



Master regulators of biological systems in higher dimensions

Holger Eble^a, Michael Joswig^{a,b,1} , Lisa Lambert^{c,d}, and William B. Ludington^{e,f,1}

Edited by Eugene I. Shakhnovich, Harvard University, Cambridge, MA; received January 12, 2023; accepted October 23, 2023 by Editorial Board Member Robert Tycko

A longstanding goal of biology is to identify the key genes and species that critically impact evolution, ecology, and health. Network analysis has revealed keystone species that regulate ecosystems and master regulators that regulate cellular genetic networks. Yet these studies have focused on pairwise biological interactions, which can be affected by the context of genetic background and other species present, generating higher-order interactions. The important regulators of higher-order interactions are unstudied. To address this, we applied a high-dimensional geometry approach that quantifies epistasis in a fitness landscape to ask how individual genes and species influence the interactions in the rest of the biological network. We then generated and also reanalyzed 5-dimensional datasets (two genetic, two microbiome). We identified key genes (e.g., the *rhs* locus and *pykF*) and species (e.g., *Lactobacilli*) that control the interactions of many other genes and species. These higher-order master regulators can induce or suppress evolutionary and ecological diversification by controlling the topography of the fitness landscape. Thus, we provide a method and mathematical justification for exploration of biological networks in higher dimensions.

epistasis | higher-order interaction | fitness landscape | microbiome | lifespan

Master regulators are nodes in a network that control the rest of the network. They are often identified as highly connected nodes. For example, in eukaryotic cells, the protein, target of rapamycin (TOR), interacts with many other proteins and pathways to control cellular metabolism (1), which is important for understanding cellular metabolism as well as for treating diseases including cancer, autoimmunity, and metabolic disorders (1). Ecological master regulators are called keystone species, a classical example being the starfish, *Pisaster*, which regulates the biodiversity of the intertidal zone by eating many other species (2). Identifying key nodes in biological networks provides control points that further basic science and enable applications, for instance, cancer therapy (through TOR) or ecological restoration (through starfish).

Epistasis is a standard framework to quantify biological networks, specifically gene networks, where the genes are nodes and interactions are specified by edges. Constructing a gene network using epistasis works by iteratively mutating a set of individual genes and pairs of these genes, and then using the phenotypes of the mutants to construct the network. For instance, if genes *A* and *B* both affect a phenotype, *C*, we make the single mutants *a* and *b* and the double mutant *ab* and denote the phenotypes of these mutants as C_a , C_b , and C_{ab} . By measuring the effects on a phenotype, e.g., fitness, it can be determined if *A* and *B* operate in parallel to affect *C* (i.e., $A \rightarrow C$ and $B \rightarrow C$) or in serial (i.e., $A \rightarrow B \rightarrow C$). These two possibilities are differentiated based on the degree of non-additivity: If the C_a and C_b add up to C_{ab} , the genes do not interact and thus operate in parallel. If they are non-additive, the genes interact and thus operate in serial. More specifically, if $A \rightarrow B \rightarrow C$, then mutants *a*, *b*, and *ab* will each produce the same phenotype, thus, $C_a + C_b \neq C_{ab}$, indicating non-additivity or epistasis.

Applying epistasis to genome-wide measurement of pairwise genetic interactions has revealed biochemical pathways composed of discrete sets of genes (3, 4) as well as complex traits, such as human height, that are affected by almost every gene in the genome (5, 6). New innovations have applied epistasis to broader data types (7, 8) and at different scales, making epistasis a widely valuable tool. The concept has been used to map pairwise interactions for protein structure (9), genetics (3, 4, 7, 10, 11), microbiomes (12, 13), and ecology (14–16).

Epistatic interactions are important in nature (17), for instance, when mutations occur (18–20) or when sex, recombination, and horizontal gene transfer bring groups of genes together (11, 21–25), making multiple loci interact. Epistasis between bacteria in the microbiome has functional consequences (12, 13, 26–29) when community assembly

Significance

Some parts of biological networks exert more regulatory control than others. Identifying these is essential to understanding biology. For instance, master regulator genes control essential cellular processes in metabolism and development, and keystone species control ecosystem stability. These key regulators have more direct, pairwise interactions with other genes or species than average, but this approach misses interactions that change in different contexts. Because context-dependent effects are prevalent in biology, we developed an approach that identifies regulators of interactions in the entire network. The approach uses the mathematical concept of epistasis on fitness landscapes. This approach reveals master regulator genes and keystone microbiome species that affect evolutionary trajectories and lifespan of the host, respectively.

Author contributions: H.E., M.J., L.L., and W.B.L. designed research; performed research; contributed new reagents/analytic tools; analyzed data; and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. E.I.S. is a guest editor invited by the Editorial Board.

Copyright © 2023 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

¹To whom correspondence may be addressed. Email: joswig@math.tu-berlin.de or ludington@carnegiescience.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2300634120/-DCSupplemental>.

Published December 14, 2023.

combines groups of species in a fecal transplant. In this case, the nodes in the network are bacterial species instead of genes.

Master regulators of biological networks are identified by their position in the network, often as nodes with a higher degree of edges than average (30). A known challenge of biological networks is that they are high dimensional. Consequently, the interactions can change depending on the biological context or the genetic background (31); c.f. ref. 32 and references therein. This is important because such networks cannot be fully captured by pairwise interactions. Higher-order epistatic interactions are interactions that require three or more interacting parts. From a network standpoint, such loci that affect the interactions between many other loci play a key role in regulation of network structure.

Identifying such regulators requires a high-dimensional formulation of epistatic interactions. Fitness landscapes are one such approach (33, 34). They depict biological fitness as a function of genotype space (18, 19, 35). Wright defined the genotype space as a hypercube with each genetic locus represented as an independent dimension (35). Previous work formalized the fitness landscape of this genotype space and quantified epistasis on the fitness landscape (23, 36–38). We developed the *epistatic filtration* technique, which provides a global measure of interactions in higher dimensions.

Here, we develop that framework further in order to apply it to identify master regulators of high-dimensional interactions. Rather than the traditional approach of assigning significance to a gene or species based on its pairwise interactions (2, 3, 39–41), we assign significance based on how the presence of that gene or species influences the structure and magnitude of interactions in the rest of the network. In order to compare interaction magnitudes across different dimensions, we develop a dimensionally normalized definition of epistasis. We also develop a graphical approach to determine whether high-dimensional epistasis has lower-dimensional roots and what they are. We then analyze four datasets that are 5 dimensional. Two are genetic datasets for i) mutations that arose in *E. coli* evolution (42), herein called *Ecoli_{evo}*, and ii) β -lactamase antibiotic resistance (43), herein called *Ecoli_{plac}*. Two are microbiome datasets measuring the impact of bacterial interactions on *Drosophila* lifespan, with one previously published (12), herein called *Dmel_{Gould}* after the lead author, and another generated here, called *Dmel_{Eble}*. Our framework identifies regulators of higher-dimensional network structure in both the genetics and microbiome datasets. We find that specific genes and bacterial species change interactions in the rest of the network, meaning they regulate the higher-order network structure.

Results

Epistatic Filtrations Describe Higher-Dimensional Biological Networks. Our goal is to identify master regulators of biological interactions in higher dimensions. Our approach is to first measure epistasis on the high-dimensional fitness landscape and then ask how individual loci, e.g., genes, change the shape of the landscape. We use the epistatic filtration technique to quantify epistasis on the fitness landscape. We use parallel epistatic filtrations to quantify the changes in the landscape due to each locus.

First, we describe epistatic filtrations. Epistatic filtrations are analogous to analyzing the drainage sectors within a watershed (Fig. 1), which is a real physical landscape with altitude as a

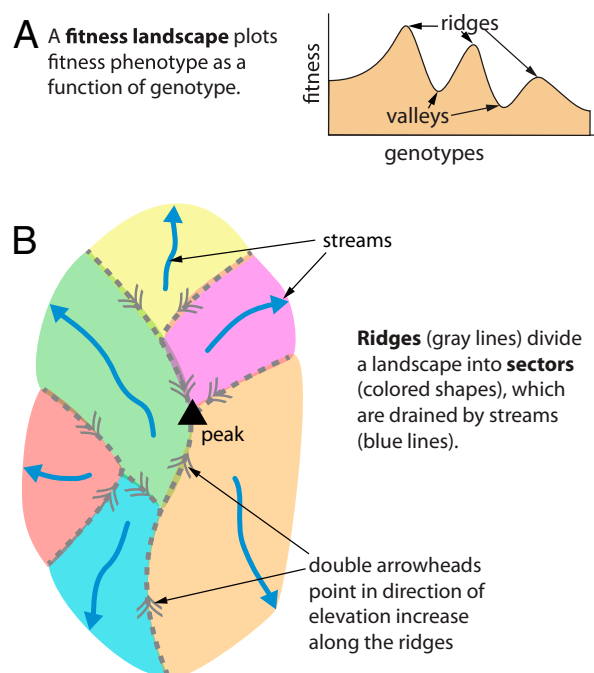


Fig. 1. Landscapes can be divided into sectors based on the ridges. (A) A fitness landscape. (B) Ridges divide a physical landscape into sectors (colored regions). Adjacent sectors share a common ridge.

function of latitude and longitude. Boundaries of a sector are set by ridges, which determine where water will flow.

We can think of a fitness landscape as having sectors as well. In a fitness landscape, the ridges are set not by altitude but by measurements of organismal fitness as a function of genotype. The latitude and longitude of a watershed correspond to genotypes in the fitness landscape and the altitudes to phenotype. However, unlike latitude and longitude, which are continuous, genes are discrete (i.e., a gene is either wild type or mutant). Our framework is discrete too.

We represent each gene with a separate dimension as proposed by Wright (35); if the gene is wild type, its position is 0; if the gene is mutant, its position is 1. The space of all genotypes has many dimensions, one per mutated gene (32, 35). This high-dimensional space is called a *genotype hypercube* (18, 19, 35).

We next quantify the epistasis of the fitness landscape. This requires that we define sets of genotypes to compare. We do so by segmenting the genotype cube into sectors (Fig. 2). Each sector contains information about the steepness of the fitness landscape at that location.

This approach is different from a previous approach that defined epistasis based on paths, called circuits, that traverse the landscape (37). An advantage of our approach is that there are orders of magnitude fewer sectors in a landscape than there are circuits, reducing the search space and the associated statistical constraints from multiple testing comparisons (36).

The sectors are sets of adjacent genotypes in the hypercube. Geometrically speaking, the shape of a sector is a simplex, meaning each vertex (genotype) is directly connected to every other vertex in the set. For instance, a 2D simplex is a triangle. To segment the fitness landscape into sectors, we use a triangulation. In Fig. 3 A–C, we illustrate how a two-dimensional fitness landscape is triangulated using the phenotypes of the genotypes. The phenotypes form a third dimension that we depict as a

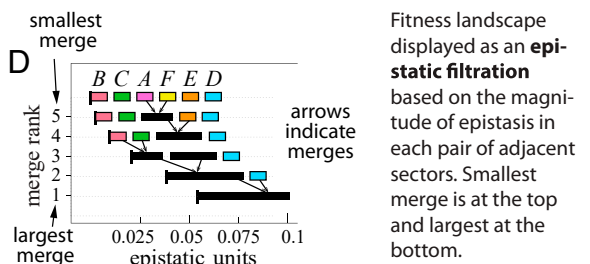
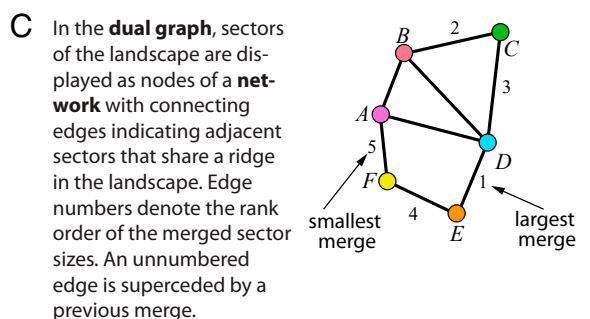
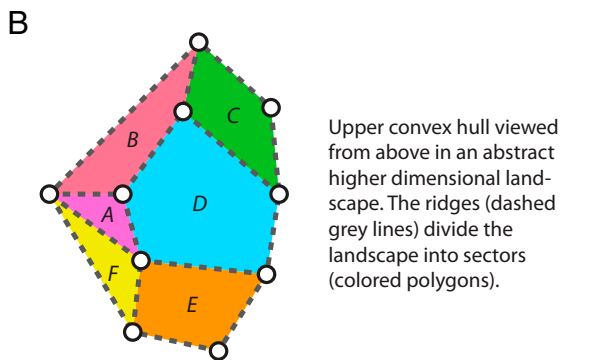
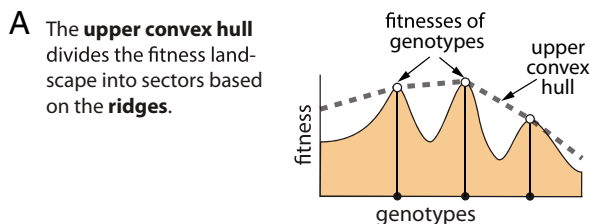


Fig. 2. Conceptual introduction to epistatic filtrations. An epistatic filtration depicts the epistasis of a fitness landscape. By analogy with a watershed, producing the filtration can be conceptualized in four steps: (A) The fitness landscape defines the topography based on the ridges; (B) the landscape is segmented into sectors based on these ridges, which connect adjacent sectors; (C) epistasis is calculated as the shared area of adjacent sectors and displayed on a dual graph, which depicts the adjacency relationships of sectors; (D) the epistatic filtration depicts the rank order of epistasis magnitude in the adjacent sectors as a set of merges. The black tick marks on the left-most sector of each row indicate the magnitude of the merge. Widths of sectors and gaps between them carry no magnitude information. Formal definitions follow in Fig. 3 and *SI Appendix, Text*.

height function on the vertical axis. The phenotype data uniquely determine the ridges of the landscape. Projecting these ridges back to the $2D$ genotype plane forms a triangulation of the genotypes, which delineates the sectors. This diagram is similar to previous illustrations of epistasis on a two-dimensional landscape (c.f. refs. 31 and 32), but our approach is unique in that we use the geometry to sector the fitness landscape.

Next, we construct a network representation of the sectorized genotype space to depict the pairwise adjacency of neighboring simplices (nodes) (36). These simplices are groups of genotypes. An edge in this network indicates that two simplices are adjacent, meaning they share a face. This shape with two simplices sharing a common base but having separate apexes is called a *bipyramid*. Thus, the edges in our network correspond to bipyramids.

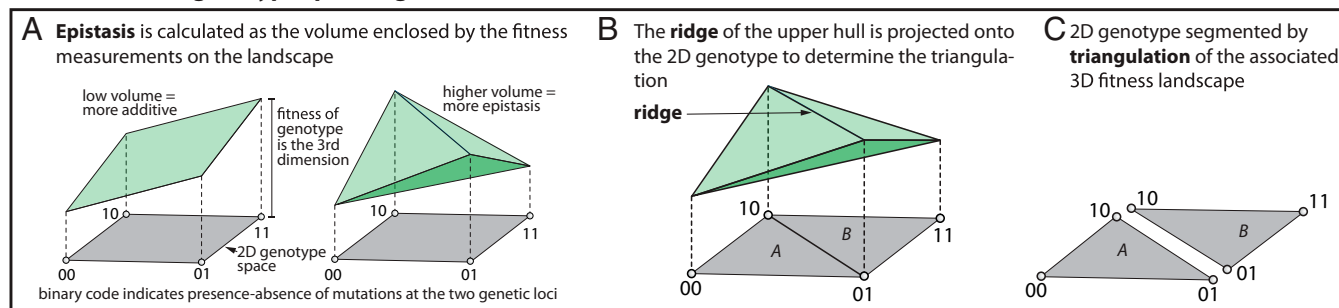
We calculate the magnitude of epistasis of each pair of adjacent sectors in the triangulation by calculating the volume spanned by the fitness phenotypes that correspond to the genotypes of the vertices of the bipyramids. This definition of epistasis is unique yet consistent with previous ones in lower dimensions (Fig. 3). We additionally normalize the fitness phenotypes so that we can compare epistasis between different dimensions and different datasets. Scaling the phenotype height function h by a positive constant does not change the regular triangulation, and thus, it does not change the dual graph. To normalize, we read the height function h as a vector of length 2^D , one for each vertex of the D -dimensional hypercube, and rescale to Euclidean norm 1. Effectively, this amounts to reading the height function as a direction in 2^D -dimensional space. The epistatic volumes are rescaled accordingly (*SI Appendix, section 1*), allowing direct comparisons of epistasis magnitude across dimensions. Our definition of epistasis makes the framework self-consistent when applying it to higher dimensions.

We next rank the volumes of the bipyramids from smallest to largest. These merged sectors can be visualized as a network graph, where the nodes are sectors and the edges are bipyramids. Because there are multiple adjacent sectors, we denote the first merge of a sector as a *critical edge*, consistent with Morse theory (44). If an edge joins a cluster of simplices in the graph due to a separate critical edge, this edge that was brought along by the critical edge is called *noncritical*. Biologically, a critical edge supports gene flow on a fitness landscape. Plotting the rank order of the critical edges gives an epistatic filtration (Figs. 2 and 3). Together, the network graph and the filtration represent the increasing magnitude of epistasis and the connectivity between the regions of the epistatic landscape where the significant epistasis occurs, which is important because it helps identify sets of interacting genotypes. We note that merging small sectors first is a choice based on Theorem 8 in ref. 36 that emphasizes the finer details of the landscape in the visualization by epistatic filtration. The merge order can be reversed to give greater influence to larger sectors in determining the filtration order.

To determine how an individual locus, e.g., gene or species, affects the interactions in the rest of the network, we compare the epistasis for each pair of adjacent sectors with the locus of interest added or removed. This *parallel filtration* quantifies how adding or removing a locus affects the epistasis of the individual sectors on the high-dimensional fitness landscape (36). Discovering loci that have out-sized effects on the fitness landscape allows a new approach to identify master regulators that operate in higher dimensions.

A Volume-Based Definition of Epistasis Is Valid across Many Dimensions. In this section, we explain the $2D$ genotype case of the definition of epistasis that we employ throughout. With two loci and two alleles (0 or 1) at each locus, we plot the genotypes as a unit square in the x - y plane and the measured phenotypes of each genotype on the z -axis (Fig. 3A). The phenotypes thus lift the genotypes into one dimension higher, here going from $2D$ to $3D$. Connecting the four phenotypes gives a simplex, shown as the green polytope in Fig. 3A. Depending on the relative magnitudes

2-locus bi-allelic genotype space segmentation



3-locus bi-allelic genotype space segmentation and epistatic filtration calculation

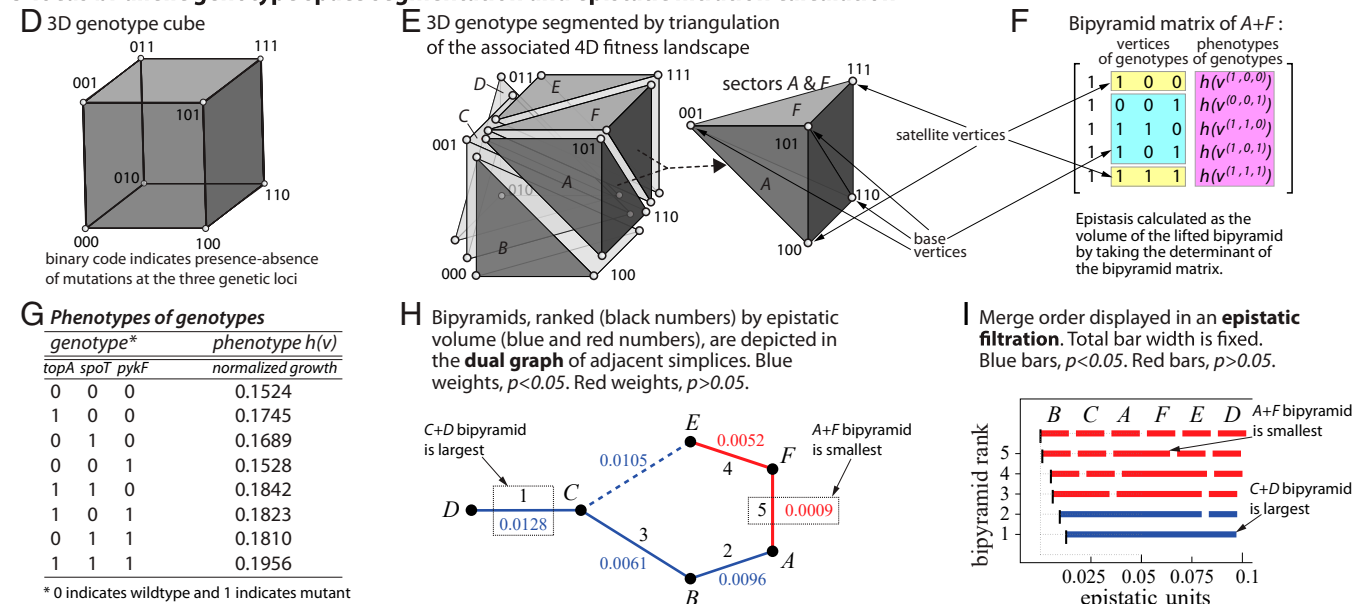


Fig. 3. Definition of epistasis and how triangulation is used to segment the fitness landscape. (A) The 2D genotype set has two loci, each of which can be 0 or 1: {00, 01, 10, 11}. Each genotype gets *lifted* into 3D space by appending the phenotype $h(v)$ to each genotype coordinate in the set, $v \in \{(00), (01), (10), (11)\} \subset \mathbb{R}^2$. Connecting these lifted phenotype points forms a *convex hull*, depicted as the green 3D body $G^{(3)}$ above the gray genotype set. The Euclidean volume of the 3D body $G^{(3)}$ yields a measure for epistasis (c.f. ref. 36). A lower degree of epistasis produces a lower volume of the green body, and higher epistasis produces a larger volume. (B) The upper surface of the green body is two green triangles, which are divided by the *ridge*. The ridge sets a triangulation of the genotype space in gray. This is done by removing the phenotype dimension from the ridge vertices, which projects it back to the 2D genotype space. (C) The ridge thus splits the space into sectors, which are two adjacent triangles, {00, 01, 10}, and {01, 10, 11}, denoted as A and B. We note that in two dimensions, our volumetric definition is equivalent to the absolute value of the established formula $\epsilon = h(00) + h(11) - (h(10) + h(01))$ for epistasis, scaled by a dimension related constant factor. This approach also works in higher dimensions, with the triangles becoming simplices. (D) For the 3D case, the genotype set forms a cube. (E) Using the phenotypes from the first three loci of the *E. coli* dataset, the ridges produce a *regular triangulation*, the subdivision S , which consists of the six tetrahedra, A, B, C, D, E, and F. Different data would produce a different triangulation. Epistasis is calculated from the union of adjacent tetrahedra, for example, sectors A and F, which share a face. The vertices of the shared face are called *base vertices*. The unshared vertices of the two tetrahedra are called *satellites*. (F) The lifted *bipyramid* with its corresponding fitness values is represented in matrix form, which allows us to calculate the volume by taking its determinant. (G) Phenotypes of the 3D genotype cube taken from the *E. coli* dataset. (H) The adjacency relations of the tetrahedra give rise to a network, which is the *dual graph* of S . The edge 5 refers to the bipyramid comprised of A and F with vertices {100} + {001, 110, 101} + {111} [2]. The set {001, 101, 110} is the base where A and F meet, and it separates the two satellites 100 and 111. The edge (C, E) (dotted) is non-critical (see definitions in *SI Appendix, Text*). (I) The *epistatic filtration* of the genotype-phenotype map depicts the iterative process of gluing bipyramids in a non-redundant manner, going from lowest to highest epistatic volume. For example, rank 5 is the merge between A and F and has the lowest epistasis, rank 4 is the merge between E and F, and so forth. The black vertical tick mark at the left end of each row of blocks gives the epistasis added to the filtration at that rank.

of the phenotypes, the green simplex can be larger or smaller, with the perfectly additive (no epistasis) case giving zero volume. We define epistasis as the Euclidean volume of the green simplex. In the 2D case, this is proportional to the absolute value of the established formula for epistasis, $\epsilon = h(00) + h(11) - (h(10) + h(01))$ (37). We call our definition the *epistatic volume* and note that it is of one dimension higher than the genotype space due to the measured phenotype (Fig. 3). This definition of epistasis based on volume is important because it applies equally well in higher dimensions, as we discuss in *SI Appendix*: “A primer on epistatic filtrations” and “Epistatic filtrations: the n -locus case.”

Epistatic Filtrations Reveal Higher-Order Structure in *E. coli* Evolution. To illustrate our approach, we examined *E. coli* evolution, an existing dataset from Lenski's (45) classic experimental evolution of *E. coli* with each combination of the first five beneficial mutations to fix in the population (42) (Fig. 4A). The key finding was that epistatic interactions reduce the overall fitness benefit expected from additivity of the individual mutations (42), with the potential to slow the rate of adaptation (46). But an exception to this rule was that a single locus, pyruvate kinase (*pykF*), which is critical for control of metabolic flux, showed an increase in fitness benefit with other mutations. A remaining question is

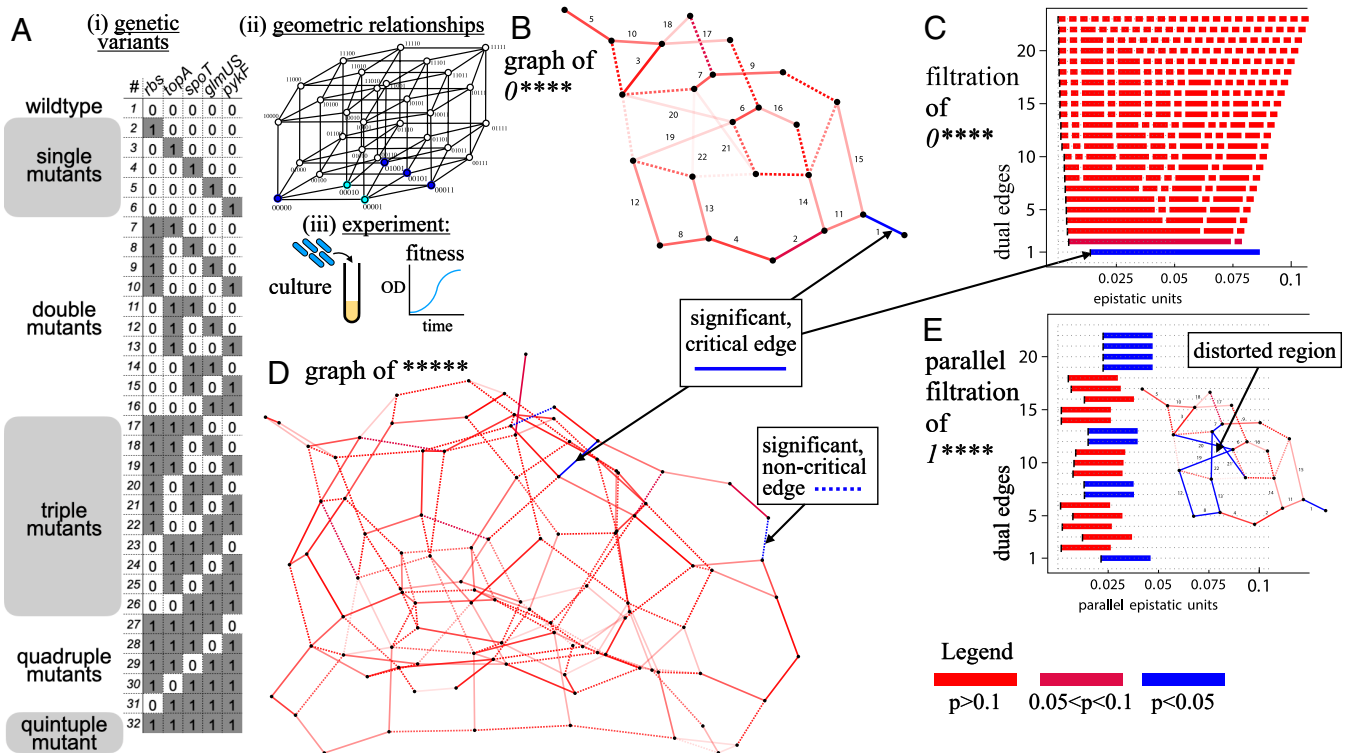


Fig. 4. *E. coli* evolution is guided by epistatic landscape distortions. (A) (i) *E. coli* mutants examined (42), (ii) their geometric relationships, and (iii) experimental approach to measure fitness. In (ii), light blue vertices are the satellites and dark blue are the base for the significant bipyrmaid indicated in B, C, and D. (B) Edge labeled dual graph and (C) epistatic filtration restricted to $n = 4$ mutations in *topA* (locus 2), *spoT* (locus 3), *glmUS* (locus 4) and *pykF* (locus 5). Locus 1, *rbs*, is fixed 0 (wild type). Note that the left edge of the bars in (C) indicates there is very little epistatic volume added to the filtration except for the final merge, where the single genotype 00001 gives weight to the entire filtration. This final interaction corresponds to the vertices {00001} + {00000, 01001, 00101, 00011} + {00010}. (D) Dual graph for the complete *E. coli* dataset. Black indices in (B) label the critical edges of the dual graph of $S(h)$. (E) In the parallel filtration, for 1****, where the *rbs* mutation is present, the landscape is distorted by a concentrated area of higher epistasis. Inset: graph in (B) recolored with weights from (E). The lengths of the bars in the parallel transport figure (E) have no meaning. Only the horizontal position of the black marks, the vertical position of the bars, and its coloring encode information. The horizontal shift represents the value of the epistatic volume, the vertical position of the bar indicates which dual edge is transported and the color expresses if the epistatic volume is significant after parallel transport.

why bacteria in stable habitats continuously evolve even after reaching a near optimum fitness (47).

We first examine $n = 3$ loci, corresponding to mutations in *topA*, *spoT*, and *pykF*. Epistasis was generally low in magnitude (42, 48) and occurs in two ways: i) either from merging groups of simplices (c.f. *BC* + *AFE* in line #3 of Fig. 3I), or ii) from merging a single simplex, c.f. *D*, with the group of the rest of the simplices (c.f. line #2 of Fig. 3I). Geometrically, this merge constitutes a vertex split (49). Biologically, the significant sector is evidence that this set of loci has a strong epistatic interaction.

We next add a fourth mutation, in the *glmUS* locus (Fig. 4 B and C), encoding peptidoglycan availability, which is an essential component of the cell wall. The filtration reveals an additive landscape with one dominant sector where epistasis arises only in the final merge of the filtration (Fig. 4C), meaning the epistatic topography of the entire landscape (Fig. 4D) rests upon the single vertex, 00001, *pykF*. Consistent with the published analysis, which detected a significant, marginal effect of *pykF* (42), filtrations reveal the geometric structure in terms of which specific combinations of loci are responsible for the effect (Fig. 4E): We establish an interaction between *glmUS*, {00010}, and *pykF*, {00001}. The interaction depends on the genotypes {00000, 01001, 00101, 00011} in the bipyrmaid base (Fig. 4 A, ii).

Interestingly, the four locus context involves genotypes with the wild type and only up to double mutants. But these double mutants must be present together to yield a higher dimensional

interaction. This conclusion is consistent with recent genome-wide work on trans-gene interactions (5), suggesting that complex traits may arise from genome-wide sign epistasis, where each mutation's contribution to the trait depends on the presence of other mutations. Additionally, we observe that the interaction of {00001}, {00000, 01001, 00101, 00011}, {00010} in the 4D case (with the first locus wild type) remains significant in the full 5-locus setting: See the blue critical edge in the dual graph of Fig. 4D, which indicates that this interaction in lower dimensions is unaffected when the mutation is introduced in the first locus.

Parallel Epistatic Filtrations Reveal Master Regulators in *E. coli* Evolution. To discern the role of each locus on the 4D network structure, we applied parallel filtrations (Fig. 5) (36, section 6.6). This technique measures context-dependence in the fitness landscape by assessing how the epistasis of sectors changes when a particular locus is mutated versus wild type. Changes in the fitness landscape induced by a single locus are important because they identify master regulators. To discuss parallel filtrations, we introduce the notation, *, to indicate that a locus is variable in the analysis as opposed to being fixed. For example, the epistatic filtration can be calculated for 0****, where the first locus is fixed as wild type and the filtration is performed for the remaining 4 loci, which are varied in the analysis. This yields a set of bipyrmaids for which the epistasis is calculated. In the parallel filtration, we compare the epistasis for 0**** with the epistasis for 1**** using the bipyrmaids set by 0**** as well as

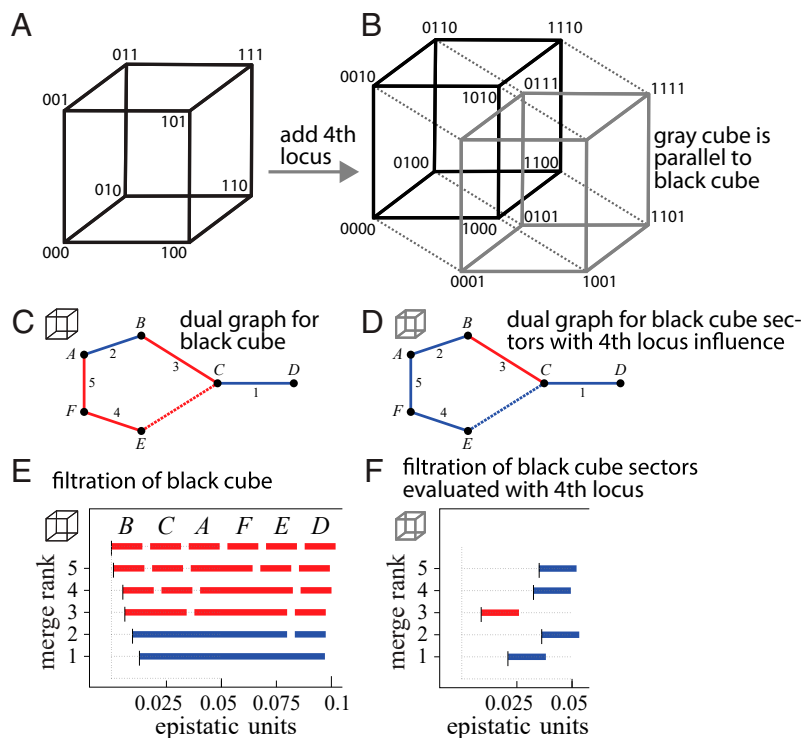


Fig. 5. Parallel epistatic filtration for three loci when a fourth locus is modified. (A) The 3D genotype space. (B) Adding a locus produces a 4D genotype space that can be visualized as two parallel 3D genotype spaces, depicted in black and gray, where the gray genotype space has a mutation in the fourth locus and the black is wild type at the fourth locus. (C) The dual graph of S for the black genotype space. (D) The parallel dual graph for the gray genotype space. Note several edges in C (black cube) shift to significant in D (gray cube), indicating the context of the fourth locus influences the interactions. (E) The epistatic filtration of the black genotype space. (F) The parallel filtration calculates epistasis of the black genotype sectors with the phenotypes of the parallel cube (i.e., when the fourth locus is present). This approach measures the influence of the fourth locus on the rest of the epistatic interactions in the network. Specifically, note the shift in the x-values of the black vertical tick marks on the left sides of the left-most colored bars in E versus the corresponding tick mark and bar in F.

the rank order. In this way, two parallel faces of the 5-cube are compared (Fig. 5 and *SI Appendix*, Fig. S1). Parallel filtrations extend the concepts of conditional, marginal, and sign epistasis (17, 50) into the geometric framework of epistatic filtrations.

Examining the *E. coli* data with and without the *pykF* mutation (42) (*SI Appendix*, Fig. S2) showed increased significance in 8 out of 22 of the dual edges, when *pykF* was mutated. Each bipyrmaid in *SI Appendix*, Fig. S2, *Left* matches a bipyrmaid in *SI Appendix*, Fig. S2, *Right* via the parallel transport operation (36). Examining the restoration of *pykF* to wild type (*SI Appendix*, Fig. S3), only 3 of 22 edges changed significance and just one critical edge lost significance, emphasizing that effects on the fitness landscape are not always bidirectional (43).

The biological interpretation of the parallel transport operation is simple. It measures the effect of genetic background on the fitness landscape. For Fig. 4E, this means that epistatic volumes with wild-type *rbs* are different when *rbs* is mutated. Since *rbs* is fixed in the parallel transport operation, we call this locus the bystander. Here, changing the bystander state from wild type to mutant modifies the magnitude and significance status of the epistatic volumes (Fig. 4C and E), with epistatic volumes generally higher when *rbs* is mutated. Thus, mutating the *rbs* locus distorts the fitness landscape with new epistatic interactions, which in turn opens up new evolutionary trajectories. We note that the precise locations of the distortions are concentrated as a set of adjacent blue edges in the dual graph (Fig. 4E, *Inset*). We call loci that change the fitness landscape in higher dimensions high-dimensional master regulators.

Lactobacilli and Acetobacters Are Master Regulators of the Microbiome. Up to this point, we have focused on genetic epistasis, but our framework is equally valid for interactions of environmental parameters, including ecological interactions between bacterial species in the gut microbiome. Like the genome, which is composed of many genes that interact to determine organismal fitness, the microbiome is also composed of

many smaller units, i.e., bacterial species, which carry their own sets of genes. Hosts are known to select and maintain a certain core set of microbes (51, 52); the interactions of these bacteria can affect host fitness (12); and it is debated to what extent these interactions are of higher order, c.f. ref. 27. See also ref. 32 for a broad overview of papers elaborating on possible meanings and instances of higher-order epistasis. While vertebrates have a gut taxonomic diversity of ~ 100 to 1,000 species, precluding study of all possible combinations, the laboratory fruit fly, *Drosophila melanogaster*, has naturally low diversity of ≈ 5 stably associated species (53).

To assess how bacterial interactions affect fly health, we made gnotobiotic flies inoculated with each combination of a set of $n = 5$ bacteria ($2^5 = 32$ combinations) that were isolated from a single wild-caught *D. melanogaster*, consisting of two members of the *Lactobacillus* genus (*L. plantarum* and *L. brevis*) and three members of the *Acetobacter* genus (Fig. 6A). We measured fly lifespan, which we previously identified as a reproducible phenotype that is changed by the microbiome (12). Overall a reduction of microbial diversity (number of species) led to an increase in fly lifespan as with a taxonomically similar set of bacteria we examined previously, which came from multiple hosts (12). The key finding of the previous work was that higher-order interactions between bacterial species change the fitness of the host fly.

The dual graph for the 5-loci genotype space revealed a single significant and critical epistatic interaction (Fig. 6B). Abundant, significant, non-critical edges were distributed throughout the graph (Fig. 6C) indicating prevalent interactions that weakly affect the fitness landscape. We note that such interactions were absent from the *E. coli* fitness landscape (compare the number of blue edges in Fig. 6B vs. Fig. 4D).

Using parallel filtrations to measure the role of individual bacterial species on the overall network, we found that the *Lactobacilli* drive changes in the global structure (Fig. 6D and F). In 46 out of 128 (36%) interactions, significance changed due to

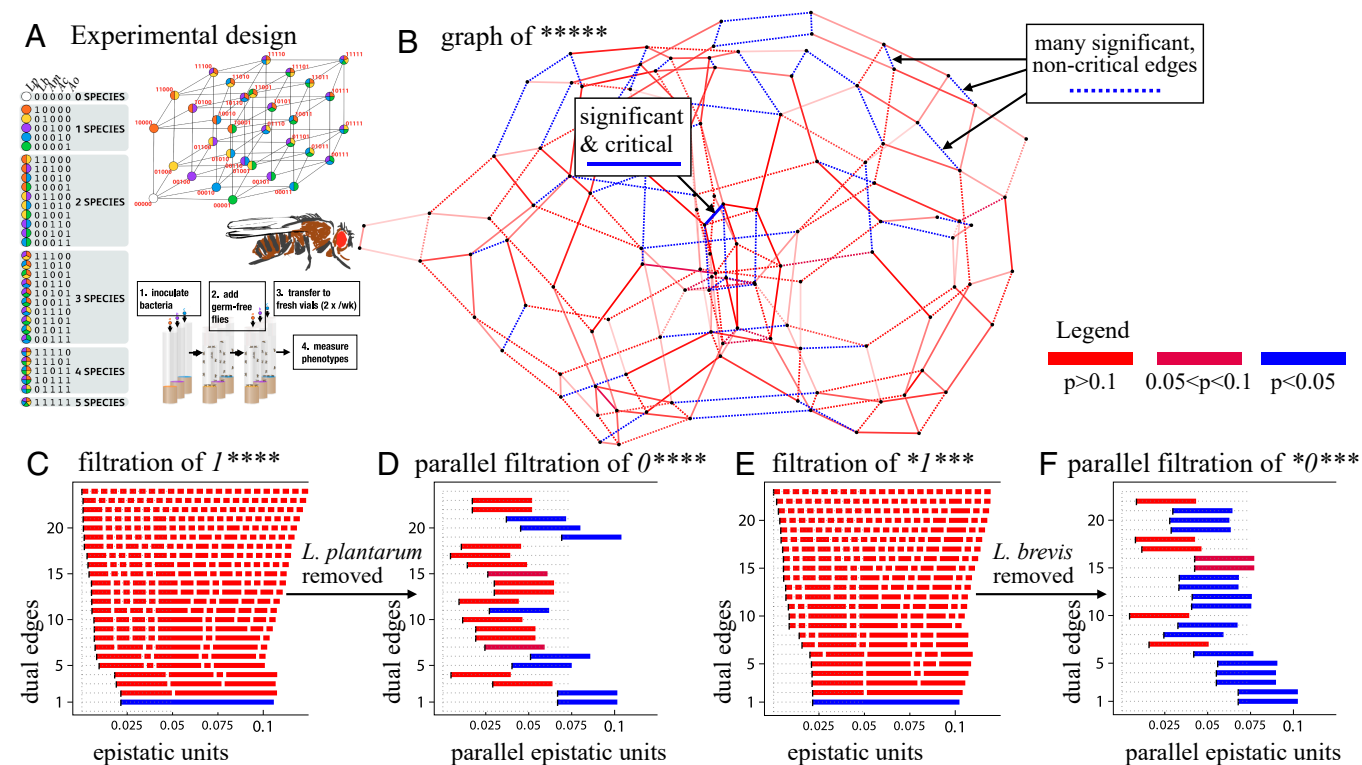


Fig. 6. Loss of lactobacilli causes global distortion of the microbiome epistatic landscape. (A) Experimental design for *DmelEble* and *DmelGould* (12) microbiome manipulations in flies. (B) Full graph of **** for the *DmelEble* data. (C) Filtration of $S(h)$ for the 4-face, 1****, of *DmelEble* data, where *L. plantarum* is present, indicates epistasis where two clusters of sectors merge. (D) Parallel filtration with *L. plantarum* removed shows a landscape distortion. (E) Filtration for *1***, where *L. brevis* is present has similar structure to 1****. (F) Parallel filtration with *L. brevis* removed shows a landscape distortion.

adding or removing a *Lactobacillus* (Fig. 6 C–F and *SI Appendix*, Figs. S7 and S8). These changes in significance primarily derive from non-significant interactions when *L. brevis* is present that become significant when it is removed and vice versa, indicating *L. brevis* is a master regulator that suppresses epistatic interactions that affect fly lifespan.

However, *L. brevis* is not unique as a master regulator. In the *DmelEble* dataset, each of the species changes interactions that affect lifespan when it is removed from the 5–species group (Table 1). A similar pattern exists in the *DmelGould* dataset, with *L. plantarum* showing the most prevalent effects and the *Acetobacter*s showing lower overall impact. Thus, both *Lactobacilli* and *Acetobacter*s can be master regulators of fly lifespan.

The gut microbiome has been suggested as a driver of human aging (54), and it is well documented to influence aging in *Drosophila* (55), including an increase in bacterial abundance as flies age (56). Microbiome interactions affecting abundances could drive the effects on host lifespan; however, comparing the epistatic landscapes for CFUs and lifespan, we found that only 2 of 99 dual edges were significant for both the bacterial abundance

and fly lifespan datasets (*SI Appendix*, Figs. S9, S10, S11, and S12 and Tables S2, S3, S4, and S5), and there was a lack of correlation between the epistatic volumes of the bipyramids (Spearman rank correlations: $P = 0.7$, $P = 0.5$, $P = 0.3$, and $P = 0.3$ respectively). This discord between the epistatic landscapes for microbiome fitness and host fitness could, e.g., diminish the rate of co-evolution between host and microbiome because the interactions influencing microbiome fitness (i.e., abundance) are different from the ones influencing host fitness through aging.

The Epistatic Landscape within a Single Enzyme Is Rugged. As a point of comparison with the *Ecoli_{evo}* dataset for genome evolution, we re-analyzed the *Ecoli_{lac}* dataset, a fully factorial combination of 5-mutations in a single gene, β -lactamase, which confers antibiotic resistance. Each mutation is in a separate residue of the same enzyme (43, 57). Weinreich et al. (57) found that there was a single fitness peak with a low number of evolutionary paths to reach it. Tan et al. (43) found limited potential for reverse evolution despite the relative simplicity of

Table 1. Role of each microbiome species as a regulator of interactions influencing fly lifespan

Species	<i>DmelEble</i> 1 $\rightarrow$ 0	<i>DmelGould</i> 1 $\rightarrow$ 0	<i>DmelEble</i> 0 $\rightarrow$ 1	<i>DmelGould</i> 0 $\rightarrow$ 1
<i>L. plantarum</i>	7	8	2	8
<i>L. brevis</i>	14	5	0	7
<i>Acetobacter</i> sp. 1	12	6	3	4
<i>Acetobacter</i> sp. 2	13	2	2	4
<i>A. orientalis</i>	7	0	4	6

Number of interactions out of ≈ 22 that become significant in the parallel filtration by removing ($1 \rightarrow 0$) or adding ($0 \rightarrow 1$) the bacterial species.

the fitness landscape. In our reanalysis, we note that the data are discrete (growth/no growth for a given set of antibiotic concentrations), and this type of microbiology experiment does not show variation in general. Thus, we can generally treat the calculated interaction magnitudes as accurate. Our computations are based on the reported mean values from Tan (43).

Examining the filtration, the epistasis arises in many steps with increasing magnitude in each step (*SI Appendix, Figs. S5 and S6*), consistent with the low number of possible evolutionary paths observed by Weinreich et al. (57), and distortions are apparent in the shifted magnitude of epistasis by parallel transport (*SI Appendix, Figs. S5 and S6*), consistent with the irreversible evolutionary paths observed by Tan et al. (43). The filtration also reveals a tiered structure to the epistasis, c.f. the largest weight merges two clusters of simplices (see dual edge #2 in *SI Appendix, Figs. S5 and S6*) in contrast to the *Ecoli_{evo}* dataset, where epistasis came from one individual simplex, indicating that epistasis arises due to interactions between sets of mutations in the β -lactamase, which could relate to the subdomains of this enzyme.

Normalization of Phenotypes Allows Comparison of Epistasis across Dimensions and across Datasets. To gain a perspective on the generality of higher-order interactions, it is desirable to compare epistatic landscapes, such as between the point mutations in β -lactamase, the genomic mutations in *E. coli*, and the microbiome communities that we examine here. Do these very different biological systems have similar higher-order interactions? And do the interactions vary across dimensions? Different phenotypes have different metrics, making comparisons difficult for current approaches to epistasis. Filtrations are well suited in this sense based on our unique normalization approach. The resulting normalization of epistatic volumes provides a standardized metric for epistasis across different dimensions and across different datasets.

Comparing the magnitudes of interactions between the different datasets (Fig. 6D), the epistatic volume (i.e. magnitude) for the microbiome data generated a $\approx 5\%$ effect, roughly three times the weight in the *Ecoli_{evo}* data and half that in the *Ecoli_{lac}* landscape (43) (c.f. *x*-axis between Fig. 6 and *SI Appendix, Figs. S4 and S5*), indicating the magnitude of epistasis varies by over two doublings across the different scales and different selection pressures of these biological systems. Epistatic filtrations provide a tool to explore the reasons for these differences.

Interactions Are Sparse in Higher Dimensions. Previous literature is somewhat equivocal on the extent to which higher-order interactions are important. We used epistatic filtrations to systematically evaluate the prevalence of higher-order interactions as a function of the number of dimensions. Critical, significant, higher-order interactions were less frequent than pairwise interactions ($P < 10^{-6}$, *Z-test*) for each of the *Ecoli_{evo}*, *Dmel_{Eble}*, and *Dmel_{Gould}* datasets, with a decreasing probability as a function of the face dimension (Table 2). This occurs for three primary reasons. First, the degrees of freedom increase fast in higher dimensions. Second, the probability of selecting a significant interaction from the set of all possible interactions decreases because the total number of interactions increases dramatically with increasing dimensions. Finally, the absolute number of significant interactions decreases in higher dimensions (Table 2), meaning they are biologically less prevalent (11). Overall, $\approx 10\%$ of possible dual edges were significant at higher order, with $\approx 1\%$ significant for $n = 5$ dimensions (Table 2), suggesting limits to the dimensions of biological complexity.

Table 2. Prevalence of interactions at different levels of complexity in genetics and microbiome datasets

Interaction dimension	Dataset: <i>Ecoli_{evo}</i>	Dataset: <i>Dmel_{Eble}</i>	Dataset: <i>Dmel_{Gould}</i>
2:	20/80 (25%)	24/80 (30%)	22/80 (28%)
≥ 3 :	29/508 (5.7%)	58/540 (10%)	21/520 (4.0%)
3:	21/194 (11%)	35/199 (17%)	14/194 (7.2%)
4:	7/214 (3.2%)	22/226 (10%)	6/216 (2.7%)
5:	1/100 (1.0%)	1/115 (0.8%)	1/110 (0.9%)
total:	49/588 (8.3%)	82/620 (13%)	43/600 (7.1%)

Significant critical dual edges ($P < 0.05$).

Higher-Order Interactions Can Arise from Lower-Order Interactions. Lower-order interactions can produce interactions in higher dimensions (48). In examining the higher-order epistasis present in our datasets, we noted that the clusters where significant epistatic volumes occur are often preceded in lower dimensions by clusters with nearly significant epistatic volumes (*SI Appendix, Fig. S4*). More explicitly, higher-dimensional bipyramids with significant epistatic volumes often have lower dimensional projections that also have significant epistatic volumes. Considering that epistasis is calculated as the volume of the bipyramid, it is intuitive that high-dimensional shapes with large volumes might have projections in lower dimensions that also have large volumes. However, that is not guaranteed.

We developed a graphical approach to distinguish these cases from those that arise de novo (*SI Appendix, Fig. S14 B and C and Meta-Epistatic Charts*). These charts are intended to answer the questions of i) to what extent are higher-order epistatic effects induced by lower dimensional ones and ii) which lower dimensional interactions maintain significance when embedded into higher dimensions? The reasoning is similar to regression-based epistasis calculations, where one can assign a certain portion of a higher-order interaction into the fitted coefficients of lower-order interactions.

In *SI Appendix, Fig. S14B*, we exhibit an example for the *Dmel_{Eble}* dataset, with 5 loci, where we consider the three 4-dimensional faces 0****, *0*** and **0**. For each such face, we computed the corresponding filtration of epistatic volumes. We then repeated this procedure for the relevant 3-dimensional projections (*SI Appendix, Fig. S14B* second row) and 2-dimensional projections (*SI Appendix, Fig. S14B* last row). We performed the same operations on the *Dmel_{Gould}* and *Ecoli_{evo}* data, where there are overall fewer significant epistatic volumes.

We observed examples of lower-order interactions inducing higher-order ones (*SI Appendix, Fig. S14C*). We also observed that several higher-order interactions in the *Dmel_{Eble}*, *Dmel_{Gould}* and *Ecoli_{evo}* data could not be attributed to lower-order effects (*SI Appendix, Fig. S14 B and C* as well as *SI Appendix, Table S6*) in four, three, or two interacting loci inside the 5-locus system, regardless of their significance (c.f. *SI Appendix, Fig. S14C*). Thus, some interactions arise only in the higher-dimensional context and cannot be discovered or predicted by studying lower-order interactions. Specifically, our analysis suggests that $\lesssim 10\%$ of higher-order interactions cannot be predicted by studying their components in lower dimensions (Table 2).

As we noted previously, the 4-dimensional interaction in the *E. coli* evolution experiment involved loci with two genes mutated (Fig. 4), whereas in the microbiome, interactions involved

loci with four species present, suggesting different underlying geometries for the interactions between genes in evolution versus between species in the microbiome.

The Analysis Tools Are Available through an Online Client. In order to enable other researchers access to these tools, we made an online client at https://www3.math.tu-berlin.de/combi/dmg/data/epistatic_filtrations/, where users can upload their datasets and receive the complete analysis. The site calculates the filtration and then iteratively computes the parallel filtrations for each locus separately. The outputs are the dual graph, the epistatic filtration, and a table of the results for each locus. From the summary table, users can quickly determine whether any loci are master regulators. In order for users to be able to test out the tools, the *Dmel^{Eble}*, *Dmel^{Gould}*, and *Ecoli^{evo}* datasets can also be downloaded from the site and analyzed. Users may also wish to test synthetic datasets.

Discussion and Conclusions

New Biological Findings. From an evolutionary perspective, the Red Queen hypothesis emphasizes how conflicts with other organisms can drive continuous genetic innovation (58). In our analysis of the shapes of fitness landscapes, we find that epistasis in higher dimensions reshapes the fitness landscape (59). Thus, the continuous diversification observed in long-term evolution experiments (47) could be generated by the continuously changing fitness landscape as new mutations occur. In particular, we identify master regulators that operate in higher dimensions by significantly enhancing or suppressing interactions in the rest of the biological network. In the microbiome, both *Lactobacilli* and *Acetobacters* can be master regulators. It is interesting to note that the *Dmel^{Eble}* dataset shows stronger effects of the individual species than the *Dmel^{Gould}* dataset. A possible contributing factor is that the *Dmel^{Eble}* bacteria were all isolated from the same individual wild fly, whereas the *Dmel^{Gould}* bacteria were isolated from multiple flies. Thus, the established ecological community may have stronger interactions. In *E. coli* evolution, we identified *rbs* and *pykF* as master regulators that open up new evolutionary trajectories. While it would require future experiments, it might be expected that higher-order master regulator genes may similarly regulate the onset and progression of evolutionary diseases such as cancers.

The prevalence and importance of higher-order interactions is debated, with some studies suggesting pairwise interactions predict the vast majority of interactions in complex communities (27) and others suggesting a large influence of context-dependent effects (12, 29), which would make higher-order interactions unpredictable. Ample evidence that higher-order epistasis has at least some evolutionary impact was established in recent publications; see ref. 32 and its references. Our analyses indicate that higher-order interactions exist at least as high as 5 dimensions but that the prevalence of epistasis in higher dimensions is quite low. Thus, we find that the unpredictability of higher-order interactions exists, but it is rare to the extent that lower-order interactions should be predictive a majority of the time.

Relation between Epistatic Filtrations and Other Measures of Epistasis. Simple regression-based approaches detect higher-order interactions in the same datasets that we analyze here (12, 42), but they are limited in that they are locus-focused, meaning that they do not detect interactions between genotypes, which are the basis for the fitness landscape. For example,

simple regression approaches would miss the interactions between sets of double mutants such as we detect for the *Ecoli^{evo}* data (Fig. 4).

From a methodological point of view, the present work lays the geometric groundwork for detecting epistasis via interactions of higher-order as well as other geometric properties of large fitness landscapes. Our work relies on polytope theory, following the shape approach of (37, 60), as this is a natural framework allowing a mathematical definition of epistasis in a fine-grained manner for a general n -locus system. By this we mean, that our interactions involve a minimal number of genotypes in the sense of a minimal set of dependent points, e.g., simplices (61). The motivation for this is that these sets generalize the notion of adjacent triangles in a 2-locus system to an n -locus system. Additionally, in this way, interactions have a geometric meaning, which makes them comparable across datasets. Although our method has similarities with (37, 60), it also has significant theoretical and computational differences and improvements. For example, our analyses heavily rely on studying the dual graph of the induced triangulation together with the filtration. This is a novelty in the theory and provides a number of new biological findings. For example, we localize regions of epistasis in four fitness landscapes, we quantify the sparsity of these regions, and we compare portions of fitness landscapes via the parallel transport operation, where we examine the effects of different bystanders on the rest of the interactions. We also further develop (36) by providing a new framework to detect and interpret how higher-order epistasis arises from lower-order epistasis via meta-epistatic charts.

More specifically, epistatic volumes capture new properties of fitness landscapes even in the 3-locus case. In this case, there are between four and six epistatic volumes, as these are the number of adjacent pairs of simplices in the subdivision of the 3D cube, which appear as edges in the dual graph (62, figure 3). In contrast, there are 20 circuit interactions (37, Ex. 3.9) and many more possible and potentially relevant interactions that must be checked in a randomized, exhaustive search. In addition to reducing the search space, epistatic volumes can be localized in the fitness landscape, allowing the occurrence of mutations to be linked to changes in the topography of the epistatic landscape. Furthermore, we can link these changes across dimensions, tracking the source of the interactions.

Our method relates to other measures of epistasis, for example to linear regression approaches, as we explain in *SI Appendix: Comparison with a simple linear regression approach*; see also the recent work (8). It also relates to methods originating from harmonic analysis, c.f. refs. 48, 63, and 11; and to correlations between the effects of pairwise mutations, as we pointed out in ref. 36. More concretely, in a 2-locus system, all these methods can easily be recovered from one another; some of them even agree. This is also true for some ecological approaches, including the generalized Lotka–Volterra equations, which yield a mathematically equivalent form of epistasis for certain situations; c.f. see equation 9 of ref. 14. In higher dimensional systems, these methods remain conceptually closely related but they generally yield different insights about the problem, such as which interactions are considered, whether the interactions are significant, what their magnitude is, and what their sign is. Because these previous methods make specific, a priori assumptions about the forms of interactions, they are limited by these assumptions. Epistatic filtrations add a global perspective, determining the structure of interactions from the shape of the fitness landscape in a parameter-free approach.

Finally, rank orders play an important role in the recent fitness landscape theory (38, 64). For an overview and for references to relevant work in the theory, see the review article (32). It is straightforward to recast the fitness landscapes presented here into a rank-order fitness graph and then count the number of peaks, i.e., the number of sinks in a fitness graph. The technical details are beyond the scope of the present paper.

Interactions in Higher Dimensions. We found that biologically significant epistatic interactions in four and five dimensions are sparse and sometimes rooted in lower order, meaning that a limited number of regions of epistasis and hence of distortion exist in these fitness landscapes. This extends to higher dimensions the trend that 3-way interactions are often predicted from 2-way interactions (7, 12, 27). However, our finding that key genes and species cause distortions emphasizes the need to identify the significant higher-order interactions from the vast number of possible ones, a task that epistatic filtrations enable.

In a five-loci case, we also found that the fitness landscape in the *Dmel^{Eble}* dataset is much more distorted, i.e., non-linear, than the *Ecoli^{evo}* or *Dmel^{Gould}* fitness landscapes. We also found the precise locations of distortions inside the corresponding fitness landscapes and contextualize them in terms of distortions visible in their lower dimensional sub-landscapes through meta-epistatic charts. This graphical approach to isolate the source of higher-order interactions is new and cannot be established with previous methods.

Strength and Limitations of Epistatic Filtrations. A major advance of this work is that we provide a way to discover high-dimensional regulators of biological networks. Rather than identifying key nodes as having a high number of low-dimensional edges, we developed a method to identify nodes that regulate the higher-dimensional interactions in the rest of the network. This operation is performed by the parallel transport function, and we provide a web-based tool to perform the analysis. The implications of these findings are that certain genes and species modulate the interactions in the rest of the network, and perturbing these loci can destabilize the network. Destabilizing an unhealthy biological network could be crucial to restoring a degraded ecosystem, a sick microbiome, or curing a cancer, while destabilization of a healthy biological network could have the opposite consequences.

Methodologically, we also improve the framework in which higher-order epistasis can be mathematically formalized and analyzed geometrically. We provide concrete tools to find epistatic interactions in the fitness landscape and to distinguish whether the landscape is locally flat, i.e., close to a hyperplane of a certain dimension. Our work additionally allows us to localize and contextualize regions inside the fitness landscape which are not flat and hence distorted.

Our approach does not require a distinction between positive and negative epistasis, but only between presence and absence of epistasis. We note that the sign in traditional versions of epistasis depends on the choice of the origin, i.e., wild type for the interaction space, which does not have a clear biological motivation, particularly if different races are compared in the dataset. The 2-locus case provides an elementary example. In traditional terms, the epistasis in Fig. 3 is negative since the fitness of 11 lies below the plane spanned by the fitnesses of 00, 10, and 01. Picking that particular plane for choosing the sign rests on picking 00 as wild type. If instead we use the genotype 10 as the origin, then the fitness of that genotype and those of its

two neighbors, 00 and 11, span a plane such that the fitness of 10 lies above that plane of reference. This is important because the same data can yield positive or negative epistasis depending on the choice of origin. In our approach, the location of epistasis is concisely encoded in the regular triangulation induced by the phenotypes as described (c.f. Fig. 3). In this sense, the lack of sign is not a limitation of epistatic filtrations but a consequence of the high-dimensional approach.

The bipyramids that occur in an epistatic filtration are special cases of the circuits studied by Beerenwinkel et al. (37). In this way, our method only considers some of the epistatic effects visible via more general circuits c.f. *SI Appendix, Table S1*. There are two reasons for this restriction. First, the simplices that are at the foundations of epistatic filtrations are characterized via fittest populations because they are defined by the ridges, which are the highest points of the convex hull (see ref. 36, Section 2.2), meaning that interactions of low fitness are ignored because we do not examine the valleys. It is a simple operation to shift focus to the valleys (36). Second, with an increasing number of genes, the total number of circuits to consider from ref. 37 grows dramatically. These are too many for a statistical study, even for as few as $n = 5$ genes; see *SI Appendix, Table S1*.

Finally, we note that the phenotypes that we consider in this manuscript are roughly linear. When non-linear phenotypes are considered, there is a risk of detecting spurious epistasis (48). Appropriate transforms can be applied, or a rank-based approach could be developed for epistatic filtrations. These issues are beyond the scope of the present work.

Outlook. This geometric approach could be extended, e.g., to GWAS (5, 10, 65), ecosystems (14, 15), or neuronal networks (66), to discover non-additive higher-order structures at different scales. Resolving these structures should become increasingly possible as higher throughput methodologies enable generation of larger datasets, approaches where each experimental factor is combinatorially included with respect to each other factor, as with the *E. coli* genes and microbiome species datasets here. To facilitate researchers analyzing their own data, we developed a web-based interface tool that allows researchers to upload their data and receive the results. The datasets analyzed in this work are provided and can be directly uploaded so that prospective users can explore the method.

It should be noted that the polyhedral geometry methods for analyzing epistasis deserve to be developed further from the mathematical point of view. We believe that more concepts related to curvature for piece-wise linear manifolds will be useful (67).

Taken together, our approach offers a number of insights into higher-dimensional properties of fitness landscapes and their biological implications, and we think these will be useful as higher throughput experiments enable more combinatorial approaches. **SI datasets.** Three datasets are supplied in *SI Appendix* that can be directly analyzed using the online client: *KhanFitnessNormalized.csv* (68), *EbleSurvivalDataNormalized.csv*, and *GouldSurvivalDataNormalized.csv* (12).

Materials and Methods

Fly Husbandry. Flies were reared germ-free and inoculated with one combination of bacteria on day 5 after eclosion. $N \geq 100$ flies were assayed for lifespan in $n \geq 5$ independent vials per bacterial combination for a total of 3,200 individual flies. Food was 10% autoclaved fresh yeast, 5% filter-sterilized glucose, 1.2% agar, and 0.42% propionic acid, pH 4.5. Complete methods are described in Gould et al. (12).

Bacterial Cultures. Bacteria were cultured on MRS or MYPL, washed in PBS, standardized to a density of 10^7 CFU/mL and 50 μ L was inoculated onto the fly food. Strains are indicated in *SI Appendix, Table S7*. See Gould et al. (12) for complete methods.

Genetics Data. Existing genetics datasets were gotten from Sailer and Harms 2017 (48) GitHub repository (<https://github.com/harmslab/epistasis>) or from Tan et al. (43). For the *Ecoli_{plac}* dataset, we focused on the piperacillin with clavulanate data from ref. 43 as it is the better behaved. More details are provided in *SI Appendix*.

Microbiome Datasets. In this work, *Drosophila* microbiome fitness landscapes consist of experimental measurements on germ-free *Drosophila* flies inoculated with different bacterial species. The lifespans of approximately 100 individual flies were measured for each combination of bacterial species, giving roughly 3,200 individual fly lifespans for each of the two datasets presented. The experimental methods are described in ref. 12. The first dataset is the exact data presented in ref. 12. The second dataset is the second set of species with exactly the same methods used in ref. 12. Each bacterial composition consists of all possible combinations of five species selected from a set of seven that occur naturally in the gut of wild flies: *Lactobacillus plantarum* (LP), *Lactobacillus brevis* (LB), *Acetobacter pasteurianus* (AP), *Acetobacter tropicalis* (AT), *Acetobacter orientalis* (AO), *Acetobacter cerevisiae* (AC), *Acetobacter malorum* (AM). The 5-member communities both stably persist in the fly gut. For the purposes of this work, we define stable as maintaining colonization of the gut when ≤ 20 flies are co-housed in a standard fly vial and transferred daily to fresh, sterile food. By this method, we observe that the total number of species found stably associated with an individual fly is typically between 3 and 8. Consistently, *Lactobacillus plantarum* and *Lactobacillus brevis* are found with two to three *Acetobacter* species. Less consistently, species of *Enterobacteria* and *Enterococci* occur and these have been described as pathogens. While more strains may be present, for each of the two datasets in the present work, a set of five non-pathogen species

was chosen, including the two *Lactobacilli* and three *Acetobacter* species. The combinations of species are shown in *SI Appendix, Table S7*. Different strains of the same species were used in the two datasets.

Computational Analysis. The filtrations code is available as a polymake (69) package (c.f. <https://github.com/holgereble/EpistaticFiltration>), and the analysis pipeline is available as a Jupyter notebook. We also provide an online client, which processes raw csv data sheets, https://www3.math.tu-berlin.de/combi/dmg/data/epistatic_filtrations/ (70).

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. Portions of this work appear in the dissertation of H.E., which was supervised by M.J. We are indebted to Bernd Sturmfels and Niko Beerenwinkel for introducing us and inspiring this work. We thank L. J. Holt, O. Brandman, J. Derrick, and two anonymous reviewers for their valuable comments. M.J. has been supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): The Berlin Mathematics Research Center MATH⁺ (EXC-2046/1, project ID 390685689); "Symbolic Tools in Mathematics and their Application" (TRR 195, project ID 286237555). W.B.L. acknowledges NIH grants DP5OD017851 and R01DK128454, NSF IOS grants 2032985 and 2144342, SCIALOG grant 27912 from Research Corporation for Science Advancement (RCSA) and the Allen Foundation, and the Carnegie Institution for Science Endowment.

Author affiliations: ^aChair of Discrete Mathematics/Geometry, Technical University Berlin, Berlin 10623, Germany; ^bMax Planck Institute for Mathematics in the Sciences, Leipzig 04103, Germany; ^cDepartment of Biosystems Science and Engineering, Federal Institute of Technology (ETH Zürich), Basel 4058, Switzerland; ^dSwiss Institute of Bioinformatics, Basel 4058, Switzerland; ^eDepartment of Biosphere Sciences and Engineering, Carnegie Institution for Science, Baltimore, MD 21218; and ^fDepartment of Biology, Johns Hopkins University, Baltimore, MD 21218

1. S. Wullschlegler, R. Loewith, M. N. Hall, TOR signaling in growth and metabolism. *Cell* **124**, 471–484 (2006).
2. R. T. Paine, A note on trophic complexity and community stability. *Am. Nat.* **103**, 91–93 (1969).
3. M. Costanzo et al., The genetic landscape of a cell. *Science* **327**, 425–431 (2010).
4. S. R. Collins et al., Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
5. X. Liu, Y. I. Li, J. K. Pritchard, Trans effects on gene expression can drive omnigenic inheritance. *Cell* **177**, 1022–1034.e6 (2019).
6. E. A. Boyle, Y. I. Li, J. K. Pritchard, An expanded view of complex traits: From polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
7. E. Kuzmin et al., Systematic analysis of complex genetic interactions. *Science* **360**, 1–9 (2018).
8. J. Zhou, D. M. McCandlish, Minimum epistasis interpolation for sequence-function relationships. *Nat. Commun.* **11**, 1782 (2020).
9. N. C. Wu, L. Dai, C. A. Olson, J. O. Lloyd Smith, R. Sun, Adaptation in protein fitness landscapes is facilitated by indirect paths. *eLife* **5**, 1–21 (2016).
10. Ö. Carlborg, C. S. Haley, Epistasis: Too often neglected in complex trait studies? *Nat. Rev. Genet.* **5**, 618–625 (2004).
11. D. M. Weinreich, Y. Lan, J. Jaffe, R. B. Heckendorn, The influence of higher-order epistasis on biological fitness landscape topography. *J. Stat. Phys.* **172**, 208–225 (2018).
12. A. L. Gould et al., Microbiome interactions shape host fitness. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11951–E11960 (2018).
13. A. Sanchez-Gorostiaga, D. Bajić, M. L. Osborne, J. F. Poyatos, A. Sanchez, High-order interactions distort the functional landscape of microbial consortia. *PLoS Biol.* **17**, e3000550 (2019).
14. T. J. Case, E. A. Bender, Testing for higher order interactions. *Am. Nat.* **118**, 920–929 (1981).
15. I. Billick, T. Case, Higher order interactions in ecological communities: What are they and how can they be detected? *Ecology* **75**, 1529–1543 (1994).
16. J. Grilli, G. Barabás, M. J. Michalska-Smith, S. Allesina, Higher-order interactions stabilize dynamics in competitive network models. *Nature* **548**, 210–213 (2017).
17. D. M. Weinreich, R. A. Watson, L. Chao, Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2005).
18. J. M. Smith, Natural selection and the concept of a protein space. *Nature* **225**, 563–564 (1970).
19. S. Kauffman, S. Levin, Towards a general theory of adaptive walks on rugged landscapes. *J. Theor. Biol.* **128**, 11–45 (1987).
20. D. M. McCandlish, Long-term evolution on complex fitness landscapes when mutation is weak. *Heredity* **121**, 449–465 (2018).
21. M. Schumer et al., Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656–660 (2018).
22. D. M. Weinreich, Y. Lan, C. S. Wylie, R. B. Heckendorn, Should evolutionary geneticists worry about higher-order epistasis? *Curr. Opin. Genet. Dev.* **23**, 700–707 (2013). Genetics of system biology.
23. K. Crona, Rank orders and signed interactions in evolutionary biology. *eLife* **9**, 1–12 (2020).
24. Z. R. Sailer, M. J. Harms, High-order epistasis shapes evolutionary trajectories. *PLoS Comput. Biol.* **13**, 1–16 (2017).
25. M. J. McDonald, D. P. Rice, M. M. Desai, Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature* **531**, 233–236 (2016).
26. C. Ratze, J. Barrere, J. Gore, Strength of species interactions determines biodiversity and stability in microbial communities. *Nat. Ecol. Evol.* **4**, 376–383 (2020).
27. J. Friedman, L. M. Higgins, J. Gore, Community structure follows simple assembly rules in microbial microcosms. *Nat. Publ. Group* **1**, 1–7 (2017).
28. P. Piccardi, B. Vessman, S. Mitri, Toxicity drives facilitation between 4 bacterial species. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15979–15984 (2019).
29. D. Sundaraman et al., Higher-order interactions dampen pairwise competition in the zebrafish gut microbiome. *mBio* **11**, 1–15 (2020).
30. M. Padi, J. Quackenbush, Integrating transcriptional and protein interaction networks to prioritize condition-specific master regulators. *BMC Syst. Biol.* **9**, 1–17 (2015).
31. D. Weinreich, R. Watson, L. Chao, Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evolution* **59**, 1165–1174 (2007).
32. J. Krug, Epistasis and evolution (2021).
33. I. Fragata, A. Blancaert, M. A. Dias Louro, D. A. Liberles, C. Bank, Evolution in the light of fitness landscape theory. *Trends Ecol. Evol.* **34**, 69–82 (2019).
34. F. J. Poelwijk, D. J. Kiviet, D. M. Weinreich, S. J. Tans, Empirical fitness landscapes reveal accessible evolutionary paths. *Nature* **445**, 383–386 (2007).
35. S. Wright, The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proc. Sixth Int. Congr. Genet.* **1**, 356–366 (1932).
36. H. Eble, M. Joswig, L. Lamberti, W. B. Ludington, Cluster partitions and fitness landscapes of the *Drosophila* fly microbiome. *J. Math. Biol.* **79**, 1–39 (2019).
37. N. Beerenwinkel, L. Pachter, B. Sturmfels, Epistasis and shapes of fitness landscapes. *Statist. Sinica* **17**, 1317–1342 (2007).
38. K. Crona, A. Gavryushkin, D. Greene, N. Beerenwinkel, Inferring genetic interactions from comparative fitness data. *eLife* **6**, 1–28 (2017).
39. K. Venkatesan et al., An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).
40. M. Schuldiner et al., Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
41. R. T. Paine, Food-web analysis through field measurement of per capita interaction strength. *Nature* **355**, 73–75 (1992).
42. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011).
43. L. Tan, S. Serene, H. X. Chao, J. Gore, Hidden randomness between fitness landscapes limits reverse evolution. *Phys. Rev. Lett.* **106**, 1–4 (2011).
44. R. Forman, Morse theory for cell complexes. *Adv. Math.* **134**, 90–145 (1998).

45. J. E. Barrick, R. E. Lenski, Genome dynamics during experimental evolution. *Nat. Rev. Genet.* **14**, 827–839 (2013).
46. H. H. Chou, H. C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).
47. B. H. Good, M. J. McDonald, J. E. Barrick, R. E. Lenski, M. M. Desai, The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45–50 (2017).
48. Z. R. Sailer, M. J. Harms, Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics* **205**, 1079–1088 (2017).
49. S. Herrmann, M. Joswig, Splitting polytopes. *Münster J. Math.* **1**, 109–141 (2008).
50. J. L. Gill, Effects of finite size on selection advance in simulated genetic. *Austr. J. Biol. Sci.* **18**, 599–617 (1965).
51. R. E. Ley *et al.*, Evolution of mammals and their gut microbes. *Science* **320**, 1647–1651 (2008).
52. A. Risely, Applying the core microbiome to understand host-microbe systems. *J. Anim. Ecol.* **89**, 1549–1558 (2020).
53. W. B. Ludington, W. W. Ja, *Drosophila* as a model for the gut microbiome. *PLoS Pathog.* **16**, 1–6 (2020).
54. M. J. Blaser, G. F. Webb, Host demise as a beneficial function of indigenous microbiota in human hosts. *mBio* **5**, e02262–14 (2014).
55. A. Arias-Rojas, I. Iatsenko, The role of microbiota in *Drosophila melanogaster* aging. *Front. Aging* **3**, 1–10 (2022).
56. N. A. Broderick, N. Buchon, B. Lemaître, Microbiota-induced changes in *Drosophila melanogaster* host gene expression and gut morphology. *mBio* **5**, e01117–14 (2014).
57. D. M. Weinreich, N. F. Delaney, M. A. Depristo, D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114 (2006).
58. L. Van Valen, Molecular evolution as predicted by natural selection. *J. Mole. Evol.* **3**, 89–101 (1974).
59. D. Bajic, J. C. C. Vila, Z. D. Blount, A. Sanchez, On the deformability of an empirical fitness landscape by microbial evolution. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11286–11291 (2018).
60. N. Beerenwinkel, L. Pachter, B. Sturmfels, S. F. Elena, R. E. Lenski, Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evol. Biol.* **7**, 60 (2007).
61. J. A. De Loera, J. Rambau, F. Santos, *Triangulations, Algorithms and Computation in Mathematics* (Springer-Verlag, Berlin, 2010), vol. 25, p. xiv+535. Structures for algorithms and applications.
62. P. Huggins, B. Sturmfels, J. Yu, D. S. Yuster, The hyperdeterminant and triangulations of the 4-cube. *Math. Comput.* **77**, 1653–1679 (2008).
63. E. D. Weinberger, Fourier and Taylor series on fitness landscapes. *Biol. Cybernet.* **65**, 321–330 (1991).
64. C. Lienkaemper, L. Lamberti, J. Drain, N. Beerenwinkel, A. Gavryushkin, The geometry of partial fitness orders and an efficient method for detecting genetic interactions. *J. Math. Biol.* **77**, 951–970 (2018).
65. G. Fang *et al.*, Discovering genetic interactions bridging pathways in genome-wide association studies. *Nat. Commun.* **10**, 1–18 (2019).
66. M. W. Reimann *et al.*, Cliques of neurons bound into cavities provide a missing link between structure and function. *Front. Comput. Neurosci.* **11**, 1–68 (2017).
67. J. M. Sullivan, "Curvatures of smooth and discrete surfaces" in *Discrete Differential Geometry, Oberwolfach Semin* (Birkhäuser, Basel, 2008), vol. 38, pp. 175–188.
68. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011).
69. E. Gawrilow, M. Joswig, "polymake: A framework for analyzing convex polytopes" in *Polytopes–Combinatorics and Computation (Oberwolfach, 1997)*, *DMV Sem.* (Birkhäuser, Basel, 2000), vol. 29, pp. 43–73.
70. H. Eble, "Epistasis, regular subdivisions and spanning trees" PhD thesis, Technischen Universität Berlin (2022).