(\epsilon^C)$ of the chimeric measure ϵ^C : $\epsilon^C > 0$ corresponds to a positive interaction while $\epsilon^C < 0$ corresponds to a negative interaction.

Although it is often described in terms of a multiplicative fitness model, the chimeric epistasis measure ϵ^C is *not* equal to the multiplicative measure ϵ^M . The chimeric epistasis measure ϵ^C in Equation (13) is similar to equation (11), but the deviation between the observed double-mutant fitness f_{11} and the expected fitness $f_{01}f_{10}$ under a multiplicative null model is computed using subtraction instead of division. Equivalently, the (log-)multiplicative epistasis measure $\log \epsilon^M = \log f_{11} - \log f_{01}f_{10}$ computes the difference between the observed and expected logarithm of the fitness of the double mutant, while the chimeric epistasis measure $\epsilon^C = f_{11} - f_{01}f_{10}$ computes the difference directly (Fig. 1A). In this way, the chimeric epistasis measure may overstate or understate the strength of a pairwise interaction in a multiplicative fitness model (Fig. 1A); see Supplementary Note 2 for a numerical example highlighting this issue with the chimeric measure.

We note that the differences between the multiplicative epistasis measure ϵ^M_{ij} and chimeric epistasis measure ϵ^C_{ij} do not appear to be widely appreciated in either the applied or theoretical literature. Almost every study that uses the chimeric epistasis measure ϵ^C_{ij} does not consider the multiplicative measure ϵ^M_{ij} . On the other hand, while many in the statistics literature draw a distinction between additive and multiplicative interaction effects (e.g., refs. 12,13), none of these papers discuss the chimeric interaction measure ϵ^C_{ij} that is frequently used in the genetics and drug interaction literature. An exception is Gao, Granka, and Feldman⁸⁷ who refer to the multiplicative formula (2) as a “rescaling of [the chimeric] formula”, but we take the stronger view that “rescaling” obscures consequential implications of the two formula.

Nevertheless, we show that the chimeric measure ϵ^C measures the same *sign* of interaction as the multiplicative measure ϵ^M .

Proposition 1. Let $f_{01}, f_{10}, f_{11} \in \mathbb{R}$ be real numbers. Let $\epsilon^M = \frac{f_{11}}{f_{01}f_{10}}$ and $\epsilon^C = f_{11} - f_{01}f_{10}$. Then $\text{sgn}(\epsilon^C) = \text{sgn}(\log \epsilon^M)$.

Proof. $f_{11} - f_{01}f_{10} > 0 \iff \frac{f_{11}}{f_{01}f_{10}} > 1$.

Choosing an appropriate null model. The appropriate choice of null fitness model depends on the quantitative trait being used to approximate fitness. Cellular growth rate (i.e., the Malthusian parameter⁸⁸, which is often used as a measure of the fitness of microbial populations) is typically described with an additive fitness model; this is because in the population genetics literature, it is often assumed that mutations that independently effect survival and reproduction probabilities combine multiplicatively within individual cells^{89,90}, and so these mutations will combine additively in their effect on the growth rate of a continuously-growing clonal population^{11,71}. However, multiplicative models are sometimes still used inappropriately in this setting⁹. On the other hand, the relative frequency of a microbial (or protein) population is typically modeled multiplicatively, as the growth rate of a population is proportional to the logarithm of its relative frequency. In general, Wagner^{13,46} suggests that one should use the model that preserves the scale on which single-mutant fitness effects were measured (i.e., additive or fold differences from wild type).

Higher-order epistasis

We next generalize our discussion to genotypes with $L \geq 2$ loci, where we demonstrate that the differences between the multiplicative measure and the chimeric measure become even more pronounced when analyzing *higher-order epistasis*, or interactions between three or more loci.

There are 2^L genotypes $x_1 \dots x_L$, where $x_\ell \in \{0, 1\}$ indicates a mutation in locus ℓ , with each genotype $x_1 \dots x_L$ having a corresponding fitness value $f_{x_1 \dots x_L}$, e.g., f_{010} is the fitness of genotype 010 with a mutation in the second locus and no mutations in the first and third loci. However, because writing out the 2^L genotypes is infeasible for large L , we use the following notational shorthand. We use f_i to refer to the fitness of the genotype with a single mutation in locus i , f_{ij} to refer to the fitness of the genotype with mutations in loci i, j , and so on. For example, for $L = 3$ loci, f_2 corresponds to f_{010} while f_{12} corresponds to f_{110} . Without loss of generality, we assume the wild-type fitness f_\emptyset is equal to 0 for the additive fitness model and equal to 1 for the multiplicative and chimeric fitness models. We also define $\epsilon_{ij}^A, \epsilon_{ij}^M, \epsilon_{ij}^C$ as the additive, multiplicative, and chimeric pairwise epistasis measure, respectively, between the i -th locus and the j -th locus, i.e., $\epsilon_{ij}^A = f_{ij} - f_i - f_j$, $\epsilon_{ij}^M = \frac{f_{ij}}{f_i f_j}$ and $\epsilon_{ij}^C = f_{ij} - f_i f_j$. For example, for $L = 3$ loci, ϵ_{12}^M corresponds to $\frac{f_{110}}{f_{100} f_{010}}$.

Additive fitness model. We start by quantifying three-way epistasis in the additive fitness model. When there is no pairwise epistasis, the fitness f_{ijk} of a triple mutant is equal to $f_i + f_j + f_k$, i.e., the fitness from of each of the single mutants. When there is pairwise epistasis, then the triple mutant fitness f_{ijk} also includes pairwise interaction measures, i.e.,

$$f_i + f_j + f_k + \epsilon_{ij}^A + \epsilon_{ik}^A + \epsilon_{jk}^A \quad (14)$$

Three-way epistasis is computed by measuring the difference between the observed triple-mutant fitness f_{ijk} and the expected fitness in (14) when only pairwise interactions are included. Thus, the three-way additive epistasis measure ϵ_{ijk}^A is given by

$$\begin{aligned} \epsilon_{ijk}^A &= f_{ijk} - [f_i + f_j + f_k + \epsilon_{ij}^A + \epsilon_{ik}^A + \epsilon_{jk}^A] \\ &= f_{ijk} - f_{ij} - f_{ik} - f_{jk} + f_i + f_j + f_k. \end{aligned} \quad (15)$$

As in the pairwise case, the sign of the three-way epistatic measure ϵ_{ijk}^A determines the sign of the interaction: if $\epsilon_{ijk}^A > 0$, then there is a *positive* three-way interaction between loci i, j, k – in the sense that the fitness f_{ijk} of the triple mutant is larger than the expected fitness in (14)

when only pairwise interactions are present – while if $\epsilon_{ijk}^A < 0$, then there is a *negative* three-way interaction between loci i, j, k .

Our derivation of the three-way epistasis measure ϵ_{ijk}^A is easily extended to higher-order interactions. The additive K -way epistasis measure $\epsilon_{i_1 \dots i_K}^A$ is defined recursively as

$$\epsilon_{i_1 \dots i_K}^A = f_{i_1 \dots i_K} - \left[\left(\sum_{j=1}^K f_{i_j} \right) + \left(\sum_{1 \leq j_1 < j_2 \leq K} \epsilon_{i_{j_1} i_{j_2}}^A \right) + \dots + \left(\sum_{1 \leq j_1 < \dots < j_{K-1} \leq K} \epsilon_{i_{j_1} \dots i_{j_{K-1}}}^A \right) \right]. \quad (16)$$

The K -way epistasis measures $\epsilon_{i_1 \dots i_K}^A$ are proportional to two other measures of epistasis: (1) the K -th order Walsh coefficient used to quantify background-averaged epistasis among K genetic loci^{30–32} and (2) the K -th order interaction coefficients of a linear regression model, which we discuss in more detail in the following section.

Multiplicative fitness model. We derive formulae for epistasis in a multiplicative fitness model by using the equivalence between multiplicative fitness and additive log fitness. For example, the 3-way epistasis measure ϵ_{ijk}^M in the multiplicative model is given by

$$\epsilon_{ijk}^M = \frac{f_{ijk}}{f_i f_j f_k} - \frac{f_{ijk}}{f_i f_j f_k} \epsilon_{ij}^M \epsilon_{ik}^M \epsilon_{jk}^M = \frac{f_{ijk} f_i f_j f_k}{f_{ij} f_{ik} f_{jk}}. \quad (17)$$

As in the pairwise setting, the sign of interaction is determined by the difference between the multiplicative measure ϵ_{ijk}^M and 1, or equivalently by the $\text{sgn}(\log \epsilon_{ijk}^M)$ of the logarithm of the epistasis measure ϵ_{ijk}^M .

Using (16), then the K -way epistasis measure $\epsilon_{i_1 \dots i_K}^M$ in the multiplicative model is defined recursively by

$$\epsilon_{i_1 \dots i_K}^M = \frac{f_{i_1 \dots i_K}}{\left(\prod_{j=1}^K f_{i_j} \right) \left(\prod_{1 \leq j_1 < j_2 \leq K} \epsilon_{i_{j_1} i_{j_2}}^M \right) \dots \left(\prod_{1 \leq j_1 < \dots < j_{K-1} \leq K} \epsilon_{i_{j_1} \dots i_{j_{K-1}}}^M \right)}. \quad (18)$$

Recent work in the genetics^{24,26} and drug interaction²⁷ claim to measure three-way epistasis using a multiplicative fitness model. However, they do not measure three-way epistasis the multiplicative epistasis formula (17) but instead derive a chimeric formula using both additive and multiplicative measurement scales:

$$\epsilon_{ijk}^C = f_{ijk} - f_i f_j f_k - \epsilon_{ij}^C f_k - \epsilon_{ik}^C f_j - \epsilon_{jk}^C f_i. \quad (19)$$

We call ϵ_{ijk}^C the *chimeric* three-way epistasis measure. In these applications, the sign of the interaction is determined by the $\text{sgn}(\epsilon_{ijk}^C)$ of the chimeric measure ϵ_{ijk}^C .

Despite the claim that the chimeric measure ϵ_{ijk}^C is derived from a multiplicative fitness model, it is clear by inspection that the three-way chimeric measure ϵ_{ijk}^C is not equal to the multiplicative three-way epistasis measure ϵ_{ijk}^M . However, unlike in the pairwise setting, even the *signs* of these two measures disagree (Fig. 1B). We demonstrate in Supplementary Note 2 that even when $\epsilon_{ijk}^M = 1$ – that is, there is no three-way epistasis – the chimeric three-way epistasis measure ϵ_{ijk}^C may still indicate either positive or negative three-way epistasis.

Tekin et al.²⁷ extended the three-way chimeric epistasis formula (19) by heuristically deriving chimeric formulae for 4-way and 5-way epistasis. For example, their chimeric formula for 4-way epistasis is given by

$$\begin{aligned} \epsilon_{ijkl}^C &= f_{ijkl} - f_i f_j f_k f_l - f_i f_j f_k \epsilon_{il}^C - f_i f_j f_l \epsilon_{ik}^C - f_i f_j \epsilon_{kl}^C - f_i f_k \epsilon_{jl}^C - f_i f_l \epsilon_{jk}^C \\ &\quad + 2f_i f_j f_k \epsilon_{il}^C + 2f_i f_j f_l \epsilon_{ik}^C + 2f_i f_j \epsilon_{kl}^C + 2f_i f_k \epsilon_{jl}^C + 2f_i f_l \epsilon_{jk}^C - 6f_i f_j f_k f_l. \end{aligned}$$

As in three-way epistasis, the sign of the 4-way and 5-way chimeric epistasis measures derived by Tekin et al.²⁷ do not match the signs of the corresponding multiplicative epistasis measures (Fig. 1C). This fundamental disagreement motivates a deeper mathematical

understanding of these epistasis measures, which we explore in the following section.

Multivariate Bernoulli distribution

In the previous section, we defined quantitative measures of epistasis for two standard null models for fitness: the additive model and the multiplicative model. Nevertheless, some recent papers use a multiplicative fitness model but instead use an epistasis measure which is a *chimera* of both multiplicative and additive measurement scales. Here, we unify these different epistasis measures using the *multivariate Bernoulli* distribution from probability theory²⁹.

The multivariate Bernoulli distribution describes any distribution on $\{0, 1\}^L$, i.e., binary strings of length L , for $L \geq 2$. The multivariate Bernoulli distribution has three different parameterizations which are used throughout the literature^{29,47}. We start by describing these parameterizations for the simplest such distribution: a *bivariate* Bernoulli distribution over binary strings of length $L = 2$.

Bivariate Bernoulli distribution. Suppose that $X = (X_1, X_2) \in \{0, 1\}^2$ is distributed according to a bivariate Bernoulli distribution. A distribution on X is specified by the parameters $p_{00}, p_{01}, p_{10}, p_{11}$, where $p_{x_1, x_2} = P(X_1 = x_1, X_2 = x_2)$ is the probability of (x_1, x_2) . The parameters $\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11})$ are sometimes called the *general* parameters²⁹. Note that since $p_{00} + p_{01} + p_{10} + p_{11} = 1$, only three such parameters are needed to define the distribution.

The probability density function (PDF) $P(X_1, X_2)$ of $X = (X_1, X_2)$ has the form

$$\begin{aligned} P(X_1, X_2) &= p_{00}^{(1-X_1)(1-X_2)} p_{01}^{(1-X_1)X_2} p_{10}^{X_1(1-X_2)} p_{11}^{X_1X_2} \\ &= \exp \left[\log p_{00} + \left(\log \frac{p_{10}}{p_{00}} \right) X_1 + \left(\log \frac{p_{01}}{p_{00}} \right) X_2 + \left(\log \frac{p_{11} p_{00}}{p_{10} p_{01}} \right) X_1 X_2 \right]. \end{aligned} \quad (20)$$

In other words, the PDF $P(X_1, X_2)$ follows a log-linear model of the form

$$\log P(X_1, X_2) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 \quad (21)$$

for constants $\beta_0, \beta_1, \beta_2, \beta_{12} \in \mathbb{R}$. There is a one-to-one correspondence between the general parameters $\mathbf{p} = (p_{00}, p_{01}, p_{10}, p_{11})$ and the constants $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_{12})$. Thus, a bivariate Bernoulli distribution is also parametrized by the parameters $\boldsymbol{\beta}$, also known as the *natural* parameters of the distribution²⁹. As with the general parameters \mathbf{p} , we note that only three out of the four parameters $\beta_0, \beta_1, \beta_2, \beta_{12}$ are needed to fully specify a distribution. We also note that independence between the random variables X_1 and X_2 is described by the parameter β_{12} , where X_1 and X_2 are independent if and only if $\beta_{12} = 0$.

Equation (21) demonstrates that $X = (X_1, X_2)$ follows an *exponential family* distribution, a wide class of distributions that includes many common distributions including normal distributions or Poisson distributions. In particular, using the terminology of exponential families, equation (21) shows that the *sufficient statistics* of X are X_1, X_2 , and $X_1 X_2$, with corresponding *canonical* parameters β_1, β_2 , and β_{12} ^{48,91}. As a result, the distribution $P(X)$ is uniquely defined by the *expected values* $E[X_1], E[X_2], E[X_1 X_2]$ of the sufficient statistics, sometimes called the *moments* or the *mean parameters* of the distribution⁴⁸. Thus, we obtain a third parametrization of the distribution $P(X)$ using the *moments* $\mu_0 = 1, \mu_1 = E[X_1], \mu_2 = E[X_2], \mu_{12} = E[X_1 X_2]$. The elements of the vector $\boldsymbol{\mu} = (1, \mu_1, \mu_2, \mu_{12})$ of moments are sometimes called the *mean parameters* of the distribution.

Multivariate Bernoulli distribution. The three parametrizations we derived for the bivariate Bernoulli distribution extend to the multivariate Bernoulli distribution. Suppose that $(X_1, \dots, X_L) \in \{0, 1\}^L$ is distributed according to a multivariate Bernoulli distribution. Then the

distribution $P(X)$ of the random variables X is uniquely specified by one of the three following parametrizations.

1. **General parameters:** These are 2^L non-negative values $\mathbf{p} = (p_{x_1 \dots x_L})_{(x_1, \dots, x_L) \in \{0, 1\}^L}$ satisfying

$$p_{x_1 \dots x_L} = P(X_\ell = x_\ell \text{ for } \ell = 1, \dots, L). \quad (22)$$

For example if $L = 3$, then $p_{010} = P(X_1 = 0, X_2 = 1, X_3 = 0)$ and $p_{110} = P(X_1 = 1, X_2 = 1, X_3 = 0)$. Note that since $\sum_{(x_1, \dots, x_L) \in \{0, 1\}^L} p_{x_1 \dots x_L} = 1$, only $2^L - 1$ values $p_{x_1 \dots x_L}$ are necessary to define the distribution.

2. **Natural/canonical parameters:** These are 2^L real numbers $\boldsymbol{\beta} = (\beta_S)_{S \subseteq [L]} \in \mathbb{R}$ satisfying

$$\log P(X_1, \dots, X_L) = \sum_{S \subseteq [L]} \beta_S \cdot \prod_{i \in S} X_i. \quad (23)$$

Similar to the general parameters p_i , only $2^L - 1$ values β_S are necessary to uniquely define the distribution. Typically, the parameter β_\emptyset , often called a normalizing constant or a *partition function* of the distribution, is left unspecified. As noted in the bivariate setting, equation (23) shows that the multivariate Bernoulli is an exponential family distribution with $2^L - 1$ sufficient statistics of the form $\prod_{i \in S} X_i$ for subsets S with $|S| > 0$. Moreover, by rewriting (23) as

$$\log p_{x_1 \dots x_L} = \beta_\emptyset + \left(\sum_{i=1}^L \beta_i x_i \right) + \left(\sum_{1 \leq i_1 < i_2 \leq L} \beta_{i_1 i_2} \cdot x_{i_1} x_{i_2} \right) + \dots + (\beta_{1 \dots L} \cdot x_1 \dots x_L), \quad (24)$$

we observe that the natural parameters $\boldsymbol{\beta}$ correspond to interaction coefficients in a log-linear regression model with response variables \mathbf{p} . For example, the natural parameter β_{12} is the coefficient of the interaction term $x_1 x_2$.

3. **Moments/mean parameters:** These are 2^L real numbers $\boldsymbol{\mu} = (\mu_S)_{S \subseteq [L]}$ satisfying

$$\mu_S = E \left[\prod_{i \in S} X_i \right]. \quad (25)$$

For example if $L = 3$, then $\mu_{13} = E[X_1 X_3]$ while $\mu_{12} = E[X_1 X_2]$. The mean parameters $\{\mu_S\}_{|S| > 0}$ are sufficient statistics for the multivariate Bernoulli distribution, as seen in the exponential family form (23) of the multivariate Bernoulli distribution.

We note that all three parametrizations, as well as the fitness values f and epistasis measures ϵ , can be defined either in terms of subsets $S \subseteq [L]$ as with the natural parameters $\boldsymbol{\beta}$ and moments $\boldsymbol{\mu}$, or in terms of binary strings $x_1 \dots x_L$ as with the general parameters \mathbf{p} . We use both definitions interchangeably, with the convention that a subset $S \subseteq [L]$ corresponds to the binary string $x_1 \dots x_L$ with $x_i = 1_{i \in S}$.

Moreover, when written as vectors indexed by binary strings, the three parametrizations $\boldsymbol{\beta}, \boldsymbol{\mu}, \mathbf{p}$ of the multivariate Bernoulli are related to each other through different linear transformations involving a matrix operation known as the *Kronecker product* (see Supplementary Note 3 for specific formulae). Interestingly, several papers quantify epistasis using the *Walsh-Hadamard transform* which is also defined in terms of Kronecker products^{30–32}. This connection is not a coincidence; in the next section we show that the Walsh-Hadamard transform is closely related to the parametrizations of the multivariate Bernoulli.

Unifying epistasis measures with the multivariate Bernoulli

The multivariate Bernoulli distribution provides an elegant means by which to describe the different epistasis formulae in the literature. We model the genotype $(X_1, \dots, X_L) \in \{0, 1\}^L$ as a random variable

distributed according to a multivariate Bernoulli distribution. The parametrizations of the multivariate Bernoulli correspond to different features of the genotype, as we demonstrate next.

Multiplicative and additive epistasis measures. We start by relating the multiplicative epistasis formula (18) to the multivariate Bernoulli distribution. A careful reader may observe that the natural parameter $\beta_{12} = \log \frac{p_{11}p_{00}}{p_{10}p_{01}}$ in the bivariate Bernoulli distribution (21) bears close resemblance to the multiplicative epistasis measure in equation (10). Specifically, if the fitness $f_{x_1x_2}$ of each genotype $(x_1, x_2) \in \{0, 1\}^2$ is proportional to the probability $p_{x_1x_2}$ of that genotype in the multivariate Bernoulli, then the natural parameter β_{12} is equal to the logarithm $\log \epsilon_{12}^M$ of the multiplicative epistasis measure ϵ_{12}^M . Thus, for $L = 2$ loci, epistasis is measured by the natural parameters β of a bivariate Bernoulli distribution.

We prove that this observation is not specific to the bivariate Bernoulli distribution with $L = 2$ loci, and in fact generalizes to any number L of loci. Specifically, we prove that if the fitness $f_{x_1 \dots x_L}$ of genotype (x_1, \dots, x_L) is proportional to the probability $p_{x_1 \dots x_L}$ of observing the genotype, then for each subset $S \subseteq [L]$ of loci, the natural parameter β_S equals the logarithm $\log \epsilon_S^M$ of the corresponding multiplicative epistasis measure as defined in Equation (18).

Theorem 1. Let $f_x \in \mathbb{R}$ be fitness values for genotypes $\mathbf{x} = (x_1, \dots, x_L) \in \{0, 1\}^L$ such that $f_x = c \cdot p_x$ for some constant $c > 0$ and for some multivariate Bernoulli random variable (X_1, \dots, X_L) with general parameters $\mathbf{p} = (p_x)_{\mathbf{x} \in \{0, 1\}^L}$. Then for all subsets $S \subseteq \{1, \dots, L\}$ of loci, the log-multiplicative epistasis measure $\log \epsilon_S^M$ is equal to the interaction parameter β_S of the random variable (X_1, \dots, X_L) .

By using the equivalence between multiplicative fitness values and additive log-fitness values, we also derive a similar probabilistic interpretation of the additive epistasis formula. Specifically, if fitness $f_{x_1 \dots x_L}$ is proportional to the log-probability $\log p_{x_1 \dots x_L}$ of observing the genotype (x_1, \dots, x_L) , then for each subset $S = \{i_1, \dots, i_k\} \subseteq [L]$ of loci, the natural parameter β_S equals the logarithm $\log \epsilon_{i_1, \dots, i_k}^A$ of the corresponding additive epistasis measure as defined in Equation (16). We formalize this observation as the following Corollary of Theorem 1.

Corollary 1. Let $f_x \in \mathbb{R}$ be fitness values for genotypes $\mathbf{x} = (x_1, \dots, x_L) \in \{0, 1\}^L$ such that $f_x = c \cdot \log p_x$ for some constant $c > 0$ and for some multivariate Bernoulli random variable (X_1, \dots, X_L) with general parameters $\mathbf{p} = (p_x)_{\mathbf{x} \in \{0, 1\}^L}$. Then for all subsets $S \subseteq \{1, \dots, L\}$ of loci, the log-additive epistasis measure $\log \epsilon_S^A$ is equal to the interaction parameter β_S of the random variable (X_1, \dots, X_L) .

We note that Theorem 1 follows from Lemma 3.1 in ref. 29 which states a formula relating the general parameters \mathbf{p} and the natural parameters β of a multivariate Bernoulli distribution. Theorem 1 follows by showing that the right-hand side of the formula in Lemma 3.1 is equal to the multiplicative epistasis measure.

The assumption that the probability $p_{x_1 \dots x_L}$ of observing a genotype (x_1, \dots, x_L) is derived from its fitness $f_{x_1 \dots x_L}$ is often used in generative models for estimating the fitness of protein structures from sequence data³⁹. Moreover, many real-world fitness datasets – including the yeast fitness data and many of the protein datasets analyzed in the Results – measure the fitness of a genotype \mathbf{x} in terms of its relative frequency in a large population of genotypes, i.e., its probability p_x .

We also note that the statistical problem of estimating the natural parameters β or mean parameters μ of a multivariate Bernoulli distribution from samples (X_1, \dots, X_L) of the distribution is computationally hard⁴⁸. The reason why we are able to use relatively simple formulae (16), (18) to compute the natural parameters β is because in this setting, we have both samples (X_1, \dots, X_L) and their corresponding probabilities $P(X_1, \dots, X_L)$, i.e., the fitness values f .

Relationship with (log-)linear regression. Under the assumption that the fitness values $f_{x_1 \dots x_L}$ are proportional to the genotype probabilities $p_{x_1 \dots x_L}$, then (24) is a log-linear regression model of the form

$$\log f_{x_1 \dots x_L} = \beta_{\emptyset} + \left(\sum_{i=1}^L \beta_i x_i \right) + \left(\sum_{1 \leq i_1 < i_2 \leq L} \beta_{i_1 i_2} \cdot x_{i_1} x_{i_2} \right) + \dots + (\beta_{1 \dots L} \cdot x_1 \dots x_L). \quad (26)$$

Thus, Theorem 1 shows that computing the multiplicative epistasis measure ϵ^M is equivalent to computing the interaction parameters of the log-linear regression in (26). The interaction parameters of regression are a standard approach for quantifying epistasis in GWAS and QTL analyses⁵. Equation (26) is sometimes also called the “Taylor series expansion” of a fitness landscape⁹².

Similarly, Corollary 1 demonstrates the equivalence between the additive epistasis measure ϵ^A and the coefficients of a linear regression model with response variables equal to the fitness values. Specifically, under the assumption that the fitness values $f_{x_1 \dots x_L}$ are proportional to the logarithm $\log p_{x_1 \dots x_L}$ of the genotype probabilities, then computing the additive epistasis measures ϵ^A is equivalent to computing the interaction parameters β of the following linear regression model

$$f_{x_1 \dots x_L} = \beta_{\emptyset} + \left(\sum_{i=1}^L \beta_i x_i \right) + \left(\sum_{1 \leq i_1 < i_2 \leq L} \beta_{i_1 i_2} \cdot x_{i_1} x_{i_2} \right) + \dots + (\beta_{1 \dots L} \cdot x_1 \dots x_L). \quad (27)$$

In this way, Theorem 1 and Corollary 1 provide a connection between the multiplicative and additive epistasis measures and the interaction coefficients of log-linear and linear regression models, respectively.

Relationship with case-control GWAS. We also prove that the natural parameters β of an MVB with three variables are closely related to the two standard approaches for measuring pairwise SNP-SNP interactions in a case-control GWAS: logistic regression and conditional independence testing⁴⁹. Specifically, suppose we are given genotype $(X_1, X_2) \in \{0, 1\}^2$ and (binary) disease status $D \in \{0, 1\}$. Then the joint random variable (X_1, X_2, D) follows an MVB distribution, where the log-probability $\log P(X_1, X_2, D)$ is given by the following expression in terms of the natural parameters β :

$$\log P(X_1, X_2, D) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_d D + \beta_{12} X_1 X_2 + \beta_{1d} X_1 D + \beta_{2d} X_2 D + \beta_{12d} X_1 X_2 D. \quad (28)$$

We note that there is a natural approach for representing GWAS data from diploid genomes (with $\{0, 1, 2\}$ -valued allelic states) using binary random variables X_1, X_2 , as described in ref. 93.

We show that the logistic regression approach for measuring pairwise interactions is equivalent to computing β_{12d} , while the conditional independence test is equivalent to testing the null hypothesis $H_0: \beta_{12} = \beta_{12d} = 0$. See Supplementary Note 5 for details.

Chimeric epistasis measure. The multivariate Bernoulli also provides a way of rigorously defining the pairwise and higher-order chimeric epistasis measures using *joint cumulants*. Joint cumulants are a concept from probability theory used to quantify higher-order interactions between random variables. For example, the 2nd order joint cumulant $\kappa(X, Y)$ of two random variables X, Y is given by

$$\kappa(X, Y) = E[XY] - E[X]E[Y], \quad (29)$$

and is equal to the covariance $\text{Cov}(X, Y)$. The 3rd order joint cumulant $\kappa(X, Y, Z)$ of three random variables is given by

$$\kappa(X, Y, Z) = E[XYZ] - \kappa(X, Y)E[Z] - \kappa(X, Z)E[Y] - \kappa(Y, Z)E[X]. \quad (30)$$

Under the assumption that the fitness $f_{x_1 \dots x_L}$ of a genotype (X_1, \dots, X_L) is equal to the corresponding *moment* μ_{x_1, \dots, x_L} , we define the K -way chimeric epistatic measure $\epsilon_{i_1 \dots i_K}^C$ as the K -th order *joint cumulant* $\kappa(X_{i_1}, \dots, X_{i_K})$ of the random variables X_{i_1}, \dots, X_{i_K} .

Definition 1. Let $f_{\mathbf{x}} \in \mathbb{R}$ be fitness values for genotypes $\mathbf{x} = (x_1, \dots, x_L) \in \{0, 1\}^L$ such that $f_{x_1 \dots x_L} = c \cdot \mu_{x_1, \dots, x_L}$ for some constant $c > 0$ and for some multivariate Bernoulli random variable (X_1, \dots, X_L) with moments $\mu_{x_1, \dots, x_L} = E[X_1^{x_1} \dots X_L^{x_L}]$. The *chimeric epistasis measure* $\epsilon_{i_1 \dots i_K}^C$ is the joint cumulant $\kappa(X_{i_1}, \dots, X_{i_K})$ of the random variables X_{i_1}, \dots, X_{i_K} .

Our definition of the K -th order chimeric epistasis measure $\epsilon_{i_1 \dots i_K}^C$ as the K -th order joint cumulant formalizes the heuristic derivation of the chimeric measure in previous literature. Almost every paper that uses the chimeric epistasis measures ϵ^C does not even mention the joint cumulant. Two notable exceptions are refs. 40,45, which use the joint cumulants to derive formulae for 3-way, 4-way, and 5-way interactions between drugs. However, refs. 40,45 do not rigorously define a probability distribution nor the random variables whose joint cumulant they compute.

At the same time, our formal definition of the chimeric epistasis measure ϵ^C reveals two critical issues with the chimeric formula. First, the assumption that the fitness values f are equivalent to the moments of an MVB random variable is not biologically reasonable for higher-order interactions between three or more loci. This assumption implies that the fitness of a particular genotype depends on the probability of many other genotypes. For example, making this assumption for $L = 4$ loci, the fitness f_{1100} of a double mutant is equal to the moment $E[X_1 X_2]$, which is equal to

$$E[X_1 X_2] = P(X_1 = 1, X_2 = 1) = p_{1100} + p_{1101} + p_{1110} + p_{1111}. \quad (31)$$

However, it is not clear why the fitness f_{1100} of a *single* genotype, 1100, should equal depend on the probabilities of *four different* genotypes, 1100, 1101, 1110, and 1111.

The second issue is that joint cumulants are not necessarily an appropriate measure of higher-order interactions between *binary* random variables. The differences between the joint cumulants and natural parameters β have been previously investigated in the neuroscience literature, as both quantities have been used to quantify higher-order interactions in neuronal data. For example, Staude et al.^{34,35} write that the joint cumulants κ and natural parameters β measure mathematically distinct types of higher-order interactions, and that each quantity may be appropriate for different applications. In particular, Staude et al.^{34,35} note that the joint cumulants measure higher-order interactions between random variables in terms of “additive common components”, while the natural parameters β measure “to what extent the probability of certain binary patterns can be explained by the probabilities of its sub-patterns”. It follows that for binary mutation data, the natural parameters β correspond exactly with the epistasis we aim to measure, i.e., how the fitness of a binary pattern can be explained by the fitness of its “sub-patterns”, while the joint cumulants do not.

Walsh coefficients and background-averaged epistasis. The multivariate Bernoulli distribution also provides a probabilistic interpretation of the *Walsh coefficients* that are used to measure “background-averaged” epistasis^{30–32,42,43}. The Walsh coefficients $\mathbf{u} = [u_{x_1 \dots x_L}] \in \mathbb{R}^{2^L}$,

i.e., a vector indexed by binary strings, are defined by

$$\mathbf{u} = \Psi \mathbf{f} \quad (32)$$

where $\Psi = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{\otimes L} \in \mathbb{R}^{2^L \times 2^L}$ is a *Hadamard matrix*³² and $\mathbf{f} = [f_{x_1 \dots x_L}] \in \mathbb{R}^{2^L}$ is the vector of fitness values indexed by binary strings. Equation (32) is known as the *Walsh-Hadamard transformation*, sometimes also called the *Walsh* or *Fourier-Walsh transform*; see refs. 30,32 for more details.

We prove that if the fitness values \mathbf{f} are equal to probabilities \mathbf{p} of a multivariate Bernoulli random variable (X_1, \dots, X_L) , then the Walsh coefficients \mathbf{u} are equal to the *moments* of $(1 - 2X_1, \dots, 1 - 2X_L) \in \{-1, 1\}^L$, i.e., a linear transformation of the random variable (X_1, \dots, X_L) such that it takes values in $\{-1, 1\}^L$ instead of $\{0, 1\}^L$.

Theorem 2. Let $(X_1, \dots, X_L) \in \{0, 1\}^L$ be distributed according to a multivariate Bernoulli distribution with general parameters \mathbf{f} , and define $Y_\ell = 1 - 2X_\ell \in \{-1, 1\}$ for $\ell = 1, \dots, L$. Define $\mathbf{u} = [u_{x_1 \dots x_L}] \in \mathbb{R}^{2^L}$ as in (32). Then $u_{x_1 \dots x_L} = E[Y_1^{x_1} \dots Y_L^{x_L}]$.

Theorem 2 gives a probabilistic interpretation of the Walsh coefficients \mathbf{u} . Interestingly, the Walsh coefficients \mathbf{u} assume an *additive* fitness model^{30,32} while Theorem 2 requires that the fitness values \mathbf{f} are equal to the probabilities \mathbf{p} , an assumption corresponding to the *multiplicative* fitness model (Table 1).

Relationship to theoretical genetics models. We note that some previous works in theoretical genetics by Barton and Turelli (e.g., ref. 94) also model the genotype with an MVB. However, their approach is substantially different from ours. Barton and Turelli model linkage disequilibrium between k loci X_{i_1}, \dots, X_{i_k} using the k -way central moment $\tilde{\mu}(X_{i_1}, \dots, X_{i_k}) = E\left[\prod_{j=1}^k (X_{i_j} - E[X_{i_j}])\right]$ of the genotype distribution $P(X_1, \dots, X_n)$. Barton and Turelli model epistasis using coefficients that are not related to the genotype distribution $P(X_1, \dots, X_n)$. In contrast, we model epistasis with the natural parameters β of the genotype distribution $P(X_1, \dots, X_n)$, as described in Multiplicative and additive epistasis measures.

Interestingly, Barton and Turelli’s 3-way linkage disequilibrium term, i.e., $\tilde{\mu}(X_{i_1}, X_{i_2}, X_{i_3})$, is equal to the 3-way joint cumulant $\kappa(X_{i_1}, X_{i_2}, X_{i_3})$ implicitly used by Kuzmin et al.^{24,26} to measure 3-way epistasis. This equivalence is because the k -way central moment $\tilde{\mu}(X_{i_1}, \dots, X_{i_k})$ is equal to the k -way joint cumulant $\kappa(X_{i_1}, \dots, X_{i_k})$ for $k = 1, 2, 3$. However, for $k \geq 4$, the k -way linkage disequilibrium term used by Barton and Turelli is not equal to the k -way joint cumulant.

Simulating fitness values

We simulate fitness values $f_{\mathbf{x}}$ for genotypes $\mathbf{x} = (x_1, \dots, x_L)$ with $L = 10$ loci and K -way interactions using the following two different approaches. For both models, we divide all of the fitness values \mathbf{f} by f_{\emptyset} so that $f_{\emptyset} = 1$.

Multiplicative fitness model. We draw interaction parameters $\beta_S \sim \text{Uni}(-0.5, 0.5)$ for each subset $S \subseteq \{1, \dots, L\}$ of loci with size $|S| \leq K$. We set the fitness $f_{\mathbf{x}}$ of genotype $\mathbf{x} = (x_1, \dots, x_L)$ as

$$\log f_{\mathbf{x}} = \left(\sum_{\substack{S \subseteq \{1, \dots, L\} \\ |S| \leq K}} \beta_S \left(\prod_{i \in S} x_i \right) \right) + \epsilon_{\mathbf{x}} \quad (33)$$

where $\epsilon_{\mathbf{x}} \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed Gaussian random variables with mean zero and variance σ^2 . We note that

our noise model ϵ differs from the widely-used Rough Mount Fuji fitness model⁹⁵, where the noise terms are the source of epistasis^{96–98}.

NK model. We simulate fitness values \mathbf{f} according to the NK model with the code used by ref. 31. Because ref. 31 uses an additive fitness model, we exponentiate the fitness values from the NK model.

Epistasis between protein mutations

The analysis in Epistasis between protein mutations was performed using publicly available DMS data for the following proteins/RNA molecules:

- the *E. coli* metabolic protein FoaA⁷², where the fitness of a genotype is measured by the logarithm of its relative frequency in a large population;
- the *Streptococcus pyogenes* Cas9 (SpCas9) nuclease³⁸, where the fitness of a genotype is measured by the logarithm of its relative frequency in a large population;
- the immunoglobulin-binding protein G domain B1 (GB1) expressed in *Streptococcal* bacteria^{73,74}, where the fitness of a genotype is measured by its relative frequency in a large population;
- the Omicron BA.1 variant of the SARS-CoV-2 virus⁷⁵, where the fitness of a genotype is measured by the logarithm of its binding affinity relative to the Wuhan Hu-1 strain;
- the *Entacmaea quadricolor* fluorescent protein (eqFP611)⁷⁶, where the fitness of a genotype is measured by (a normalized version of) its relative frequency in a large population;
- the *Aequorea victoria* green fluorescent protein (avGFP)⁷⁷, where the fitness of a genotype is measured by the logarithm of its fluorescence;
- the green fluorescent proteins (GFPs) from ref. 10, where the fitness of a genotype is measured by its fluorescence;
- yeast tRNA⁷⁸, where the fitness of a genotype is measured by the logarithm of its relative frequency in a large population; and
- the *Chlamydomonas reinhardtii* flavin mononucleotide (FMN)-based fluorescent protein (CreiLOV)⁷⁹, where the fitness of a genotype is measured by the logarithm of its fluorescence.

We aim to use these protein fitness landscapes to directly compare the multiplicative and chimeric epistasis measures. However, the quantitative trait used to measure fitness – and thus the appropriate scale for measuring epistasis (additive or multiplicative) – varies across the different proteins. In particular, the original publications for the FoaA protein, SpCas9 protein, eqFP611 protein, COVID spike protein, avGFP protein, yeast tRNA, and CreiLOV protein assume that fitness is measured using an additive scale (e.g., by measuring fitness as the logarithm of the relative frequency of a genotype). Since our aim is to demonstrate the disagreement between the multiplicative and chimeric measures, we first transform the fitness measurements to a multiplicative scale by exponentiating the fitness values \mathbf{f} , i.e., $f \rightarrow e^f$, before computing the multiplicative measure. This allows us to directly compare the multiplicative epistasis measure ϵ^M with the chimeric epistasis measure ϵ^C , which implicitly assumes fitness values are measured using a multiplicative scale.

For each protein and each interaction order K , we compute the multiplicative (resp. chimeric) measure ϵ^M (resp. ϵ^C) across all K -tuples of mutational events. We note that for some proteins, the fitness of multiple mutations at a given locus is measured (e.g., all 3 possible base-pair substitutions at a locus, or all 19 possible amino acid substitutions); for these proteins, we consider each possible mutation at a given genetic locus as a separate mutational event. Furthermore, we only compute the epistasis measure for a given K tuple of mutational events if all fitness values f for the 2^K genotypes are greater than a threshold ϵ , which we set to $\epsilon = 0.01$ following the convention from ref. 73.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The datasets used in this study were obtained through publicly available repositories. The synthetic genetic array (SGA) data used to analyze three-way epistasis in yeast was obtained from <https://doi.org/10.5061/dryad.g79cnp5m9> and <https://boonelab.ccrb.utoronto.ca/supplement/kuzmin2018/supplement.html>. The drug response data used to analyze higher-order interactions between drug combinations was obtained from the supplementary information of ref. 40. The deep mutational scanning (DMS) data used to analyze higher-order interactions between protein mutations was obtained from: <https://doi.org/10.5281/zenodo.8228919> (FoaA); https://github.com/AWHKU/RunML_DE_SpCas9/tree/main (SpCas9); <https://github.com/J-SNACKKB/FLIP/tree/main/splits/gb1> (GB1); https://github.com/desai-lab/compensatory_epistasis_omicron/tree/main/Titeseq/results/Kds (Omicron BA.1 variant of SARS-CoV-2); the supplementary information of ref. 76 (eqFP611); <https://doi.org/10.6084/m9.figshare.3102154> (avGFP); https://github.com/aequorea238/Orthologous_GFP_Fitness_Peaks (GFPs); the supplementary information of ref. 78 (tRNA); and the supporting information of ref. 79 (CreiLOV). Source data are provided in this paper.

Code availability

The code for our analyses is located in our public [GitHub repository](#) and is available here: 10.5281/zenodo.14426370. Our code requires the following Python packages: Numpy ($\geq 1.23.4$), Matplotlib ($\geq 3.8.0$), Pandas ($\geq 2.1.1$), Scipy ($\geq 1.11.2$), Seaborn ($\geq 0.12.2$).

References

- Wright, S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proceedings Sixth International Congress Genetics* **1**, 356–366 (1932).
- Phillips, P. C. Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* **9**, 855–867 (2008).
- Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218 (2016).
- Wong, A. Epistasis and the evolution of antimicrobial resistance. *Front. Microbiol.* **8**, 246 (2017).
- Cordell, H. J. Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum. Mol. Genet.* **11**, 2463–2468 (2002).
- Lehner, B. Molecular mechanisms of epistasis within and between genes. *Trends Genet.* **27**, 323–331 (2011).
- Domingo, J. Úlia, Baeza-Centurion, P. & Lehner, B. The causes and consequences of genetic interactions (epistasis). *Annu. Rev. Genom. Human Genet.* **20**, 433–460 (2019).
- Krug, Joachim. Epistasis and evolution. *Oxford Bibliographies*, <https://doi.org/10.1093/obo/9780199941728-0137> (2021).
- Chevin, Luis-Miguel On measuring selection in experimental evolution. *Biol. Lett.* **7**, 210–213 (2011).
- Somermeyer, LouisaGonzalez et al. Heterogeneity of the gfp fitness landscape and data-driven protein design. *Elife* **11**, e75842 (2022).
- Russ, D. & Kishony, R. Additivity of inhibitory effects in multidrug combinations. *Nat. Microbiol.* **3**, 1339–1345 (2018).
- VanderWeele, T. J. & Knol, M. J. A tutorial on interaction. *Epidemiol. Methods* **3**, 33–72 (2014).
- Wagner, Günter P. Two rules for the detection and quantification of epistasis and other interaction effects. *Methods Mol. Biol.* 145–157, https://doi.org/10.1007/978-1-4939-2155-3_8 (2015).
- Elena, S. F. & Lenski, R. E. Test of synergistic interactions among deleterious mutations in bacteria. *Nature* **390**, 395–398 (1997).

15. Segrè, D., DeLuna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism. *Nat. Genet.* **37**, 77–83 (2005).
16. Schuldiner, M. et al. Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
17. St Onge, R. P. et al. Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* **39**, 199–206 (2007).
18. Mani, R., Onge, Robert P. St, Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl. Acad. Sci. USA* **105**, 3461–3466 (2008).
19. Beerenwinkel, N., Pachter, L. & Sturmfels, B. Epistasis and shapes of fitness landscapes. *Stat. Sin.* **17**, 1317–1342 (2007).
20. Chou, Hsin-Hung, Chiu, Hsuan-Chao, Delaney, N. F., Segrè, D. & Marx, C. J. Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
21. Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).
22. Breslow, D. K. et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* **5**, 711–718 (2008).
23. Baryshnikova, A. et al. Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7**, 1017–1024 (2010).
24. Kuzmin, E. et al. Systematic analysis of complex genetic interactions. *Science* **360**, eaao1729 (2018).
25. De Visser, J. Arjan G. M., Cooper, T. F. & Elena, S. F. The causes of epistasis. *Proc. R. Soc. B Biol. Sci.* **278**, 3617–3624 (2011).
26. Kuzmin, E. et al. Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science* **368**, eaaz5667 (2020).
27. Tekin, E., Yeh, P. J. & Savage, V. M. General form for interaction measures and framework for deriving higher-order emergent effects. *Front. Ecol. Evol.* **6**, 166 (2018).
28. Fowler, D. M. & Fields, S. Deep mutational scanning: a new style of protein science. *Nat. Methods* **11**, 801–807 (2014).
29. Dai, B., Ding, S. & Wahba, G. Multivariate Bernoulli distribution. *Bernoulli* **19**, 1465–1483 (2013).
30. Weinreich, D. M., Lan, Y., Wylie, C. S. & Heckendorn, R. B. Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23**, 700–707 (2013).
31. Weinreich, D. M., Lan, Y., Jaffe, J. & Heckendorn, R. B. The influence of higher-order epistasis on biological fitness landscape topography. *J. Stat. Phys.* **172**, 208–225 (2018).
32. Poelwijk, F. J., Krishna, V. & Ranganathan, R. The context-dependence of mutations: A linkage of formalisms. *PLoS Comput. Biol.* **12**, e1004771–e1004771 (2016).
33. Hu, Sau-Lon James Probabilistic independence and joint cumulants. *J. Eng. Mech.* **117**, 640–652 (1991).
34. Staude, B., Rotter, S. & Grün, S. CuBIC: Cumulant based inference of higher-order correlations in massively parallel spike trains. *J. Comput. Neurosci.* **29**, 327–350 (2010).
35. Staude, B., Sonja, G. & Rotter, S. *Analysis of Parallel Spike Trains*. (2010).
36. Schmiedel, J.örn M. & Lehner, B. Determining protein structures using deep mutagenesis. *Nat. Genet.* **51**, 1177–1186 (2019).
37. Rollins, N. J. et al. Inferring protein 3d structure from deep mutation scans. *Nat. Genet.* **51**, 1170–1176 (2019).
38. Choi, Gigi C. G. et al. Combinatorial mutagenesis en masse optimizes the genome editing activities of spcas9. *Nat. Methods* **16**, 722–730 (2019).
39. Hopf, T. A. et al. Mutation effects predicted from sequence co-variation. *Nat. Biotechnol.* **35**, 128–135 (2017).
40. Tekin, Elif et al. Prevalence and patterns of higher-order drug interactions in Escherichia coli. *Npj Syst. Biol. Appl.* **4**, 31 (2018).
41. Fisher, R. A. Xv.—the correlation between relatives on the supposition of mendelian inheritance. *Earth Environ. Sci. Trans. R. Soc. Edinburgh* **52**, 399–433 (1919).
42. Crona, K., Gavryushkin, A., Greene, D. & Beerenwinkel, N. Inferring genetic interactions from comparative fitness data. *Elife* **6**, e28629 (2017).
43. Crona, K. Rank orders and signed interactions in evolutionary biology. *ELife* **9**, e51004 (2020).
44. Yeh, P., Tschumi, A. I. & Kishony, R. Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* **38**, 489–494 (2006).
45. Lozano-Huntelman, NatalieAnn et al. Hidden suppressive interactions are common in higher-order drug combinations. *iScience* **24**, 102355 (2021).
46. Wagner, G.ünter P. The measurement theory of fitness. *Evolution* **64**, 1358–1376 (2010).
47. Teugels, J. L. Some representations of the multivariate bernoulli and binomial distributions. *J. Multivar. Anal.* **32**, 256–268 (1990).
48. Wainwright, M. J. & Jordan, M. I. Graphical models, exponential families, and variational inference. *Found. Trends® Mach. Learn.* **1**, 1–305 (2008).
49. Cordell, H. J. Detecting gene–gene interactions that underlie human diseases. *Nat. Rev. Genet.* **10**, 392–404 (2009).
50. Kauffman, S. A. & Weinberger, E. D. The nk model of rugged fitness landscapes and its application to maturation of the immune response. *J. Theor. Biol.* **141**, 211–245 (1989).
51. Wright, A. H., Thompson, R. K. & Zhang, J. The computational complexity of nk fitness functions. *IEEE Trans. Evol. Comput.* **4**, 373–379 (2000).
52. Benson, A. R., Gleich, D. F. & Leskovec, J. Higher-order organization of complex networks. *Science* **353**, 163–166 (2016).
53. Yan Tong, A. H. & Boone, C. Synthetic genetic array analysis in saccharomyces cerevisiae. *Methods Mol. Biol.* **313**, 171–192. (2006).
54. Kuzmin, E. et al. τ-sga: synthetic genetic array analysis for systematically screening and quantifying trigenic interactions in yeast. *Nat. Protoc.* **16**, 1219–1250 (2021).
55. Kafri, R., Springer, M. & Pilpel, Y. Genetic redundancy: New tricks for old genes. *Cell* **136**, 389–392 (2009).
56. DeLuna, A., Springer, M., Kirschner, M. W., & Kishony, R. Need-based up-regulation of protein levels in response to deletion of their duplicate genes. *PLoS Biol.* **8**, <https://doi.org/10.1371/journal.pbio.1000347> (2010).
57. DeLuna, A. et al. Exposing the fitness contribution of duplicated genes. *Nat. Genet.* **40**, 676–681 (2008).
58. Rajvanshi, Praveen Kumar, Arya, M. & Rajasekharan, R. The stress-regulatory transcription factors msn2 and msn4 regulate fatty acid oxidation in budding yeast. *J. Biol. Chem.* **292**, 18628–18643 (2017).
59. Yang, H. et al. Sterol esterification in yeast: a two-gene process. *Science* **272**, 1353–1356 (1996).
60. Steffen, K. K. et al. Ribosome deficiency protects against er stress in saccharomyces cerevisiae. *Genetics* **191**, 107–118 (2012).
61. Lin, D. H. et al. Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016).
62. Wente, S. R. & Blobel, G.ünter NUP145 encodes a novel yeast glycine-leucine-phenylalanine-glycine (GLFG) nucleoporin required for nuclear envelope structure. *J. Cell Biol.* **125**, 955–969 (1994).
63. Lei, E. P. et al. Sac3 is an mRNA export factor that localizes to cytoplasmic fibrils of nuclear pore complex. *Mol. Biol. Cell* **14**, 836–847 (2003).
64. Van Hoof, A. Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171**, 1455–1461 (2005).
65. Marshall, A. N., Montealegre, M. C., Jiménez-López, C., Lorenz, M. C., & van Hoof, A. Alternative splicing and subfunctionalization

- generates functional diversity in fungal proteomes. *PLoS Genet.* **9**, <https://doi.org/10.1371/journal.pgen.1003376> (2013).
66. Doma, M. K. & Parker, R. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. *Nature* **440**, 561–564 (2006).
 67. Tsuboi, T. et al. Dom34: Hbs1 Plays a general role in quality-control systems by dissociation of a stalled ribosome at the 3' end of aberrant mRNA. *Mol. Cell* **46**, 518–529 (2012).
 68. Horikawa, W., Endo, K., Wada, M. & Ito, K. Mutations in the G-domain of Ski7 cause specific dysfunction in non-stop decay. *Sci. Rep.* **6**, 1–10 (2016).
 69. Manning, B. D., Barrett, J. G., Wallace, J. A., Granok, H. & Snyder, M. Differential regulation of the Kar3p Kinesin-related protein by two associated proteins, Cik1p and Vik1p. *J. Cell Biol.* **144**, 1219–1233 (1999).
 70. Lyke, D. R., Dorweiler, J. E. & Manogaran, A. L. The three faces of Sup35. *Yeast* **36**, 465–472 (2019).
 71. Baeder, Desiree Y., Yu, Guozhi, Hozé, Nathanaël, Rolff, Jens, and Regoes, Roland R. Antimicrobial combinations: Bliss independence and loewe additivity derived from mechanistic multi-hit models. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, <https://doi.org/10.1098/rstb.2015.0294> (2016).
 72. Papkou, A., Garcia-Pastor, L., Escudero, JoséAntonio & Wagner, A. A rugged yet easily navigable fitness landscape. *Science* **382**, eadh3860 (2023).
 73. Wu, N. C., Dai, L., Olson, C. A., Lloyd-Smith, J. O. & Sun, R. Adaptation in protein fitness landscapes is facilitated by indirect paths. *Elife* **5**, e16965 (2016).
 74. Wu, Z., Kan, S. B., Jennifer, Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci. USA* **116**, 8852–8858 (2019).
 75. Moulana, A. et al. Compensatory epistasis maintains ace2 affinity in sars-cov-2 omicron ba. 1. *Nat. Commun.* **13**, 7011 (2022).
 76. Poelwijk, F. J., Socolich, M. & Ranganathan, R. Learning the pattern of epistasis linking genotype and phenotype in a protein. *Nat. Commun.* **10**, 4213 (2019).
 77. Sarkisyan, K. S. et al. Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).
 78. Domingo, J. Úlia, Diss, G. & Lehner, B. Pairwise and higher-order genetic interactions during the evolution of a trna. *Nature* **558**, 117–121 (2018).
 79. Chen, Y. et al. Deep mutational scanning of an oxygen-independent fluorescent protein creilov for comprehensive profiling of mutational and epistatic effects. *ACS Synth. Biol.* **12**, 1461–1473 (2023).
 80. De Visser, J. ArjanG. M. & Krug, J. Empirical fitness landscapes and the predictability of evolution. *Nat. Rev. Genet.* **15**, 480–490 (2014).
 81. Tang, D., Freudenberg, J. & Dahl, A. Factorizing polygenic epistasis improves prediction and uncovers biological pathways in complex traits. *Am. J. Hum. Genet.* **110**, 1875–1887 (2023).
 82. Fu, B. et al. Fast kernel-based association testing of non-linear genetic effects for biobank-scale data. *Nat. Commun.* **14**, 4936 (2023).
 83. Otwinowski, J. & Plotkin, J. B. Inferring fitness landscapes by regression produces biased estimates of epistasis. *Proc. Natl. Acad. Sci. USA* **111**, E2301–E2309 (2014).
 84. Weinreich, D. M., Watson, R. A. & Chao, L. Perspective: sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
 85. Beerwinkel, N., Pachter, L., Sturmfels, B., Elena, S. F. & Lenski, R. E. Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evol. Biol.* **7**, 1–12 (2007).
 86. Phillips, Patrick C et al. *Beyond the Average: The Evolutionary Importance of Gene Interactions and Variability of Epistatic Effects*. 20–38 (2000).
 87. Gao, H., Granka, J. M. & Feldman, M. W. On the classification of epistatic interactions. *Genetics* **184**, 827–837 (2010).
 88. Fisher, Ronald Aylmer. *The Genetical Theory of Natural Selection: A Complete Variorum Edition*. (Oxford University Press, 1999).
 89. Haldane, J. B. S. The effect of variation of fitness. *Am. Nat.* **71**, 337–349 (1937).
 90. Shnol, E. E. & Kondrashov, A. S. The effect of selection on the phenotypic variance. *Genetics* **134**, 995–996 (1993).
 91. Barndorff-Nielsen, Ole. *Information and Exponential Families: In Statistical Theory*. John Wiley & Sons, (2014).
 92. Weinberger, E. D. Fourier and Taylor series on fitness landscapes. *Biol. Cybern.* **65**, 321–330 (1991).
 93. Prabhu, S. & Pe'er, I. Ultrafast genome-wide scan for snp-snp interactions in common complex disease. *Genome Res.* **22**, 2230–2240 (2012).
 94. Barton, N. H. & Turelli, M. Natural and sexual selection on many loci. *Genetics* **127**, 229–255 (1991).
 95. Neidhart, J., Szendro, I. G. & Krug, J. Adaptation in tunably rugged fitness landscapes: the rough mount fuji model. *Genetics* **198**, 699–721 (2014).
 96. Wood, K., Nishida, S., Sontag, E. D. & Cluzel, P. Mechanism-independent method for predicting response to multidrug combinations in bacteria. *Proc. Natl. Acad. Sci. USA* **109**, 12254–12259 (2012).
 97. Zimmer, A., Katzir, I., Dekel, E., Mayo, A. E. & Alon, U. Prediction of multidimensional drug dose responses based on measurements of drug pairs. *Proc. Natl. Acad. Sci. USA* **113**, 10442–10447 (2016).
 98. Zhao, W. et al. A new bliss independence model to analyze drug combination data. *J. Biomol. Screening* **19**, 817–821 (2014).

Acknowledgements

We thank Daniel Weinreich, Sriram Sankararaman, and Boyang Fu for their helpful comments and suggestions. U.C. was supported by a National Science Foundation Graduate Research Fellowship and the Siebel Scholars program. B.J.A. gratefully acknowledges financial support from the Schmidt DataX Fund at Princeton University, made possible through a major gift from the Schmidt Futures Foundation. This research is also supported by National Cancer Institute (NCI) grants U24CA248453 and U24CA264027 to B.J.R.

Author contributions

Conceptualization: B.J.R., U.C., and B.A.; methodology: U.C. and B.J.R.; software: U.C. and B.A.; validation: B.A.; formal analysis: U.C., B.A., and B.J.R.; investigation: U.C., B.A., and B.J.R.; data curation: U.C. and B.A.; writing—original draft: U.C., B.A., and B.J.R.; writing—review and editing: U.C., B.A., and B.J.R.; visualization: U.C. and B.A.; supervision: B.J.R.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-56986-5>.

Correspondence and requests for materials should be addressed to Benjamin J. Raphael.

Peer review information *Nature Communications* thanks Kristina Crona, Suman Das, and the other anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025