High dimensional geometry of fitness landscapes identifies master regulators of evolution and the microbiome

Holger Eble,[1] Michael Joswig,[1,2]* Lisa Lamberti[3,4],William B. Ludington[5,6]*

[1]Chair of Discrete Mathematics/Geometry, TU Berlin, Germany

[2]MPI MiS Leipzig, Germany

[3]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

[4] SIB Swiss Institute of Bioinformatics, Basel, Switzerland

[5] Department of Embryology, Carnegie Institution for Science, USA

[6] Department of Biology, Johns Hopkins University, Baltimore, MD, USA

*To whom correspondence should be addressed;
E-mail: joswig@math.tu-berlin.de, ludington@carnegiescience.edu.

A longstanding goal of biology is to identify the key genes and species that critically impact evolution, ecology, and health. Yet biological interactions between genes (*1, 2*), species (*3–6*), and different environmental contexts (*7–9*) change the individual effects due to non-additive interactions, known as epistasis. In the fitness landscape concept, each gene/organism/environment is modeled as a separate biological dimension (*10*), yielding a high dimensional landscape, with epistasis adding local peaks and valleys to the landscape. Massive efforts have defined dense epistasis networks on a genome-wide scale (*2*), but these

1

have mostly been limited to pairwise, or two-dimensional, interactions (*11*). Here we develop a new mathematical formalism that allows us to quantify interactions at high dimensionality in genetics and the microbiome. We then generate and also reanalyze combinatorically complete datasets (two genetic, two microbiome). In higher dimensions, we find that key genes (e.g. *pykF*) and species (e.g. *Lactobacillus plantarum*) distort the fitness landscape, changing the interactions for many other genes/species. These distortions can fracture a "smooth" landscape with one optimal fitness peak into a landscape with many local optima, regulating evolutionary or ecological diversification (*12*), which may explain how a probiotic bacterium can stabilize the gut microbiome.

1 Introduction

A fitness landscape depicts biological fitness as a function of its many underlying parts, namely genes, each as a separate dimension (*10, 13, 14*). Interactions between genes can change their individual impacts on fitness in a non-additive way (*15*), adding local peaks and valleys to the fitness landscape, which affects the evolutionary paths through the landscape (*16*). The mathematical frameworks to quantify biological interactions, namely epistasis, determine the degree of non-additivity, and the concept has been applied to genetics (*1, 2*), microbiomes (*3*), and ecology (*4–6*). In two dimensions, epistasis calculates interactions between e.g. two genes as the degree to which a double mutant phenotype can be predicted by measuring the two single mutants independently. Applying epistasis to genome-wide measurement of pairwise (*17, 18*) and three-way (*2*) genetic interactions has revealed biochemical pathways composed of discrete sets of genes as well as complex traits, such as human height, that are affected by almost every gene in the genome (*19, 20*). New techniques allow epistasis to be applied to broader data types (*21*).

2

38   Epistatic interactions can arise due to mutations (*13, 14, 22*) or when sex, recombination, and

39 horizontal gene transfer bring groups of genes together (*1, 23–26*), making multiple dimensions

40 interact simultaneously. Interactions between bacteria in the microbiome also have functional

41 consequences (*3, 27–31*) and are prevalent in higher-dimensions (*3, 31*), where community as-

42 sembly may introduce groups of species in different combinations e.g. in a fecal transplant.

43   Interactions in higher dimensions could change the topography of the fitness landscape (*31*),

44 and their relative importance is unknown. To various extents, current approaches are limited in

45 their ability to discern the topography of interaction landscapes in high dimensions due to (*i*)

46 sign epistasis, which does not generalize well to more than two dimensions, (*ii*) a narrow ability

47 to account for genomic context, and (*iii*) statistical considerations of the false discovery rate due

48 to multiple testing (*32, 33*). Several different concepts of epistasis exist in the literature (*34*).

49 However, standard epistasis frameworks often rely on parameter fitting, which brings along

50 additional constraints (*35*). "Circuits", which can describe all possible epistatic interactions

51 (*33, 36*), introduce false discovery rate challenges.

52   Here we develop a new formulation of the *fitness cube* concept (*10, 13, 14*), where each bio-

53 logical entity (gene, organism, environmental factor) is a separate dimension (Fig. 1a). Because

54 the biological entities are discrete (i.e., either a bacterium is there or it is not), our framework is

55 discrete too. The cubes represent the landscape for interactions and can be composed of many

56 dimensions as n-dimensional hypercubes. We then develop *epistatic filtrations* to locate the

57 epistasis on this fitness landscape. Our approach solves problems of context and sign while

58 reducing multiple testing concerns, all in a parameter-free form that is consistent across many
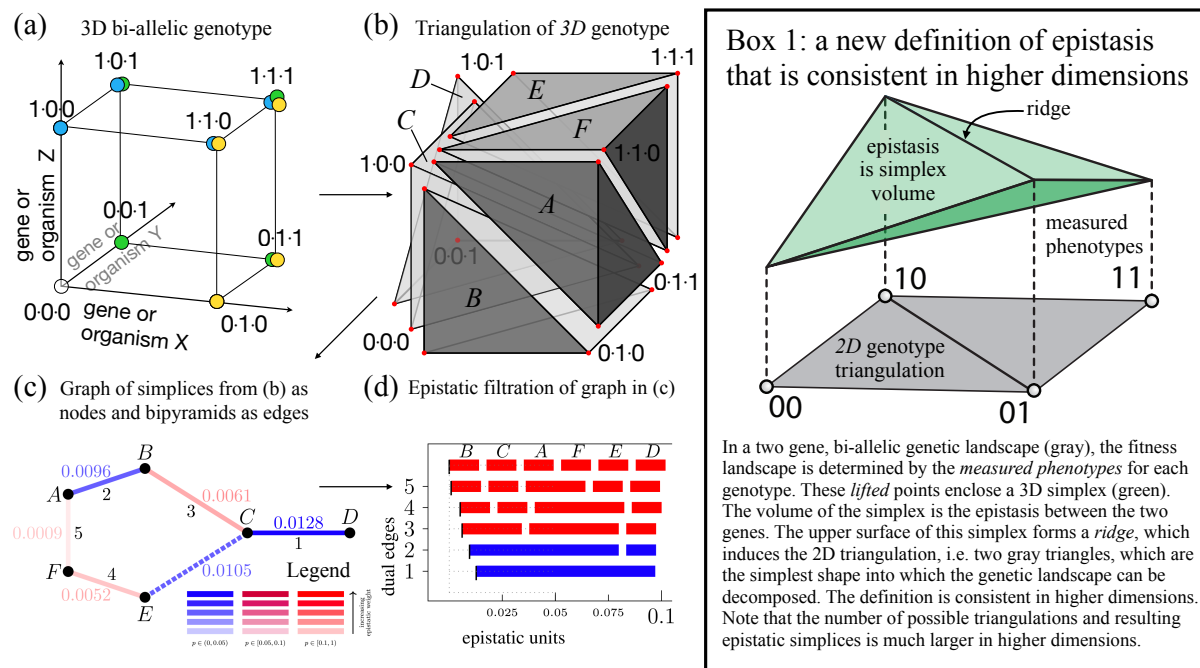
59 dimensions (Fig. 1, Box 1).

3

(a) 3D bi-allelic genotype

(b) Triangulation of *3D* genotype

(c) Graph of simplices from (b) as nodes and bipyramids as edges

(d) Epistatic filtration of graph in (c)

Box 1: a new definition of epistasis that is consistent in higher dimensions

ridge

epistasis is simplex volume

measured phenotypes

10   11

2D genotype triangulation

00   01

In a two gene, bi-allelic genetic landscape (gray), the fitness landscape is determined by the *measured phenotypes* for each genotype. These *lifted* points enclose a 3D simplex (green). The volume of the simplex is the epistasis between the two genes. The upper surface of this simplex forms a *ridge*, which induces the 2D triangulation, i.e. two gray triangles, which are the simplest shape into which the genetic landscape can be decomposed. The definition is consistent in higher dimensions. Note that the number of possible triangulations and resulting epistatic simplices is much larger in higher dimensions.

Figure 1: **Filtrations describe epistatic topography**. (a) Interacting biological entities, e.g. genes in a cell or bacterial strains in a microbiome, can be depicted as orthogonal dimensions in a unit cube, where vertices represent different genotypes or combinations of strains. (b) With 3 dimensions, the triangulation (*14*) of the fitness landscape produces *3D* simplices (labeled *A-F*) of the genotypes, and *4D* simplices of the fitness landscape (not shown) give the epistasis. (c) To map the global connectivity of the landscape, we merge adjacent simplices in a dual graph of the 3-cube triangulation, where nodes *A-F* are the simplices from (d) and the edges are the volumes of the bipyramids from the merges of neighboring simplices. The smallest bipyramid, edge 5, is formed first, followed by the next larger and so forth on up to the largest bipyramid, edge 1. The data set is from *Esherichia coli* mutations in *topA*, *spoT*, and *pykF* from (*37*). (c legend) Each dual edge has two parameters: its epistatic weight (indicated by shade) and its *p*-value (indicated by color). Black indices in (c) label the *critical* dual edges of $\mathcal{S}(h)$, where critical indicates that loss of the edge leaves nodes unconnected to the graph. (d) The sequence of merges between adjacent simplices (reading from top to bottom) shown in the dual graph is depicted by the epistatic filtration. Epistasis of the merged simplex is indicated by the thin, black vertical hatch mark on the far left bar of each row. Total width of the bars is fixed. Note the non-critical C+E merge is not depicted in the filtration because those simplices are already merged with B, A, and F.

60 2 Results

61 2.1 Defining the shapes of fitness landscapes

We first decompose the fitness cube into its most elementary parts through a triangulation (Box 1, Fig. 1b). Triangulations are used e.g. in computer vision to decompose a surface, such as a human face, into discrete parts, which are triangles. Generalizing to higher dimensions, the triangles connecting genotypes are simplices (Fig. 1b). The volume of each simplex connotes the local steepness of the landscape (Box 1). To establish the global topography of the landscape we merge adjacent simplices in a stepwise manner such that flattest parts of the landscape are merged first and the steepest parts last (Fig. 1c). Each adjacent pair of merged simplices, s and t, forms a bipyramid, (s,t) through their shared face. The **epistatic weight** of (s,t) is

$$e_h(s,t) \; := \; \left| \det E_h(s,t) \right| \cdot \frac{\mathrm{nvol}(s \cap t)}{\mathrm{nvol}\,(s) \cdot \mathrm{nvol}\,(t)} \; . \tag{1}$$

where $E_h(s,t)$ is the matrix specifying the vertices with their corresponding fitness phenotypes and nvol denotes the dimensionally normalized volume of the genotypes (Box 1; Appendix B1-B6). We use the notation

$$\{v^{(1)}\} + \{v^{(2)}, \ldots, v^{(n+1)}\} + \{v^{(n+2)}\} \tag{2}$$

62 for the bipyramid (s,t), where the first and last vertices are the apices and the middle set forms 63 the shared face. The $n+2$ genotypes of the bipyramid form a non-linear interaction of dimension 64 n when $e_h(s,t) > 0$.

65   We visualize the topography of the **epistatic landscape** by forming a **dual graph** of $\mathcal{S}(h)$, 66 where the nodes are the maximal simplices and adjacent simplices form the dual edges. Blue 67 edges indicate epistasis (Fig. 1c). The **epistatic filtration** of h (Fig. 1d) depicts the path from 68 lowest to highest epistasis by merging adjacent simplices to form a connected **cluster** c.f. (*38*). 69 In this sense, epistatic filtrations encode a global notion of epistasis in higher dimensions by

5

70 connecting adjacent bipyramids. This method has many advantages over parameter fitting,

71 including that it does not depend on the statistical constraints of determining a best fit.

72 Filtrations are also not constrained by the sign of epistasis, which depends on which geno-

73 type is considered *wildtype*, a somewhat arbitrary decision given varied ancestries (see Ap-

74 pendix B1). Studying adjacent simplices and their neighboring relationships, as we propose

75 below, allows reconstruction of the fitness landscape and its epistatic properties in high dimen-

76 sions. This process rests on the mathematical theory of linear optimization, convex polyhedra,

77 and regular subdivisions (*38*).

78 We note that bipyramids account for the majority of genomic contexts (*38*), c.f. Table S1.

79 Furthermore, the location(s) of inferred epistasis is robust to the choice of triangulation $\mathcal{S}(h)$

80 (*38*).

2.2   An evolutionary genetics example of epistatic filtrations

82 To illustrate our approach, we examined an existing data set from Lenski's (*39*) classic experi-

83 mental evolution of *Esherichia coli*, in a set of strains with each combination of five beneficial

84 mutations (*37*) (Fig. 2a). We first examine $n = 3$ loci, corresponding to biallelic mutations in

85 *topA*, *spoT*, and *pykF* (Fig. 1c,d). Epistasis was generally low in magnitude (*37,40*), and occurs

86 in two ways: (*i*) either from merging groups of groups of simplices (c.f. BC + AFE in line #2

87 of Fig. 1d), which indicates a complex interaction, or (*ii*) from merging a single simplex, c.f. D,

88 with the aggregated rest of the simplices (c.f. line #1 of Fig. 1d), much like a dominant effect in

89 the NK model (*14*). This second way is consistent with a fitness landscape distortion, which oc-

90 curs when certain mutations influence the interactions of many other genes (*41*). Geometrically,

91 such a distortion constitutes a vertex split (*42*). We next add a fourth biallelic mutation, in the

92 *glmUS* locus (Fig. 2b,c), encoding peptidoglycan availability, which is an essential component
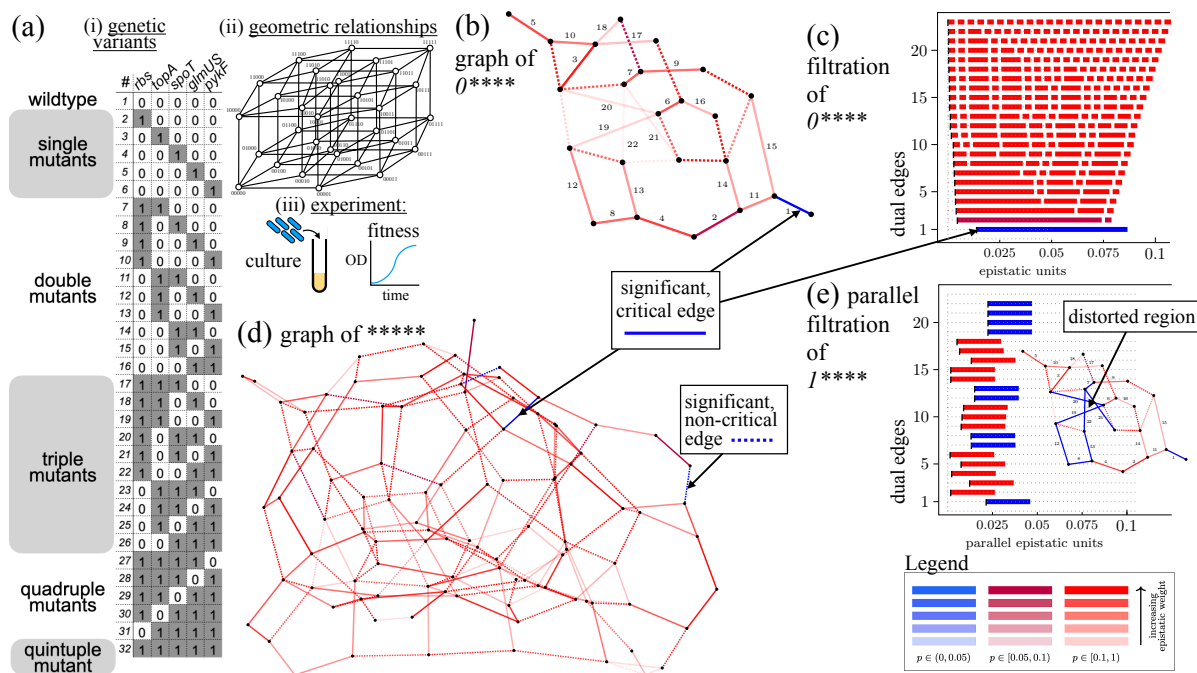
93 of the cell wall.

6

Figure 2: ***E. coli* evolution is guided by epistatic landscape distortions.** (a) (*i*) *E. coli* mutants examined (*37*), (*ii*) their geometric relationships, and (*iii*) experimental approach to measure fitness. (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 (*wildtype*). Note that the left edge of the bars in (c) indicates there is very little epistatic weight 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 Khan data set. Black indices in (b) label the critical dual edges of $\mathcal{S}(h)$. (e) In the parallel filtration, for $1****$, where the *rbs* mutation is present, the landscape is disorted by a concentrated area of higher epistasis. Inset: graph in (b) recolored with weights from (e).

The filtration reveals a smooth, additive landscape with one dominant cell where epistasis arises only in the final merge of the filtration (Fig. 2c), meaning the epistatic topography of the entire landscape (Fig. 2d) rests upon the single vertex, 00001, *pykF*. While the previous analysis detected a significant, marginal effect of *pykF* (*37*), filtrations reveal the geometric structure in terms of which specific combinations of loci are responsible for the effect (Fig. 2e): the interaction between the *glmUS*, $\{00001\}$, and *pykF*, $\{00010\}$, genes requires the context of four loci, $\{00000, 01001, 00101, 00011\}$, yet it involves only up to double mutants, suggesting high dimensional epistasis that arises from lower dimensional interactions (Fig. 2c). This conclusion

7

is consistent with recent genome-wide work on trans-gene interactions (*19*), suggesting that complex traits may arise from genome-wide epistasis, where each mutation's contribution to the trait depends on the context of other mutations.

We introduced **parallel transport** (*38, §6.6*) to give a geometric measure of context-dependence for the same set of loci with different bystanders (e.g. species or genes) (see Fig. S1), previously examined by conditional or marginal epistasis (*43*). Examining the Khan data with and without the *pykF* mutation (*37*) (Fig. S2) showed increased significance in 9 out of 20 of the dual edges (Fig. S2), when *pykF* was mutated. Examining the restoration of *pykF* (Fig. S3), only 3 of 22 edges changed significance and just one critical edge lost significance, indicating that the epistasis in this case occurs because the mutation causes new interactions. Thus, the *pykF* mutation appears to enable further evolution during the Lenski experiment (*39*) by distorting the epistatic landscape. *rbs* also generates distortions (Fig. 2e), which can be visualized as a concentrated region of epistasis on the dual graph (Fig. 2e Inset). We found similar features in another genetic data set for the β-lactamase enzyme (*44*) (Appendix B7). Filtrations can thus reveal the specific geometric structure of both the interactions and the context they rely upon.

2.3   Lactobacilli produce microbiome distortions

Up to this point, we have focused on genetic epistasis, but our framework is generalizable to interactions of environmental parameters, including the gut microbiome, for which a framework to identify complex interactions is greatly needed. Like the genome, which is composed of many genes that interact to determine organismal fitness, the microbiome is also composed of many smaller units (bacterial species in this case) that affect host fitness. Hosts are known to select and maintain a certain core set of microbes (*45, 46*); the interactions of these bacteria can affect host fitness (*3*); and it is debated to what extent these interactions are of higher-order, c.f. (*28*). While vertebrates have a gut taxonomic diversity of ≈ 1000 species, precluding study

8

126 of all possible combinations, the laboratory fruit fly, *Drosophila melanogaster*, has naturally

127 low diversity of ≈ 5 stably associated species (*47*).

128 We made gnotobiotic flies inoculated with each combination of a set of $n = 5$ bacteria ($2^5 =$

129 32 combinations) that were isolated from a single wild-caught *D. melanogaster*, consisting of

130 two members of the *Lactobacillus* genus (*L. plantarum* and *L. brevis*) and three members of

131 the *Acetobacter* genus (Fig. 3a). We measured fly lifespan, which we previously identified as a

132 reproducible phenotype that is changed by the microbiome (*3*). Overall a reduction of microbial

133 diversity (number of species) led to an increase in fly lifespan as with a taxonomically similar

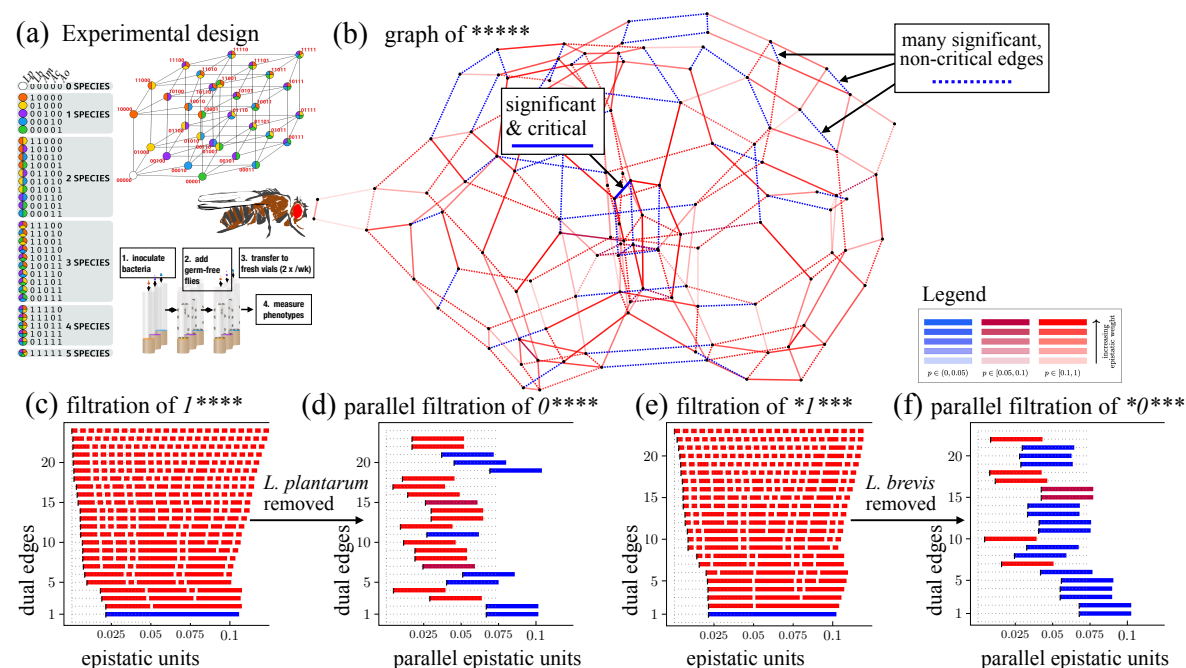134 set of bacteria we examined previously, which came from multiple hosts (*3*).

Figure 3: **Loss of lactobacilli causes global distortion of the microbiome epistatic landscape.** (a) Experimental design for Eble and Gould (*3*) microbiome manipulations in flies. (b) Full graph of $*****$ for the Eble data. (c) Filtration of $\mathcal{S}(h)$ for the 4-face, $1****$, of Eble data, where *L. plantarum* is present, indicates epistasis where two clusters of maximal cells 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.

135 Epistasis was concentrated at the center of the dual graph (Fig. 3b,c), with significant, non-

9

critical edges distributed throughout the graph (Fig. 3c). Examining the parallel transport, we found that the *Lactobacilli* drive changes in the global structure (Fig 3d,e). In 46 out of 128 (36%) interactions, significance changed due to adding or removing a *Lactobacillus* (Fig 3c-f, S7, 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* suppresses epistatic interactions that affect fly lifespan.

Microbiome 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 data sets (Fig. S9, S10, S11, S12, Tables S2, S3, S4, S5), and there was a lack of correlation between the epistatic weights 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.

2.4 Interactions are sparse in higher-dimensions

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 Khan, Eble, and Gould data sets, with a decreasing probability as a function of the face dimension (Table 1). This occurs for three primary reasons. First, the degrees of freedom increase 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 with increasing dimensions. Finally, the absolute number of significant interactions decreases in higher dimensions (Table 1), meaning they are biologically less prevalent. Overall, $\approx 10\%$ of possible dual edges were significant at higher order, with $\approx 1\%$ significant for $n = 5$ dimensions (Table 1),

10

160 suggesting limits to the dimensions of biological complexity.

161 We note that these fewer interactions in high dimensions can and do impact fitness. For ex-

162 ample, the two top 4-dimensional interactions in the Eble microbiome data produce a combined

163 9% effect on fitness (see edges 1 and 2 in (Fig. 3)) with the largest maximal cell accounting for

164 $\approx 5\%$. The relative sparsity makes for a tractable number of these interactions, where we may

165 eventually determine the mechanisms, and filtrations provide a way to identify these.

Table 1: Prevalence of interactions at different levels of complexity in genetics and microbiome data sets. Significant versus all critical dual edges ($p < 0.05$).

Interaction dimension	Dataset: Khan	Dataset: Eble	Dataset: Gould
2:	20/80 (25%)	24/80 (30%)	22/80 (28%)
all higher order:	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%)

166 2.5   Higher-order interactions can arise from lower-order interactions

167 Non-linearities of lower-order interactions can produce interactions in higher dimensions (*40*).

168 In examining the higher-order epistasis present in our data sets, we noted that the clusters where

169 significant epistatic weights occur are often preceded by clusters with nearly significant epistatic

170 weights in lower dimensions (Fig. S4). We developed a graphical approach to distinguish these

171 interactions from those that arise *de novo* (Fig. S20b,c; Appendix B11).

172 Several higher-order interactions in the Gould and Khan data could not be attributed to

173 lower-order effects (Table S6). In particular, they could not be detected from pairwise inter-

174 actions between loci, (c.f. Fig. S20c). As we noted, the 4-dimensional interaction in the *E.*

175 *coli* evolution experiment involved loci with two genes (Fig. 2), whereas in the microbiome,

11

interactions involved loci with four species, indicating different underlying geometries at these different scales of biology (Table S6).

3   Discussion and Conclusions

From an evolutionary perspective, the Red Queen's hypothesis emphasizes how conflicts with other organisms can drive continuous genetic innovation (*48*). We find that epistasis in higher dimensions generates fitness landscape distortions, which could continuously change the fitness landscape to fuel new genomic innovation even in a static environment. This could partially explain the observation of continuous diversification in long term evolution experiments (*49*). In higher dimensions, we lack simple terminology to describe the many types of interactions that may occur, whether between quadruples and singles, pairs and triples, or different genetic backgrounds. We found that biologically-significant interactions in four and five dimensions are sparse and often rooted in lower order, meaning that a limited number of such interactions exist. This extends to higher dimensions the trend that 3-way interactions are often predicted from 2-way interactions (*2, 3, 28*). 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.

This geometric approach could be extended, e.g. to GWAS (*15, 19, 50*), ecosystems (*4, 5*), or neuronal networks (*51*), to discover non-additive higher-order structures at different scales. 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 concepts of curvature for piecewise linear manifolds will be useful (*52*).

4   Acknowledgements

5   Competing interests

The authors declare no competing interests.

6   Supplementary Materials

Materials and Methods

Fig S1 – S22

Tables S1 – S9

References

1. Daniel M Weinreich, Yinghong Lan, Jacob Jaffe, and Robert B Heckendorn. The Influence of Higher-Order Epistasis on Biological Fitness Landscape Topography. *Journal of Statistical Physics*, 172(1):208–225, feb 2018.

2. Elena Kuzmin, Benjamin VanderSluis, Wei Wang, Guihong Tan, Raamesh Deshpande, Yiqun Chen, Matej Usaj, Attila Balint, Mojca Mattiazzi Usaj, Jolanda Van Leeuwen, Elizabeth N Koch, Carles Pons, Andrius J Dagilis, Michael Pryszlak, Jason Zi, Yang Wang,

13

Julia Hanchard, Margot Riggi, Kaicong Xu, Hamed Heydari, Bryan-joseph San Luis, Ermira Shuteriqi, Hongwei Zhu, Nydia Van Dyk, Sara Sharifpoor, Michael Costanzo, Robbie Loewith, Amy Caudy, Daniel Bolnick, Grant W Brown, Brenda J Andrews, Charles Boone, and Chad L Myers. Systematic analysis of complex genetic interactions. *Science (New York, NY)*, 360(April 20), 2018.

3. Alison L. Gould, Vivian Zhang, Lisa Lamberti, Eric W. Jones, Benjamin Obadia, Nikolaos Korasidis, Alex Gavryushkin, Jean M. Carlson, Niko Beerenwinkel, and William B. Ludington. Microbiome interactions shape host fitness. *Proceedings of the National Academy of Sciences*, 115(51):E11951–E11960, 2018.

4. Ted J. Case and Edward A. Bender. Testing for Higher Order Interactions. *The American Naturalist*, 118(6):920–929, 1981.

5. Ian Billick and Ted Case. Higher Order Interactions in Ecological Communities : What Are They and How Can They be Detected? *Ecology*, 75(6):1529–1543, 1994.

6. Jacopo Grilli, György Barabás, Matthew J Michalska-Smith, and Stefano Allesina. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, pages 1–5, jul 2017.

7. Casey Beppler, Elif Tekin, Zhiyuan Mao, Cynthia White, Cassandra McDiarmid, Emily Vargas, Jeffrey H. Miller, Van M. Savage, and Pamela J. Yeh. Uncovering emergent interactions in three-way combinations of stressors. *Journal of the Royal Society Interface*, 13(125), 2016.

8. Ivana Cvijović, Benjamin H. Good, Elizabeth R. Jerison, and Michael M. Desai. Fate of a mutation in a fluctuating environment. *Proceedings of the National Academy of Sciences of the United States of America*, 112(36):E5021–E5028, 2015.

14

9. Lauren E. Nicolaisen and Michael M. Desai. Distortions in genealogies due to purifying selection and recombination. *Genetics*, 195(1):221–230, 2013.

10. Sewall Wright. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*, 1:356–366, 1932.

11. Michael Costanzo, Anastasia Baryshnikova, Chad L Myers, Brenda Andrews, and Charles Boone. Charting the genetic interaction map of a cell. *Current opinion in biotechnology*, 22(1):66–74, feb 2011.

12. Djordje Bajic, Jean C C Vila, Zachary D Blount, and Alvaro Sanchez. On the deformability of an empirical fitness landscape by microbial evolution. *Proceedings of the National Academy of Sciences*, 115(44):11286–11291, oct 2018.

13. John Maynard Smith. Natural selection and the Concept of a Protein Space. *Nature*, 225(February 7):563–564, 1970.

14. S Kauffman and S Levin. Towards a general theory of adaptive walks on rugged landscapes. *Journal of Theoretical Biology*, 128(1):11–45, sep 1987.

15. Örjan Carlborg and Chris S Haley. Epistasis: too often neglected in complex trait studies? *Nature reviews Genetics*, 5(8):618–625, aug 2004.

16. Jeremy Van Cleve and Daniel B. Weissman. Measuring ruggedness in fitness landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, 112(24):7345–7346, 2015.

17. Michael Costanzo, Anastasia Baryshnikova, Jeremy Bellay, Yungil Kim, Eric D Spear, Carolyn S Sevier, Huiming Ding, Judice L Y Koh, Kiana Toufighi, Sara Mostafavi, Jeany Prinz,

15

262 Robert P St Onge, Benjamin VanderSluis, Taras Makhnevych, Franco J Vizeacoumar, Sol-
263 maz Alizadeh, Sondra Bahr, Renee L Brost, Yiqun Chen, Murat Cokol, Raamesh Desh-
264 pande, Zhijian Li, Zhen-Yuan Lin, Wendy Liang, Michaela Marback, Jadine Paw, Bryan-
265 Joseph San Luis, Ermira Shuteriqi, Amy Hin Yan Tong, Nydia van Dyk, Iain M Wallace,
266 Joseph A Whitney, Matthew T Weirauch, Guoqing Zhong, Hongwei Zhu, Walid A Houry,
267 Michael Brudno, Sasan Ragibizadeh, Balázs Papp, Csaba Pál, Frederick P Roth, Guri Gi-
268 aever, Corey Nislow, Olga G Troyanskaya, Howard Bussey, Gary D Bader, Anne-Claude
269 Gingras, Quaid D Morris, Philip M Kim, Chris A Kaiser, Chad L Myers, Brenda J An-
270 drews, and Charles Boone. The genetic landscape of a cell. *Science (New York, NY)*,
271 327(5964):425–431, jan 2010.

272 18. Sean R Collins, Kyle M Miller, Nancy L Maas, Assen Roguev, Jeffrey Fillingham,
273 Clement S Chu, Maya Schuldiner, Marinella Gebbia, Judith Recht, Michael Shales, Huim-
274 ing Ding, Hong Xu, Junhong Han, Kristin Ingvarsdottir, Benjamin Cheng, Brenda An-
275 drews, Charles Boone, Shelley L Berger, Phil Hieter, Zhiguo Zhang, Grant W Brown,
276 C James Ingles, Andrew Emili, C David Allis, David P Toczyski, Jonathan S Weiss-
277 man, Jack F Greenblatt, and Nevan J Krogan. Functional dissection of protein com-
278 plexes involved in yeast chromosome biology using a genetic interaction map. *Nature*,
279 446(7137):806–810, apr 2007.

280 19. Xuanyao Liu, Yang I Li, and Jonathan K Pritchard. Trans Effects on Gene Expression Can
281 Drive Omnigenic Inheritance. *Cell*, 177(4):1022–1034.e6, may 2019.

282 20. Evan A Boyle, Yang I Li, and Jonathan K Pritchard. An Expanded View of Complex Traits:
283 From Polygenic to Omnigenic . *Cell*, 169(7):1177–1186, jun 2017.

284 21. Juannan Zhou and David M. McCandlish. Minimum epistasis interpolation for sequence-
285 function relationships. *Nature Communications*, 11(1):1782, 2020.

22. David M. McCandlish. Long-term evolution on complex fitness landscapes when mutation is weak. *Heredity*, 121(5):449–465, 2018.

23. Molly Schumer, Chenling Xu, Daniel L Powell, Arun Durvasula, Laurits Skov, Chris Holland, John C Blazier, Sriram Sankararaman, Peter Andolfatto, Gil G Rosenthal, and Molly Przeworski. Natural selection interacts with recombination to shape the evolution ofhybrid genomes. *Science*, 360(May 11):656–660, 2018.

24. Daniel M Weinreich, Yinghong Lan, C Scott Wylie, and Robert B. Heckendorn. Should evolutionary geneticists worry about higher-order epistasis? *Current Opinion in Genetics & Development*, 23(6):700 – 707, 2013. Genetics of system biology.

25. Zachary R. Sailer and Michael J. Harms. High-order epistasis shapes evolutionary trajectories. *PLOS Computational Biology*, 13(5):1–16, 05 2017.

26. Michael J. McDonald, Daniel P. Rice, and Michael M. Desai. Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature*, 531(7593):233–236, 2016.

27. Christoph Ratzke, Julien Barrere, and Jeff Gore. Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology & Evolution*, pages 1–21, feb 2020.

28. Jonathan Friedman, Logan M Higgins, and Jeff Gore. Community structure follows simple assembly rules in microbial microcosms. *Nature Publishing Group*, 1:1–7, mar 2017.

29. Deepika Sundarraman, Edouard A Hay, Dylan M Martins, Drew S Shields, Noah L Pettinari, and Raghuveer Parthasarathy. Quantifying multi-species microbial interactions in the larval zebrafish gut. *bioRxiv*, pages 1–23, may 2020.

17

30. Philippe Piccardi, Björn Vessman, and Sara Mitri. Toxicity drives facilitation between 4 bacterial species. *Proceedings of the National Academy of Sciences of the United States of America*, 2019.

31. Alicia Sanchez-Gorostiaga, Djordje Bajić, Melisa L. Osborne, Juan F. Poyatos, and Alvaro Sanchez. High-order interactions distort the functional landscape of microbial consortia. *PLoS Biology*, 17(12):1–34, 2019.

32. Heather J. Cordell. Detecting gene-gene interactions that underlie human diseases. *Nature Reviews Genetics*, 10(6):392–404, 2009.

33. Niko Beerenwinkel, Lior Pachter, and Bernd Sturmfels. Epistasis and shapes of fitness landscapes. *Statist. Sinica*, 17(4):1317–1342, 2007.

34. Amy Berrington de González and D. R. Cox. Interpretation of interaction: A review. *Ann. Appl. Stat.*, 1(2):371–385, 12 2007.

35. Jakub Otwinowski and Joshua B. Plotkin. Inferring fitness landscapes by regression produces biased estimates of epistasis. *Proceedings of the National Academy of Sciences of the United States of America*, 111(22), 2014.

36. Ingileif B Hallgrimsdottir and Debbie S Yuster. A complete classification of epistatic two-locus models. *BMC Genetics*, 9(1):17, jan 2008.

37. Aisha I. Khan, Duy M. Dinh, Dominique Schneider, Richard E. Lenski, and Tim F. Cooper. Negative epistasis between beneficial mutations in an evolving bacterial population. *Science*, 332(6034):1193–1196, 2011.

38. Holger Eble, Michael Joswig, Lisa Lamberti, and William B. Ludington. Cluster partitions and fitness landscapes of the drosophila fly microbiome. *Journal of Mathematical Biology*, May 2019.

39. Jeffrey E. Barrick and Richard E. Lenski. Genome dynamics during experimental evolution. *Nature Reviews Genetics*, 14(12):827—839, 2013.

40. Zachary R Sailer and Michael J Harms. Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps. *Genetics*, 205(3):1079–1088, mar 2017.

41. Hiroshi C. Ito and Akira Sasaki. Evolutionary branching in distorted trait spaces. *Journal of Theoretical Biology*, 489:110152, 2020.

42. Sven Herrmann and Michael Joswig. Splitting polytopes. *Münster J. Math.*, 1:109–141, 2008.

43. J. L. Gill. Effects of Finite Size on Selection Advance in Simulated Genetic. *Australian journal of biological sciences*, 18(1508):599–617, 1965.

44. Longzhi Tan, Stephen Serene, Hui Xiao Chao, and Jeff Gore. Hidden randomness between fitness landscapes limits reverse evolution. *Phys. Rev. Lett.*, 106:198102, May 2011.

45. R E Ley, M Hamady, C Lozupone, P J Turnbaugh, R R Ramey, J S Bircher, M L Schlegel, T A Tucker, M D Schrenzel, R Knight, and J I Gordon. Evolution of Mammals and Their Gut Microbes. *Science (New York, NY)*, 320(5883):1647–1651, jun 2008.

46. Alice Risely. Applying the core microbiome to understand host–microbe systems. *Journal of Animal Ecology*, 89(7):1549–1558, 2020.

47. William B. Ludington and William W. Ja. Drosophila as a model for the gut microbiome. *PLOS Pathogens*, 16(4):1–6, 04 2020.

48. Leigh Van Valen. Molecular evolution as predicted by natural selection. *Journal of Molecular Evolution*, 3(2):89–101, 1974.

49. Benjamin H. Good, Michael J. McDonald, Jeffrey E. Barrick, Richard E. Lenski, and Michael M. Desai. The dynamics of molecular evolution over 60,000 generations. *Nature*, 551(7678):45–50, 2017.

50. Gang Fang, Wen Wang, Vanja Paunic, Hamed Heydari, Michael Costanzo, Xiaoye Liu, Xiaotong Liu, Benjamin Vandersluis, Benjamin Oately, Michael Steinbach, Brian Van Ness, Eric E Schadt, Nathan D Pankratz, Charles Boone, Vipin Kumar, and Chad L Myers. Discovering genetic interactions bridging pathways in genome-wide association studies. *Nature Communications*, 10(4274):1–18, 2019.

51. Michael W. Reimann, Max Nolte, Martina Scolamiero, Katharine Turner, Rodrigo Perin, Giuseppe Chindemi, Paweł Dłotko, Ran Levi, Kathryn Hess, and Henry Markram. Cliques of neurons bound into cavities provide a missing link between structure and function. *Frontiers in Computational Neuroscience*, 11(June), 2017.

52. John M. Sullivan. Curvatures of smooth and discrete surfaces. In *Discrete differential geometry*, volume 38 of *Oberwolfach Semin.*, pages 175–188. Birkhäuser, Basel, 2008.

53. Ewgenij Gawrilow and Michael Joswig. `polymake`: a framework for analyzing convex polytopes. In *Polytopes—combinatorics and computation (Oberwolfach, 1997)*, volume 29 of *DMV Sem.*, pages 43–73. Birkhäuser, Basel, 2000.

54. Daniel M Weinreich, Yinghong Lan, Jacob Jaffe, and Robert B Heckendorn. The influence of higher-order epistasis on biological fitness landscape topography. *Journal of statistical physics*, 172(1):208–225, 2018.

20

55. Jesús A. De Loera, Jörg Rambau, and Francisco Santos. *Triangulations*, volume 25 of *Algorithms and Computation in Mathematics*. Springer-Verlag, Berlin, 2010. Structures for algorithms and applications.

56. Daniel M Weinreich, Nigel F Delaney, Mark A Depristo, and Daniel L Hartl. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science (New York, NY)*, 312(5770):111–114, apr 2006.

57. Deepika Sundarraman, Edouard A. Hay, Dylan M. Martins, Drew S. Shields, Noah L. Pettinari, and Raghuveer Parthasarathy. Higher-order interactions dampen pairwise competition in the zebrafish gut microbiome. *mBio*, 11(5):1–15, 2020.

58. Kabir Husain and Arvind Murugan. Physical Constraints on Epistasis. *Molecular Biology and Evolution*, 2020.

59. Andrés Aranda-Díaz, Benjamin Obadia, Ren Dodge, Tani Thomsen, Zachary F. Hallberg, Zehra Tüzün Güvener, William B. Ludington, and Kerwyn Casey Huang. Bacterial interspecies interactions modulate pH-mediated antibiotic tolerance. *eLife*, 9, 2020.

60. E. D. Weinberger. Fourier and taylor series on fitness landscapes. *Biological Cybernetics*, 65(5):321–330, 1991.

61. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated, 2014.

62. Niko Beerenwinkel, Lior Pachter, Bernd Sturmfels, Santiago F. Elena, and Richard E. Lenski. Analysis of epistatic interactions and fitness landscapes using a new geometric approach. *BMC Evolutionary Biology*, 7(1):60, Apr 2007.

63. E. D. Weinberger. Fourier and taylor series on fitness landscapes. *Biological Cybernetics*, 65(5):321–330, 1991.

64. Hye-Yeon Lee, Shin-Hae Lee, Ji-Hyeon Lee, Won-Jae Lee, and Kyung-Jin Min. The role of commensal microbes in the lifespan of *Drosophila melanogaster*. *Aging*, 11(13):4611–4640, 2019.

Supplementary Materials for:
High dimensional geometry of fitness landscapes identifies master regulators of evolution and the microbiome

Holger Eble,[1] Michael Joswig,[1,2*] Lisa Lamberti[3,4],William B. Ludington[5,6*]

[1]Chair of Discrete Mathematics/Geometry, TU Berlin, Germany

[2]MPI MiS Leipzig, Germany

[3]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland

[4] SIB Swiss Institute of Bioinformatics, Basel, Switzerland

[5] Department of Embryology, Carnegie Institution for Science, USA

[6] Department of Biology, Johns Hopkins University, Baltimore, MD, USA

[*]To whom correspondence should be addressed;
E-mail: joswig@math.tu-berlin.de, ludington@carnegiescience.edu.

A Materials and Methods

A.1 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 3200 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* (*3*).

A.2 Bacterial cultures

Bacteria were cultured on MRS or MYPL, washed in PBS, standardized to a density of 10^7 CFU/mL and 50 μL was inoculated onto the fly food. Strains are indicated in Table S7. See Gould *et al* (*3*) for complete methods.

A.3 Genetics data

Existing genetics data sets were gotten from Sailer and Harms 2017 (*40*) github repository (https://github.com/harmslab/epistasis) or from Tan *et al* (*44*).

For the Khan data in Fig. 2, the fitness function h is defined for (b) by assigning the following normalized values to the 16 genotypes:

$$
\begin{array}{llll}
00000 \mapsto 0.1524 & 01000 \mapsto 0.1745 & 00100 \mapsto 0.1689 & 00010 \mapsto 0.1569 \\
00001 \mapsto 0.1528 & 01100 \mapsto 0.1842 & 01010 \mapsto 0.1756 & 01001 \mapsto 0.1823 \\
00110 \mapsto 0.1718 & 00101 \mapsto 0.1810 & 00011 \mapsto 0.1642 & 01110 \mapsto 0.1836 \\
01101 \mapsto 0.1956 & 01011 \mapsto 0.1858 & 00111 \mapsto 0.1813 & 01111 \mapsto 0.1987 \;\; .
\end{array}
$$

The Tan data set is different from the other fitness values in that only median and mean values are given, meaning we cannot compute p-values to assess the statistical significance. The fitness values are minimum inhibitory concentrations of antibiotics from a well-standardized assay with little experimental variation. Thus, the measurements and our analysis are believed to

24

be robust. We note that the regular subdivision resulting from the corresponding height function of $[0,1]^5$ is degenerate in the sense that it is not a triangulation. This degeneracy arises because the data are discrete antibiotic concentrations with 24 possible values. The repetition of exact values in several cases means a triangulation does not occur. We extended our methods to this degenerate case by restricting the analysis to the faces that do have a triangulation, broadening the application of our approach. We focused on the piperacillin with clavanulate data from (*44*) as it is the better behaved.

A.4   Computational analysis

The filtrations code is available as a `polymake` (*53*) package (cf. `https://github.com/holgereble/EpistaticFiltration`) and the analysis pipeline is available as a `jupyter` notebook.

B   Terminology

Loci (singular **locus**) refer to individual sites in the genome where a mutation may occur, or in the microbiome sense, a locus is a particular bacterial species. We write $[n] := \{1, \ldots, n\}$ for the set of all loci.

   Genotypes, $v = (v_1, \ldots, v_n)$, are vectors of loci with 0/1-coordinates that form points in some fixed Euclidean space $\mathbb{R}^n$, where n is the number of genetic loci or bacterial species considered. In this article we focus on **biallelic** n-locus systems, i.e. genotype sets of the form $V = \{0,1\}^n$ where n is the number of loci and each locus is either 0, absent, or 1, present. For instance, $v = (1, 0, 1)$ denotes a genotype in a 3-locus system $\mathbb{R}^3$, where the first and third loci are mutant and the second is wild type. The set of all genotypes will be denoted by V. The convex hull $P := \mathrm{conv}(V)$ of all genotypes is called the **genotope**. In our setting P is the n-dimensional unit cube $[0,1]^n$ (c.f. (Fig. S21) for a $2D$ projection of $[0,1]^5$).

25

448   A **fitness function** (also called **height function**) associates to each genotype $v \in V$ a quan-

449   tified **phenotype** describing the impact of the genotype on the organism. For example, if the

450   measured phenotype is fitness, h encodes the reproductive output of the genotype.

451   The **fitness landscape** is the pair (V, h), which defines the fitness $h(v)$ for each genotype

452   $v \in V$. Let $v = (v_1, \ldots, v_n) \in V$ be a genotype. Then its **lift** is given by $(v, h(v)) =$

453   $(v_1, \ldots, v_n, h(v)) \in \mathbb{R}^{n+1}$.

454   A set of points $W = \{w^{(1)}, \ldots, w^{(\ell)}\}$ is **affinely independent** if for all real scalars λ_i

455   satisfying $\sum_{i=1}^{\ell} \lambda_i = 0$ the condition $\sum_{i=1}^{\ell} \lambda_i w^{(i)} = 0$ forces $\lambda_i = 0$ for all $i \in \{1, \ldots, \ell\}$.

456   Otherwise W is **affinely dependent**.

457   An **interaction** with respect to a fitness function h occurs between a collection of $k + 2$

458   affinely dependent genotypes $v^{(1)}, \ldots, v^{(k+2)} \in V \subset \mathbb{R}^n$, for $k \leq n$, whose lifts are affinely

459   independent points in $\mathbb{R}^{n+1}$. This is in line with the standard concept of additive epistasis. The

460   number k is the **dimension** of the interaction; throughout we assume that $k \geq 2$.

     Let $U = \{v^{(1)}, \ldots, v^{(\ell)}\}$ be a set of genotypes. Its **support** is the set

$$\mathrm{supp}(U) := \left\{ k \in [n] \,\middle|\, \text{there are distinct } 1 \leq i, j \leq \ell \text{ with } v_k^{(i)} \neq v_k^{(j)} \right\} \ .$$

461   That is, the support is the set of loci where at least two of the given genotypes differ. For

462   example, if $n = 3$ and $U = \{(0, 0, 0), (1, 0, 1), (1, 0, 0)\}$ then $\mathrm{supp}(U) = \{1, 3\}$.

463   The number of loci that vary (0 vs 1) in the support is called the **order** of an interac-

464   tion; this definition agrees with, c.f., (*54*): "We designate interactions among any subset of

465   k mutations as kth-order epistasis.". We give two examples: First, let $n = 2$ and $U =$

466   $\{(0, 0), (0, 1), (1, 0), (1, 1)\} = V$ such that U is an interaction with respect to some fitness

467   function. Then U is an interaction of dimension 2 and order 2. Second, let $n = 3$ and

468   $U = \{(0, 0, 0), (0, 1, 1), (1, 0, 0), (1, 1, 1)\}$ such that, again, U is an interaction with respect to

469   some height function. Then the dimension is 2 and the order is 3. In general, the order is at least

26

470   as large as the dimension, but the two quantities may differ. We say that genes (corresponding

471   to loci) **interact** if they form the support set of an interaction of genotypes.

472   **Remark.** The dimension k of an interaction $v^{(1)}, \ldots, v^{(k+2)}$ with respect to some fitness func-

473   tion agrees with the dimension of the affine span of the given points in $\mathbb{R}^n$. This can be seen

474   as follows. By definition the lifted points $(v^{(1)}, h(v^{(1)})), \ldots, (v^{(k+2)}, h(v^{(k+2)}))$ are affinely in-

475   dependent in $\mathbb{R}^{n+1}$. So their affine span has dimension $k + 1$. As $v^{(1)}, \ldots, v^{(k+2)}$ are affinely

476   dependent, the dimension of their affine span is at most k. Now the affine dimension can only

477   increase by at most one if one coordinate is appended.

B.1   A primer on epistatic filtrations

479   We first explain the biallelic case with $n \geq 2$ loci. In the geometric framework (*33*), two in-

480   teracting loci give rise to four possible genotypes, which form the vertices of a square and may

481   be written as vectors of zeros and ones, indicating the absence (0, wildtype) or the presence (1,

482   mutant) of each locus respectively (Fig. 1b) (*33,38*). The measured phenotypes lift the genotype

483   vertices into 3-space, and there is epistasis corresponding to the volume of the simplex enclosed

484   by the lifted points (see blue simplex in Fig. 1b). Geometrically, the four genotypes involved

485   are fully symmetric, meaning that the sign of the epistasis for $n = 2$ is relative to the choice of

486   a coordinate system. Thus, the sign of epistasis depends on which genotype is considered wild-

487   type. By considering the simplex volume rather than the fold of the upper shell of the simplex,

488   epistatic filtrations do not specify a sign and thus avoid this caveat. However, directionality is

489   considered by parallel transport (see later section). Returning to our explanation, by taking the

490   upper convex hull of all 2^n lifted points and projecting back onto the genotope $[0, 1]^n$ we induce

491   a **subdivision** $\mathcal{S}(h)$; cf. (*38, 55, §2.1*), into **maximal cells** (Fig. 1b). Generically, every maximal

492   cell of $\mathcal{S}(h)$ is an n-dimensional simplex, which is the convex hull of $(n + 1)$ affinely indepen-

493   dent genotypes (Fig. 1c). Importantly, these n-dimensional simplices are the most elementary

27

494 parts into which a fitness landscape can naturally be decomposed.

Our framework generalizes to higher dimensions through a geometric shape called a **bipyramid**, where two satellite vertices, each the apex of one pyramid, are joined to a common set of base vertices. The satellites correspond in the $2D$ example (Fig. 1b) to 00 and 11 and the base to 10 and 01. This is naturally associated with $\mathcal{S}(h)$, set up by the **ridge** (Fig. 1b). For an ordered sequence of $n+2$ genotypes $(v^{(1)}, v^{(2)}, \ldots, v^{(n+2)})$ we let

$$s = \mathrm{conv}\{v^{(1)}, \ldots, v^{(n+1)}\} \quad \text{and} \quad t = \mathrm{conv}\{v^{(2)}, \ldots, v^{(n+2)}\} \ .$$

In other words, s and t form convex hulls. We call such a pair (s,t) a bipyramid with vertices $v^{(1)}, v^{(2)}, \ldots, v^{(n+2)}$. Then we can find the volume of the lifted bipyramid by forming the $(n+2) \times (n+2)$-matrix

$$E_h(s,t) := \begin{pmatrix} 1 & v_{1,1} & v_{1,2} & \cdots & v_{1,n} & h(v^{(1)}) \\ 1 & v_{2,1} & v_{2,2} & \cdots & v_{2,n} & h(v^{(2)}) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & v_{n+2,1} & v_{n+2,2} & \cdots & v_{n+2,n} & h(v^{(n+2)}) \end{pmatrix} , \tag{3}$$

where $v_{i,1}, v_{i,2}, \ldots, v_{i,n}$ are the coordinates of $v^{(i)} \in \mathbb{R}^n$. The **epistatic weight** of the bipyramid (s,t) is

$$e_h(s,t) := \big|\det E_h(s,t)\big| \cdot \frac{\mathrm{nvol}(s \cap t)}{\mathrm{nvol}(s) \cdot \mathrm{nvol}(t)} \ . \tag{4}$$

Here nvol denotes the dimensionally normalized volume. The quantity $\mathrm{nvol}(s \cap t)$ is the relative $(n-1)$-dimensional normalized volume of the **ridge** of the bipyramid, given by the intersection $s \cap t = \mathrm{conv}(v^{(2)}, \ldots, v^{(n+1)})$. We use the notation

$$\{v^{(1)}\} + \{v^{(2)}, \ldots, v^{(n+1)}\} + \{v^{(n+2)}\} \tag{5}$$

495 for the bipyramid (s,t), where the first and last vertices are the satellites and the middle set
496 forms the base. Now the $n+2$ genotypes of the bipyramid form an interaction of dimension n
497 when $e_h(s,t) > 0$.

28

In our regular triangulation $\mathcal{S}(h)$, the two n-dimensional simplices, s and t, are **adjacent** because their intersection $s \cap t$ is a common face of dimension $n-1$.

B.2   Constructing a filtration from the epistasis of adjacent simplices

We visualize the topography of the **epistatic landscape** by forming a **dual graph** of $\mathcal{S}(h)$, where the nodes are the maximal simplices and adjacent simplices form the dual edges. A rugged path is one with more blue edges (Fig. 1d). To each such dual edge we associate an epistatic weight and a label (Fig. 1c, epistatic weights are in shades of blue and red, while labels are in black). In this way, we construct an epistatic landscape that corresponds to the underlying fitness landscape with the ruggedness specified along the dual graph. The **epistatic filtration** of h (Fig. 1e) depicts the path from weakest to highest epistasis by merging adjacent simplices. These diagrams summarize the information contained in epistatic weights and dual graphs, and facilitate comparisons across data sets. But there is important new information contained in epistatic filtrations, which is not directly visible from the dual graph and its epistatic weights. Indeed, a step in the epistatic filtration merges adjacent simplices. We build the complete fitness landscape by stepwise merging of maximal cells, starting from the lowest epistatic weight and stepwise merging adjacent simplices to form a connected **cluster** c.f. (*38*). In this sense, epistatic filtrations encode a global notion of epistasis in higher dimensions by connecting adjacent bipyramids.

To see this, notice that each row of the diagram has a number of bars and a black leftmost line. In the top row the black line marks the epistatic weight of zero (x-coordinate). Each bar is red and corresponds to one maximal simplex of $\mathcal{S}(h)$. In the second row (counting from the top), we see three things: (1) the value of the lowest epistatic weight moves the x-coordinate of the black line slightly to the right. (2) The two maximal simplices of $\mathcal{S}(h)$ corresponding to this epistatic weight are merged into one. These correspond to the two bars in the previous row

29

522 above the new, longer bar in the row. The lengths of the other bars remain unchanged but are

523 shifted horizontally by the epistatic weight in (1). (3) The statistical significance of the epistatic

524 weight giving rise to the merging step, encoded by the colors of the bars; cf. Section B.4.

525   The merging procedure is then repeated for each pair of maximal simplices arising in each

526 epistatic weight until one reaches the highest epistatic weight and the last maximal simplex of

527 $\mathcal{S}(h)$ to be merged with the rest. In this way the indentation of the bar charts increases from top

528 to bottom. The total width of the bars stays constant throughout.

529   Importantly, in the epistatic filtration diagram, not every merging step is displayed; e.g.,

530 in Fig. 1d there are fewer rows than dual edges in Fig. 1c. This is because some steps do not

531 change the resulting fitness landscape (no actual new portion is merged to the previous one). The

532 reported steps are only the ones increasing the connected components of the fitness landscape

533 obtained from the previous merging steps. The epistatic weights corresponding to these steps

534 are the edges in the dual graph which we call **critical** in (*38, §.3.2*).

B.3   Normalized epistatic weights

536 To gain a perspective on the generality of higher-order interactions, it is desirable to compare

537 epistatic landscapes. Different phenotypes have different metrics, making comparisons difficult

538 for current approaches to epistasis. Filtrations are well-suited in this sense. Scaling the height

539 function h by a positive constant does not change the regular triangulation, and thus it does not

540 change the dual graph. In order to compare different data sets, we scale the height function to

541 Euclidean norm one. The epistatic weights are scaled accordingly. The resulting **normalized**

542 **epistatic weights** are measured in **epistatic units**, giving a generalized metric for epistasis.

543   Measuring the effect of context on epistatic interactions is also desirable, e.g. to detect the

544 marginal or conditional effects of a locus (*37*), and these are a natural feature of filtrations. If

545 we fix some k loci and let the remaining $n - k$ loci vary, we obtain a height function, which is

restricted to a face of the genotope $[0,1]^n$. That face has 2^{n-k} vertices, and it is an isomorphic copy of the cube $[0,1]^{n-k}$. For instance, if $n=5$ and we fix the first and the fourth locus to 0, we obtain a 3-dimensional face, which we denote $0**0*$. That is, such a face is written as a string of n symbols in the alphabet $\{0,1,*\}$, where 0 or 1 mark the fixed choices, and $*$ stands for variation. The number of $*$ symbols equals the dimension of the face. Triangulations, their dual graphs, epistatic weights, etc. are well-defined for height functions restricted to faces. This aspect of the theory allows the study of conditional epistatic effects.

B.4 Statistics of epistatic weights

We developed a statistical test to quantify the significance of an interaction associated with a fixed bipyramid; cf. (*38, §4.2*). Here we assume that $h(v)$ is the mean value of the individual phenotype measurements for some number of replicated experiments for the fixed genotype v. To each dual edge we associate a p-value, which is independent of the epistatic weight normalization. If that p-value is below 0.05 we call that dual edge **significant**. It is useful to also consider p-values, which are slightly higher because one can use the shape of the landscape to identify interesting locations for further statistical analysis. To this end we call a dual edge **semi-significant** if $0.05 \leq p < 0.1$.

While it may be possible that this approach misses some biologically relevant interactions (e.g. if they do not correspond to a bipyramid selected by our method), those interactions that we identify carry information that is robust and supported by a statistical model. The fact that not all possible interactions can be approached is an inevitable consequence of the higher dimensional nature of fitness landscapes, also reflected by a very high number of possible regular triangulations of $[0,1]^n$. That number equals 74 for $n=3$ and 87,959,448 for $n=4$, whereas the precise numbers for $n \geq 5$ are unknown; cf. (*55, §6.3*). Thus, filtrations use the data to greatly condense the number of possible interactions considered.

31

The bar colorings in the filtrations of epistatic weights, as in (Fig. S4), reflect the outcome of multiple simultaneous statistical tests (one for each epistatic weight) (*38*).

Significant dual edges at $p < 0.05$ are shown in blue, $0.05 \leq p < 0.1$ in purple, and $p \geq 0.1$ in red.

It may happen that a triangulation has a significant dual edge, which is not critical, whence it does not show in the epistatic filtration. In that case the next critical dual edge becomes blue; so a filtration encodes all significant interactions found by our method.

Remark. By funneling the analysis through the concept of regular triangulations our approach pre-selects interactions, which are most relevant with respect to fitness (*38, §2.2*). Via this major deviation from (*33*) we are able to detect interactions in many data sets, which are biologically plausible; this suggests strongly that our method is particularly good at avoiding false positives. Future work will investigate the relationship to other methods from statistics and signal processing. While most of this is beyond the scope of the present study, in Appendix B12 we offer a first step by comparing with traditional linear regression approaches.

B.5   A synthetic experiment examining how epistatic weights change as a function of the interaction order

Our method calculates significance of detected interactions and normalizes the epistatic weight to the volume of the unit cube of the same dimensionality. We used synthetic data to analyze the method performance. We first examined 468 synthetic filtrations over the 4-dimensional cube, producing 10011 critical dual edges. We found that the epistatic weight is indeed constant as a function of the interaction order, see (Fig. S19a). This indicates that the normalization method is effective. Furthermore, the number of significant interactions decreased as the standard deviation of the input data increased, indicating the statistical method is sensitive to noise, see (Fig. S19b).

32

B.6 A microbiome example in dimension 4

Here $n = 4$, and the fitness function h is defined by assigning the following values to the 16 genotypes:

$$
\begin{array}{llll}
0000 \mapsto 0.2484 ; & 1000 \mapsto 0.2320 ; & 0100 \mapsto 0.1618 ; & 0010 \mapsto 0.1698 ; \\
0001 \mapsto 0.1943 ; & 1100 \mapsto 0.1749 ; & 1010 \mapsto 0.1714 ; & 1001 \mapsto 0.1929 ; \\
0110 \mapsto 0.1668 ; & 0101 \mapsto 0.1608 ; & 0011 \mapsto 0.1617 ; & 1110 \mapsto 0.1643 ; \\
1101 \mapsto 0.1677 ; & 1011 \mapsto 0.1715 ; & 0111 \mapsto 0.1613 ; & 1111 \mapsto 0.1594 \quad .
\end{array}
$$

The vertices $U := \{v^{(1)}, \ldots, v^{(6)}\} \in V$ given by

$$
\begin{array}{lll}
v^{(1)} = (1, 1, 0, 0) ; & v^{(2)} = (0, 0, 0, 0) ; & v^{(3)} = (1, 0, 0, 0) ; \\
v^{(4)} = (1, 1, 0, 1) ; & v^{(5)} = (1, 1, 1, 1) ; & v^{(6)} = (1, 0, 0, 1)
\end{array}
$$

form a bipyramid (s, t) consisting of 4-dimensional simplices s and t as above. The simplices s and t correspond to nodes in the dual graph of $\mathcal{S}(h)$ that share a dual edge recording their adjacency relation as indicated in (Fig. 3b).

In this situation, equation (4) reads

$$
e_h(s, t) = \begin{vmatrix} 1 & 1 & 1 & 0 & 0 & 0.1749 \\ 1 & 0 & 0 & 0 & 0 & 0.2484 \\ 1 & 1 & 0 & 0 & 0 & 0.2320 \\ 1 & 1 & 1 & 0 & 1 & 0.1677 \\ 1 & 1 & 1 & 1 & 1 & 0.1594 \\ 1 & 1 & 0 & 0 & 1 & 0.1929 \end{vmatrix} \cdot \frac{\mathrm{nvol}(s \cap t)}{\mathrm{nvol}(s) \cdot \mathrm{nvol}(t)} = 0.0318 \cdot \frac{\sqrt{2}}{1 \cdot 1} \approx 0.045 \quad .
$$

Since $e_h(s, t) > 0$, the genotype set U defines a 4-dimensional interaction with full support $\{1, 2, 3, 4\}$ and of order 4, according to our terminology of Section Terminology. With a p-value of $0.0005 < 0.05$ the significance test established in (*38, §.4*) rejects the zero hypothesis for $e_h(s, t)$ and therefore proves the effect of the interaction U to be significant. We indicate this fact with the color blue both in the dual graph of $\mathcal{S}(h)$ in (Fig. 3b) and in the epistatic filtration of h in (Fig. 3c).

This example illustrates the following fact of biological interest. For the bacterial combinations $v^{(1)}, v^{(2)}, \ldots, v^{(6)}$ fitness, given by the fitness function h, varies significantly in a

33

608 non-linear way.

B.7 The epistatic landscape within a single enzyme is rugged

610 As a point of comparison with the Khan data set, we re-analyzed data from a fully factorial

611 5-mutation data set in the β-lactamase gene, where each mutation is in a separate residue of the

612 same enzyme (*44, 56*). Due to a lack of the raw replicate data, our computations are based on

613 the reported mean values, and p-values are not calculated. The filtration holds a high magnitude

614 of epistasis (Fig. S5, S6) compared with the Khan data set (Fig. S4, S2); note magnitude on the

615 x-axis. The epistasis arises in many steps (note slope of filtration on left side; (Fig. S5, S6)),

616 consistent with the low number of possible evolutionary paths observed by Weinreich (*56*),

617 and distortions are apparent in the shifted magnitude of epistasis by parallel transport. Our

618 geometric approach also reveals a tiered structure to the epistasis, c.f. the largest weight merges

619 two clusters of simplices (Fig. S5, S6), indicating a more complex epistatic landscape than the

620 Khan data set, where epistasis came from one individual simplex on the periphery of the dual

621 graph.

622 Examining the filtration (Fig. 3d), the epistatic weight (i.e. magnitude) for the microbiome

623 data generated $\approx 5\%$ effect, roughly three times the weight in the Khan data and half that in

624 the Tan β-lactamase landscapes (*44*) (c.f. x-axis between Fig. 3, S4, S5), indicating that the

625 rugosity of microbiome interactions is comparable to genetic ones.

626 To further compare the global effect of context across different datasets, we developed a

627 method to compute epistasis, based on the triangulation of dual landscapes, which we call the

628 epistatic product [Appendix *Product model for epistasic landscape rugosity*] (Fig. S13, S14,

629 S15, S16, S17, S18). The total epistasis was highest for the β-lactamase experiment (*44*), which

630 carries much higher context-dependence than either the microbiome (*3*) or *E. coli* evolution data

631 sets (*37*), indicative of overall high epistasis at the smallest, within enzyme, scale.

34

B.8 Interactions are sparse in higher-dimensions

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 (*28*), and others suggesting a large influence of context-dependent effects (*3*) (*57*), which would make higher-order interactions unpredictable. As we showed in the previous section, few such interactions are biologically meaningful in the context of fitness.

This limitation on epistasis in higher dimensions could arise due to e.g. limited phenotypic dimensions where interactions can be detected or to a lower dimensional manifold that absorbs the majority of the effects (*58*) (e.g. lifespan and fecundity are anti-correlated, making fitness robust to changes in one or the other). Regardless, our analysis shows that significant epistatic interactions are increasingly sparse as the number of dimensions for interaction increase, indicating there exist some limits to biological complexity.

We analyzed the few higher-order interactions in greater detail using a geometric approach. As we noted previously, the interactions in the Khan genetic data (Table 1) are based on a vertex split of the genotype 00001, meaning that the entire epistatic weight of the landscape is balanced by a single maximal cell (Fig. 2).

In contrast, the epistatic filtration of the Eble microbiome data in (Fig. 3) has a much richer texture. There are two significant bipyramids

$$\{01001\} + \{00000, 01000, 01101, 01111\} + \{01100\} \quad 0.0451 \quad \#2$$
$$\{01001\} + \{00000, 01000, 01011, 01111\} + \{01110\} \quad 0.0485 \quad \#1$$

given with their epistatic weights and edge id's, which form a cluster of interactions, indicating a larger topographic feature in the epistatic landscape that relates the interactions between *L. brevis* and increasing numbers of *Acetobacters*. Proximal to these significant cells are two cells with nearly significant statistical support:

$$\{01011\} + \{00000, 01001, 00111, 01111\} + \{01101\} \quad \#8$$
$$\{01011\} + \{00000, 01000, 01001, 01111\} + \{01101\} \quad \#7$$

35

648   with their edge id's (Fig. 3). This invites further research on the bacteria involved. For instance,

649   the interactions could derive from metabolic crossfeeding between the *Acetobacters*, which

650   produce many co-factors, and *L. brevis*, which produces lactate, stimulating *Acetobacter* growth

651   (*59*). Note that the support sets of the bipyramids for all four interactions contain both the wild

652   type 00000 and 01111, which are the maximum and minimum fitness respectively.

B.9   Parallel transport of epistatic weights

654   The notion of parallel transport in a fitness landscape (V, h) was introduced in (*38, §6.6*) as

655   a way to compare geometric and biological information between pairs of parallel facets of the

656   convex polytope $\operatorname{conv} V$. In this work, we extended that notion to include the case of two fitness

657   landscapes, (V, h_1) and (V, h_2), associated to different generic and normalized height functions

658   $h_i : V \to \mathbb{R}, i \in \{1, 2\}$, defined on the same vertex set $V = \{0, 1\}^n$ for some $n \in \mathbb{N}$. To

659   enable meaningful comparisons, we assume that each h_i is normalized and that there is a larger

660   fitness landscape (W, h) with a generic and normalized height function $h : W \to \mathbb{R}$ restricting

661   to h_1 and h_2 on the parallel facets V in W, such that the partition of $\operatorname{conv} W$ induced by h

662   is compatible with the one of $\operatorname{conv} V$ induced by h_1, resp. by h_2. In this setting, we define

663   **normalized epistatic weights** as with Eq. (4) with h the normalized height function and s, t

664   any adjacent simplices forming a bipyramid.

Parallel transports enable us to transport epistatic filtrations along the reflection map

$$\phi \colon V \to V; v = (v_1, v_2, \ldots, v_n) \mapsto (v'_1, v'_2, \ldots, v'_n) \ ,$$

665   with $v'_i = 1 - v_k$ if $i = k$ and $v'_i = v_i$ otherwise. More precisely, let $e_{h_1}(s, t)$ be the normalized

666   epistatic weight associated to a bipyramid of $\mathcal{S}(h_1)$ and let $\phi(e_{h_1}(s, t)) := e_{h_2}(\phi(s), \phi(t))$ be

667   the parallel normalized epistatic weight transported by ϕ. Then the filtration of normalized

668   epistatic weights induces a filtration of parallel normalized epistatic weights. Additionally, to

669   $e_{h_1}(s, t)$ and to $\phi(e_{h_1}(s, t))$ a p-value can unambiguously be associated (*38, §4.1-4.2*). Notice

36

670 that by design epistatic filtrations for $\mathcal{S}(h_1)$ only show normalized epistatic weights associated

671 to critical dual edges, defined as in (*38*). But normalized epistatic weights and their significance

672 can be defined for all bipyramids including the ones associated to noncritical dual edges. This

673 explains the labelling of the parallel transport tables below. There a row is numbered only if

674 the bipyramid corresponds to a critical dual edge in the dual graph of $\mathcal{S}(h_1)$. Noncritical dual

675 edges whose normalized epistatic weight remains non-significant after the parallel transport are

676 omitted. The normalized epistatic weight before (denoted by $e_o = e_{h_1}(s, t)$) and after (denoted

677 by $e_p = \phi(e_{h_1}(s, t))$) the parallel transport, as well as their p-values (denoted by p_o and p_p) are

678 also reported, as well as ratios of these quantities.

679 These parallel transport tables are linked to the epistatic filtration diagrams. Indeed, each

680 numbered row in the table corresponds to the row in the epistatic filtration diagram with the

681 black line set at e_o. It also corresponds to the row with black line set at e_p in the parallel

682 transported filtration diagram.

683 Recall from Section *Statistics of epistatic weights* that there may be dual edges of the tri-

684 angulations which are significant but not critical. Since only the critical dual edges are labeled

685 (by the row number in the epistatic filtration), in our tables for parallel transport these show up

686 as unlabelled rows.

687 Examples for the parallel transport of epistatic filtrations are shown in Figures S1, S2,

688 S3, S5, and S6. The magnitude of the epistasis in the left panels are roughly comparable

689 between data sets due to normalization of the input data. Compare each left panel with its

690 corresponding right panel to observe the relative change in epistasis in the parallel path. Larger

691 changes in epistasis indicate stronger context-dependence of the interaction. For instance, in

692 the first Weinreich comparison (Fig. S5), bar 10 in the right panel has a parallel epistasis greater

693 than the original filtration on the left, indicating context-dependence.

37

B.10   Product model for epistasic landscape rugosity

In this section we offer a new methodological framework to simultaneously study fitness landscapes associated to different height functions. We also provide a measure to quantify how much the height function of the combined fitness landscape differs from the sum of the height functions.

Let U and V be point configurations in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. We think of these point configurations as two sets of genotypes, which may be distinct or not. If we have height functions $\lambda : U \to \mathbb{R}$ and $\mu : V \to \mathbb{R}$, then taking the sum $\lambda + \mu$ point-wise yields a lifting function of the product $U \times V \subset \mathbb{R}^{m+n}$. The cells of the regular subdivision $\mathcal{S}_{U \times V}(\lambda + \mu)$ are products of cells of $\mathcal{S}_U(\lambda)$ with cells of $\mathcal{S}_V(\mu)$. In particular, if λ and μ are generic, i.e., $\mathcal{S}_U(\lambda)$ and $\mathcal{S}_V(\mu)$ are triangulations, then the cells of $\mathcal{S}_{U \times V}(\lambda + \mu)$ are products of simplices.

Now we consider an arbitrary height function $\nu : U \times V \to \mathbb{R}$ on the product of the point configurations. This yields height functions

$$\nu_U : U \to \mathbb{R}, \ u \mapsto \frac{1}{\ell} \sum_{v \in V} \nu(u,v) \quad \text{and} \quad \nu_V : V \to \mathbb{R}, \ v \mapsto \frac{1}{k} \sum_{u \in U} \nu(p,q) \ ,$$

where $k = \#U$, $\ell = \#V$, u is a vertex in U and v is a vertex in V.

Further we define

$$\nu' : U \times V \ , \ (u,v) \mapsto \nu(u,v) - \nu_U(u) - \nu_V(v) \ .$$

Observe that

$$(\lambda + \mu)_U(u) \ = \ \lambda(u) + \frac{1}{\ell} \sum_{v \in V} \mu(v) \quad \text{and} \quad (\lambda + \mu)_V(q) \ = \ \mu(v) + \frac{1}{k} \sum_{u \in U} \lambda(u) \ ,$$

and $(\lambda + \mu)'$ is the height function with constant value $-(\frac{1}{k} \sum_{u \in U} \lambda(u) + \frac{1}{\ell} \sum_{v \in V} \lambda(v))$. Thus $\lambda + \mu$ and $(\lambda + \mu)_U + (\lambda + \mu)_V$ induce the same regular subdivision of $U \times V$. Therefore, we propose to analyze the height function ν' to measure how much ν deviates from the sum of

38

709 two height functions. We can use the techniques from our previous paper (*38*) and apply (all

710 of) them to $\mathcal{S}_{U \times V}(\nu')$ for any given ν. For instance, this allows to measure how independent

711 two different height functions are on the same point set (this is the case $U = V$). We say that ν

712 **decomposes as a product** if $\nu' = 0$.

713 **Example 1.** If $U = V = \{0, 1\}$ are the vertices of the unit interval then $U \times V$ are the vertices

714 of the unit square $[0, 1]^2$. Analyzing $\mathcal{S}(\nu')$ for any given height function ν on the four points

715 $(0, 0), (0, 1), (1, 0)$ and $(1, 1)$ gives back the standard basic example of additive epistasis.

716 **Remark.** Two observations are in order: In (*38, §6.6*) we considered a version of parallel

717 transport to compare epistatic effects, see also Appendix B9. The connection to the product

718 model approach is as follows. Let $V = \{0, 1\}^n$, i.e., the vertex set of the n-dimensional unit

719 cube, be embedded twice, into a pair of parallel facets of the unit $(n+1)$-cube $[0, 1] \times [0, 1]^n$.

720 This occurs in the product model with $U = \{0, 1\}$. If a height function ν on $\{0, 1\} \times U$

721 decomposes as a product then the parallel transport (in both directions) is trivial. Note that the

722 number of dimensions is greater for the product model than for the parallel transport.

723   Additionally, observe that the product model differs from the marginal epistasis framework,

724 which would produce a single number testing if the mutant changes one specific interaction

725 between the genes.

726 **B.10.1   Product model for the Khan data**

727 To illustrate the product model consider the following example from the Khan data. We are

728 interested in detecting if interactions between the *topA*, *spoT*, and *pykF* genes change when the

729 *rbs* gene is mutated. To answer this question we let U and V be 3-cubes inside $[0, 1]^5$ defined

730 by three mutable loci, one for each of the above genes and indicated by $*$, and two fixed loci.

731 The first fixed locus represents the *rbs* gene. It is not mutated in U and mutated in V. The

732 height functions are compared over the three variable loci. Thus the filtration over the product

39

733 model for U and V has four dimensions in this case. A computation reveals that there are no

734 significant dual edges in the epistatic filtration on product model, see (Fig. S13). This indicates

735 that the *rbs* mutant does not affect the interaction landscape.

B.11   Meta-epistatic charts

737 This section deals with the question to which extent higher order epistatic effects are induced by

738 lower dimensional ones or, put in other terms, which lower dimension epistatic effects can be

739 seen in higher dimension. The **meta-epistatic chart** is a diagram drawn on top of the induced

740 epistatic filtrations for some selection of faces of a fixed cube; higher-order interactions induced

741 by lower order interactions are marked as corresponding.

742   In (Fig. S20b) and (Fig. S20c) we exhibit an example for the Eble data set, with 5 loci, where

743 we take the five 4-dimensional faces $0****$, $*0***$, $**0**$, $***0*$ and $****0$ into consideration.

744 Mathematically, these five 4-faces constitute the face figure of the wild type. Fix one 4-face, say

745 $0****$. The induced epistatic filtration on this face shows two blue bars corresponding to dual

746 edges labeled 1 and 2. Each of them refers to the ridge of a bipyramid, which is a 3-dimensional

747 simplex in this case. These two ridges may intersect certain 3-dimensional faces in the right

748 dimension and thus may or may not descend to significant ridges within certain 3-dimensional

749 filtrations. In case of an incidence with a lower dimensional significant ridge, the significant

750 4-dimensional effect is induced by a lower dimensional effect and one may picture this fact as

751 a directed assignment pointing from the lower towards the higher dimensional interaction.

B.12   Comparison with a simple linear regression approach

753 In the theory of fitness landscapes many linear regression approaches have been proposed to

754 study higher-order interactions, c.f. (*21, 34, 40, 60*). In this section, we compare our epistatic

755 weight method to an elementary regression approach using an example from the data.

40

The regression analysis we have in mind assumes that there is a linear relationship between the predictors $X_1, X_2, \ldots, X_n$ (one associated to each locus/dimension of the genotope) and response, or dependent, variables Y (associated to the biological measurements). That is, one assumes that $Y = f(X_1, X_2, \ldots, X_n) + \epsilon$ where $f : \mathbb{R}^n \to \mathbb{R}; (X_1, X_2, \ldots, X_n) \mapsto \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$ and where ϵ is a random error term. The coefficients $\beta_1, \beta_2, \ldots, \beta_n$ are unknown but can be estimated by minimizing the sum of squared residuals associated to the observations pairs (x, y). These observations pairs consisting of a genotype and a measurement associated to it. Notice that more than one measurements are typically associated to a single genotype. With the coefficient estimates one can make predictions for the dependent variable via

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \cdots + \hat{\beta}_n x_n \ . \tag{6}$$

756 The hat symbol $\hat{}$ indicates a prediction, for instance of Y on the basis of $x_i = X_i$, or an estimate
757 for an unknown coefficient.

758 Below, we are interested in the differences between the observed measurements y associated
759 to the genotypes of $[0, 1]^n$, expressed in terms of $x_1, x_2, \ldots x_n$ and the predicated values $\hat{y}$ on
760 the regression hyperplane (6). Notice that the regression analysis remains unchanged after
761 normalizing the height function to Euclidean norm one. Additionally, computing residues for
762 all replicated measurements (when provided) and then take averages builds on the assumption
763 that measurements associated to different genotypes are statistically independent from each
764 other. This assumption is consistent with the one underlying the computation of statistical
765 significances for epistatic weights, following (*38, §. 4.2-4.3*).

766 **Remark.** In the regression setting of (6) there are hypothesis tests (like the F-statistic, t-
767 statistics and p-value) to answer if at least one regression coefficient $\beta_j, 1 \le j \le n$ is nonzero,
768 see for example (*61*). Such statistical approaches are different from the one in (*38, §. 4.2-4.3*),
769 where other hypothesis tests for each epistatic weight were proposed.

41

B.12.1 Regression for Eble data

In the following, we perform a regression analysis focusing on the replicated measurements for the lifespan fitness landscape on $[0, 1]^5$ obtained from Eble and subspaces thereof. Numerical measures of model fit (F-statistic: 2357, with p-value essentially zero, and for 3840 observations and 5 predictors) show that the multiple linear regression model can be considered to be appropriated for this data. Since the epistatic weights of the dual edges are close to zero (≤ 0.02) and are mostly not significant, the above regression analysis conclusion is in line with what we see from the filtration of epistatic weights associated to the same fitness landscapes, see (Fig. S22).

From this example we see that the regression approach provides some general information on higher-order interactions. However, without further assumptions, only one interaction formula is given in terms of a regression hyperplane (6) while the epistatic weight approach gives more fine grained information. This example also illustrate that when the regression model fits the data well (essentially the higher the F-statistics and the more coefficients in the hyperplane equation are significantly non-zero) the epistatic filtration has little horizontal shifts and few significant epistatic weights.

We now proceed repeating the above analysis on some of the bipyramids considered in the parallel analysis for the normalized lifespan Eble data. Regressing over bipyramid 23 in Table S8

$$\{0001\} + \{0000, 1001, 1011, 0111\} + \{1111\}$$

in $0****$ and $1****$ reveals that only two average residues over $0****$ are non-zero (associated to the microbiomes 00000 and 00001), and only one is non-zero over $1****$ (associated to the microbiome 10000). This confirms the two non significant epistatic weights over bipyramid 23 in Table S8.

42

Remark. If minimally dependent sets of points in the genotope are fixed, as in the epistatic weight approach, and one regresses above these points, then the corresponding regression hyperplanes equations are learned from data and the equations generally differ from the epistatic weights given as in (4), but similar biological and geometric conclusions can be drawn. This idea could then be taken further by considering smoothing splines, instead of linear regression, and their relation to epistatic filtrations. From an application point of view, one would obtain an interesting new extension of the concept of epistasis because intermediate genotypes could be assessed, which would correspond to the case of genetically heterogeneous populations of organisms as occur in nature.

Other numerical results for the above regressions are summarized in Table S9. Over $0****$ two coefficients are significantly non-zero (for x_1 and x_4), see top part of Table S9. Similarly, over $1****$ four coefficients are significantly non-zero (x_1, x_2, x_3, x_4), see bottom part of Table S9. The fit of the linear regression models is confirmed by the relatively high values of the F-statistic. Over $0***v$ the F-statistics is 459.1 for a p-value near zero and 720 observations. Over $1****$ the corresponding F-statistics (near zero) is 52.61.

B.13   Comparison with other approaches

Currently the main lines of research to investigate higher-order epistasis in computational biology and related disciplines include the present methods, inspired from discrete polyhedral geometry (*3, 33, 38, 62*); linear regression approaches, c.f. (*21*); methods originating from harmonic analysis, c.f. (*40, 54, 63*); and using correlations between the effects of pairwise mutations, discussed in (*38*).

In a 2-locus, biallelic system, all these methods can easily be recovered from one another; some of them even agree. This is true also for some ecological approaches, including the generalized Lotka-Voleterra equations, which yield a mathematically equivalent form to epistasis for

43

814 certain situations c.f. see equation 9 of (*4*). In higher dimensional systems, these methods re-

815 main conceptually closely related but they generally yield different insights about the problem,

816 such as whether the interactions are significant, what their magnitude is, and what their sign

817 is. Because these previous methods make specific, *a priori* assumptions about the forms of in-

818 teractions, they are limited by these assumptions. Epistatic filtrations add a global perspective,

819 determining the structure of interactions from the shape of the fitness landscape.

B.14   Microbiome data sets

821 In this work, *Drosophila* microbiome fitness landscapes consist of experimental measurements

822 on germ-free *Drosophila* flies inoculated with different bacterial species. The lifespan of ap-

823 proximately 100 individual flies were measured for each combination of bacterial species, giv-

824 ing roughly 3,200 individual fly lifespans for each of the two data sets presented. The experi-

825 mental methods are described in (*3, 64*). The first data set is the exact data presented in (*3, 64*).

826 The second data set is the second set of species with exactly the same methods used in (*3, 64*).

827 The bacterial compositions considered consist of all possible combinations of five species. The

828 species considered can all occur naturally in the gut of wild flies: *Lactobacillus plantarum*

829 (LP), *Lactobacillus brevis* (LB), *Acetobacter pasteurianus* (APa), *Acetobacter tropicalis* (AT),

830 *Acetobacter orientalis* (AO), *Acetobacter cerevisiae* (AC), *Acetobacter malorum* (AM). The 5-

831 member communities both stably persist in the fly gut. For the purposes of this work, we define

832 **stable** as maintaining colonization of the gut when ≤ 20 flies are co-housed in a standard fly

833 vial and transferred daily to fresh food containing 10% glucose, 5% live yeast that has sub-

834 sequently been autoclaved, 1.2% agar, and 0.42% propionic acid, with a pH of 4.5. The total

835 number of species found stably associated with an individual fly is typically between 3 and 8.

836 Consistently, *Lactobacillus plantarum* and *Lactobacillus brevis*, are found with two to three

837 *Acetobacter* species. Less consistently, species of *Enterobacteria* and *Enterococci* occur, and

44

838   these have been described as pathogens. While more strains may be present, for each of the two

839   data sets in the present work, a set of five non pathogen species was chosen, including the two

840   *Lactobacilli* and three *Acetobacter* species. The combinations of species are shown in Table S7.

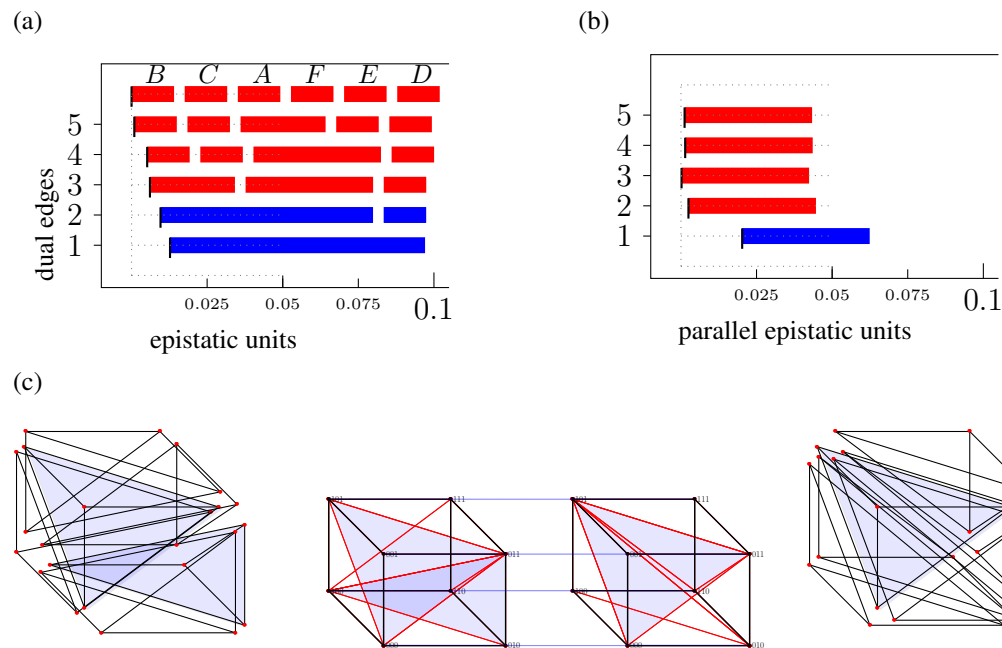841   Different strains of the same species were used in the two data sets.

Figure S1: **Parallel transport from** $0**0*$ **to** $1**0*$ **within the Khan dataset.** (a) Filtration based on the triangulation of $0**0*$. (b) Parallel epistatic weights computed from $1**0*$ for the triangulation based on $0**0*$. (c) The two parallel triangulations (and exploded copies) are depicted. The partitions in the node set are transferred from the cube on the middle left to the cube on the middle right. Exploded versions of these same triangulation on the far left and far right demonstrate the geometry of the simplices generated by the triangulations.

(a)

(b)
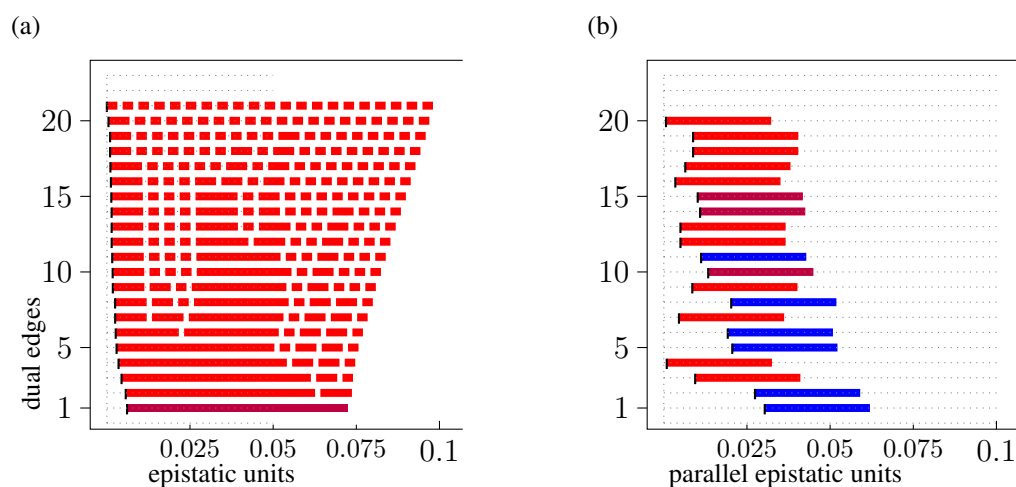
Figure S2: Epistatic filtration and parallel epistatic units for transport from ∗∗∗∗0 to ∗∗∗∗1 within the Khan data.
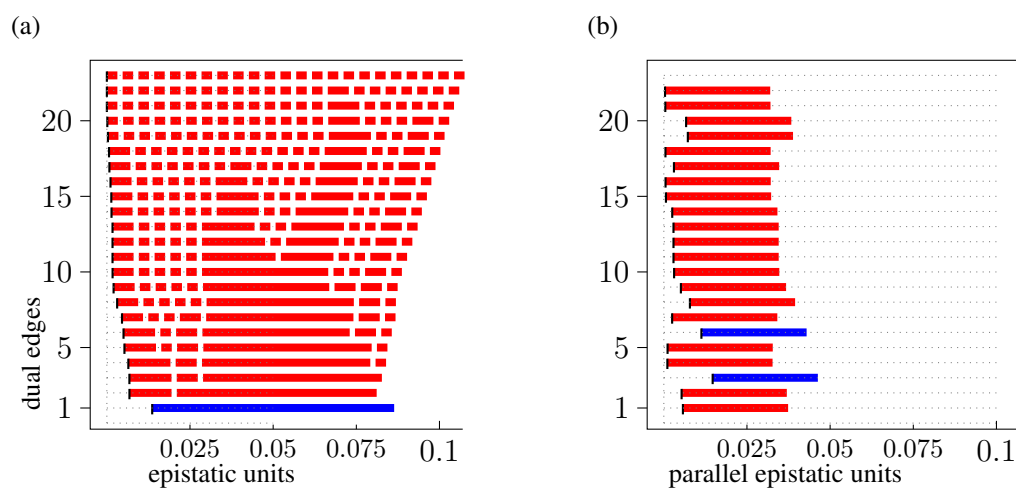
(a)

(b)

Figure S3: Epistatic filtration and parallel epistatic units for transport from ∗∗∗∗1 to ∗∗∗∗0 within the Khan data.
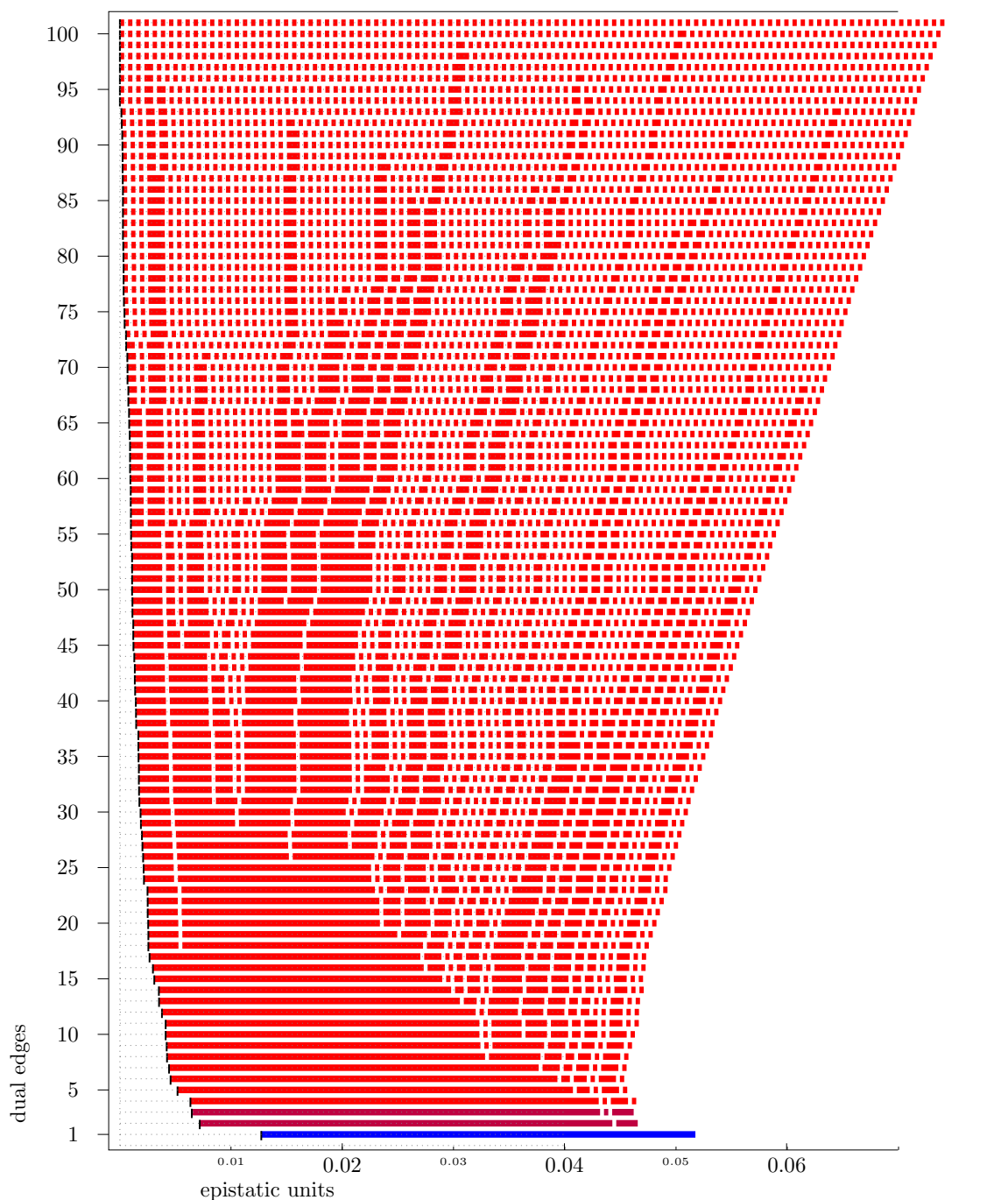
47

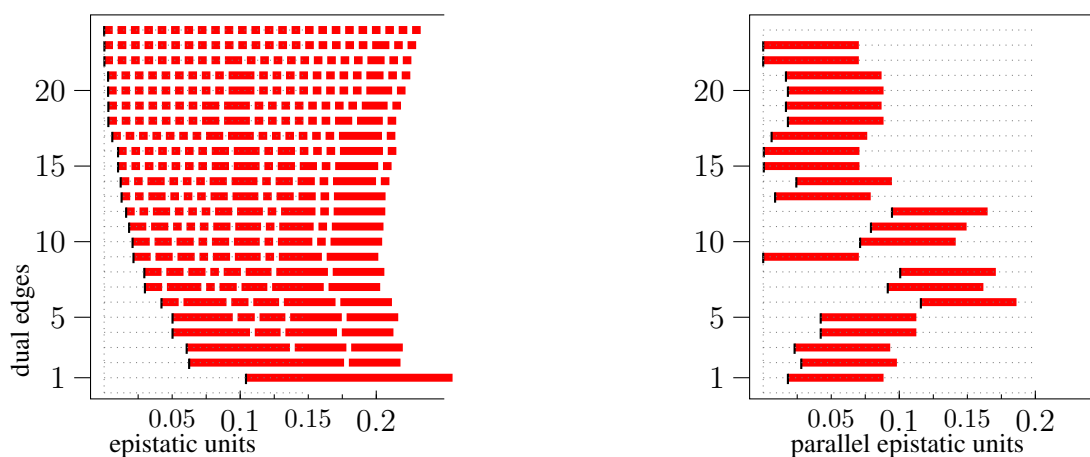Figure S4: **Complete filtration of the Khan data over the whole 5-cube.**

Figure S5: Parallel transport from 0∗∗∗∗ to 1∗∗∗∗ within the Tan data. Analysis based on mean values only; hence there is no color coding for the significance.
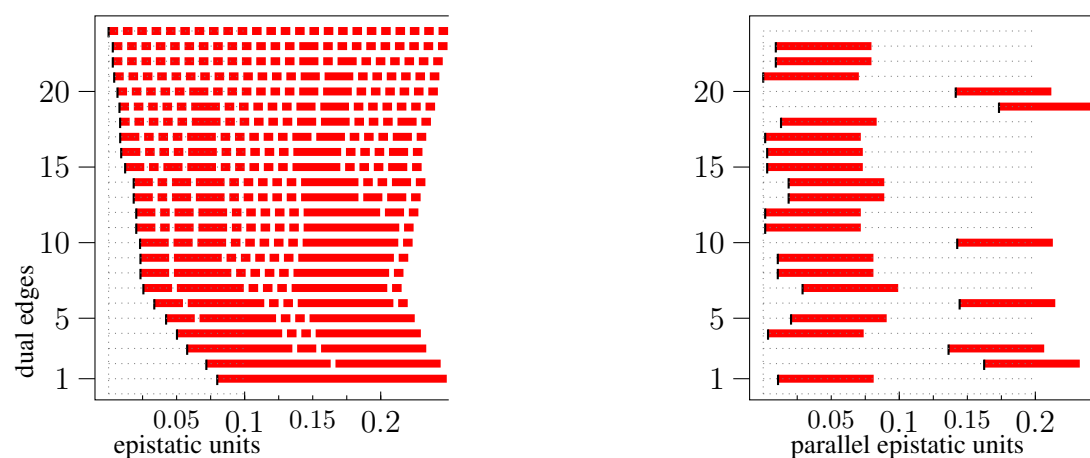

Figure S6: Parallel transport from the face ∗∗0∗∗ to the face ∗∗1∗∗ within the Tan data. Analysis based on mean values only; hence there is no color coding for the significance.
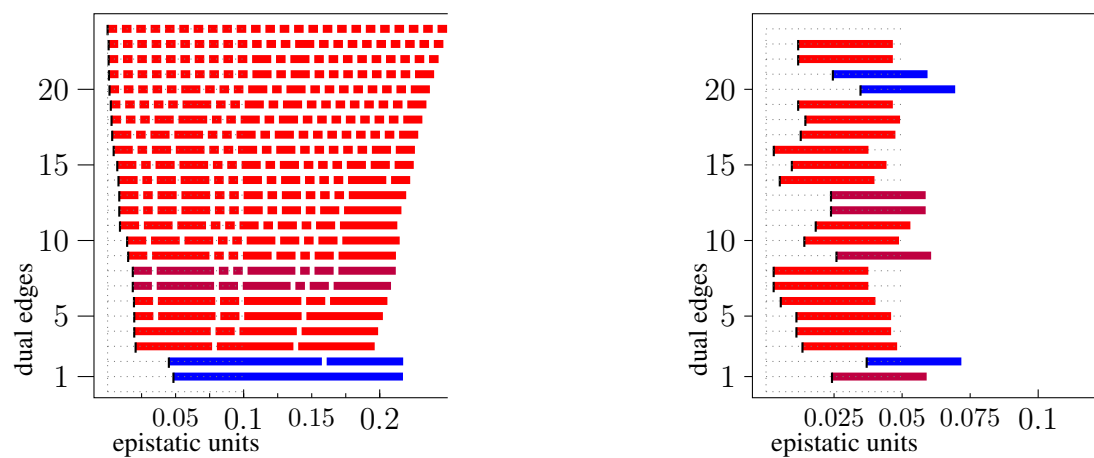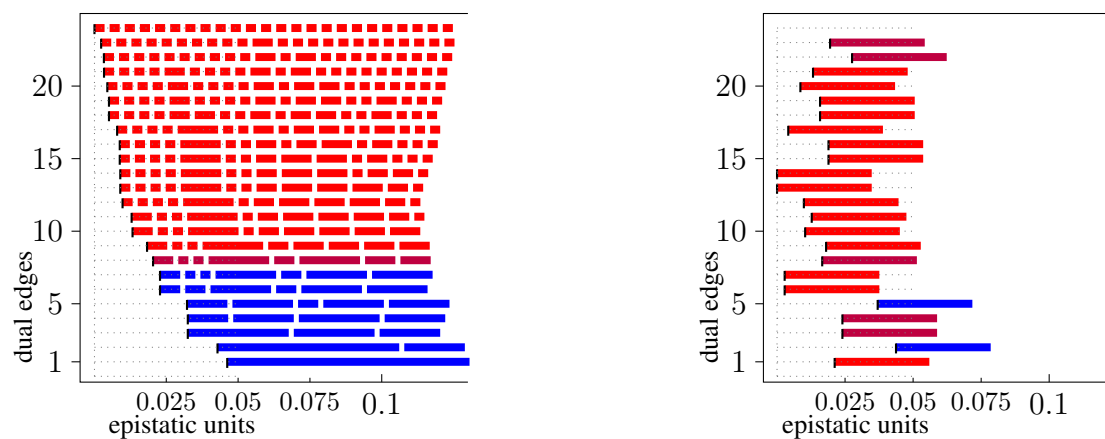
Figure S7: $0***$(Eble) to $1***$(Eble).
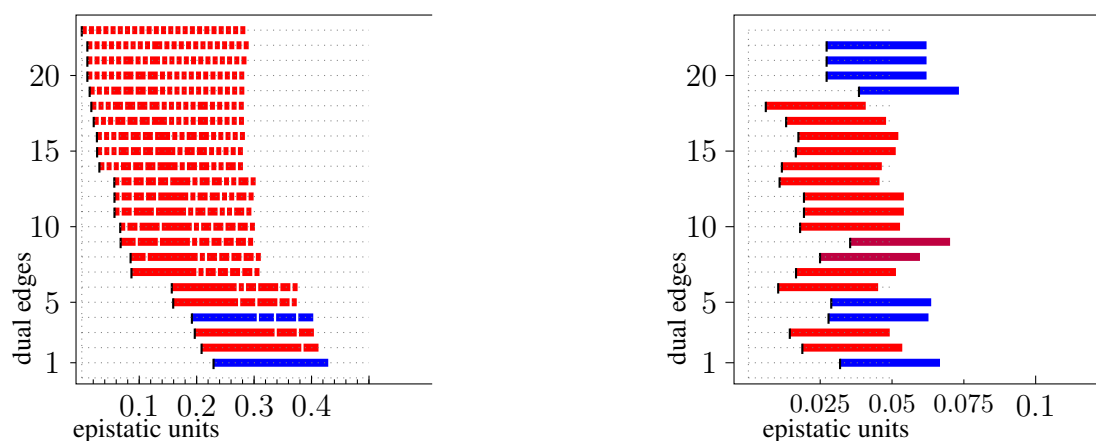

Figure S8: $*0***$(Eble) to $*1***$(Eble).

Figure S9: $0***$(GouldCFU) to $0***$(GouldTTD).
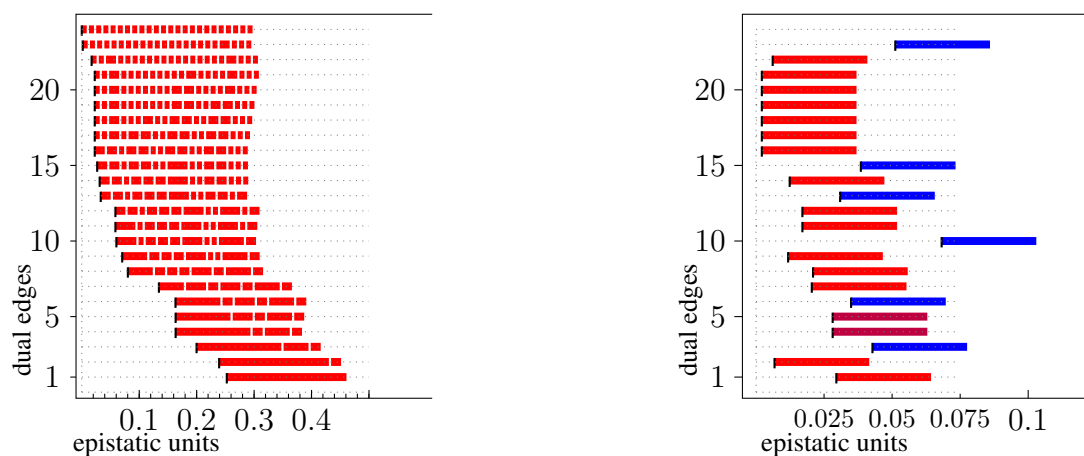
Figure S10: $1****$(GouldCFU) to $1****$(GouldTTD).
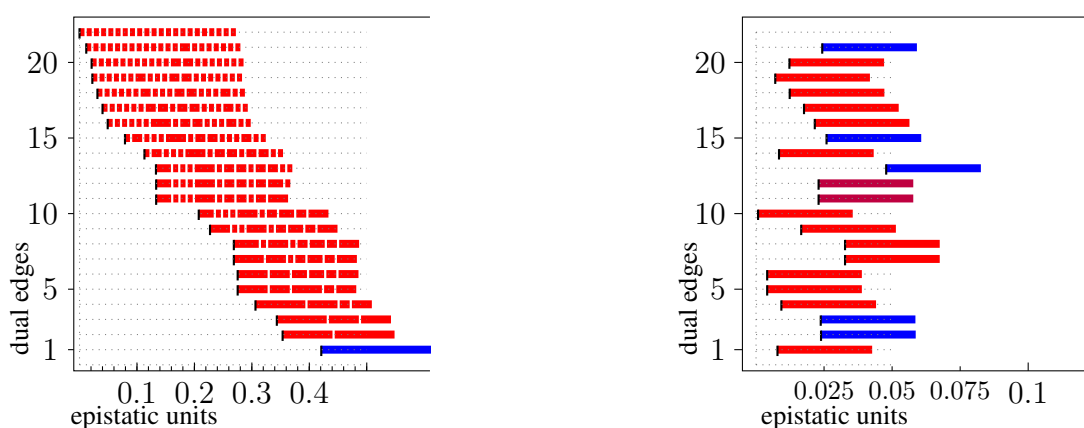
Figure S11: $*0***$(GouldCFU) to $*0***$(GouldTTD).
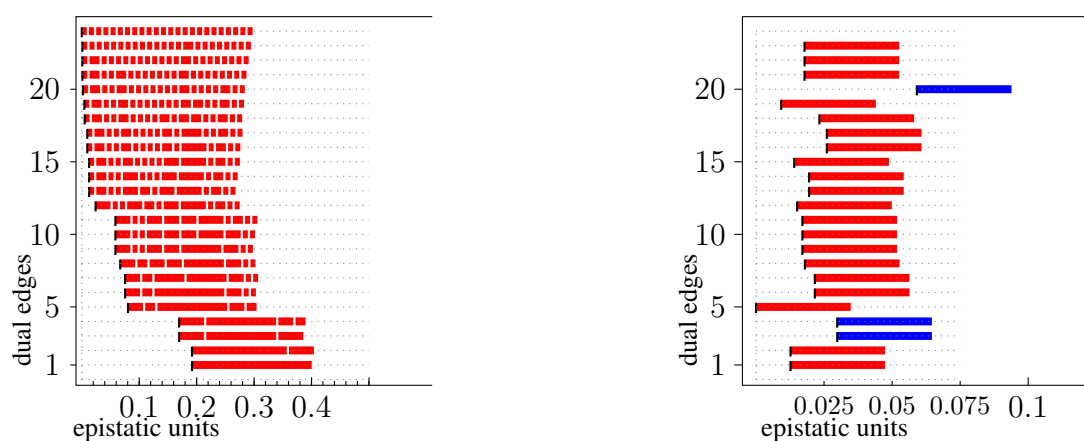
Figure S12: $*1***$(GouldCFU) to $*1***$(GouldTTD).

51
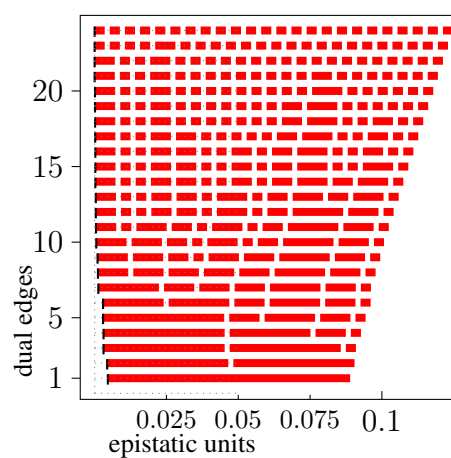
Figure S13: Product model associated to the parallel transport $0**0* \to 1**0*$ within the Khan evolution data, cf. (Fig. S1.)
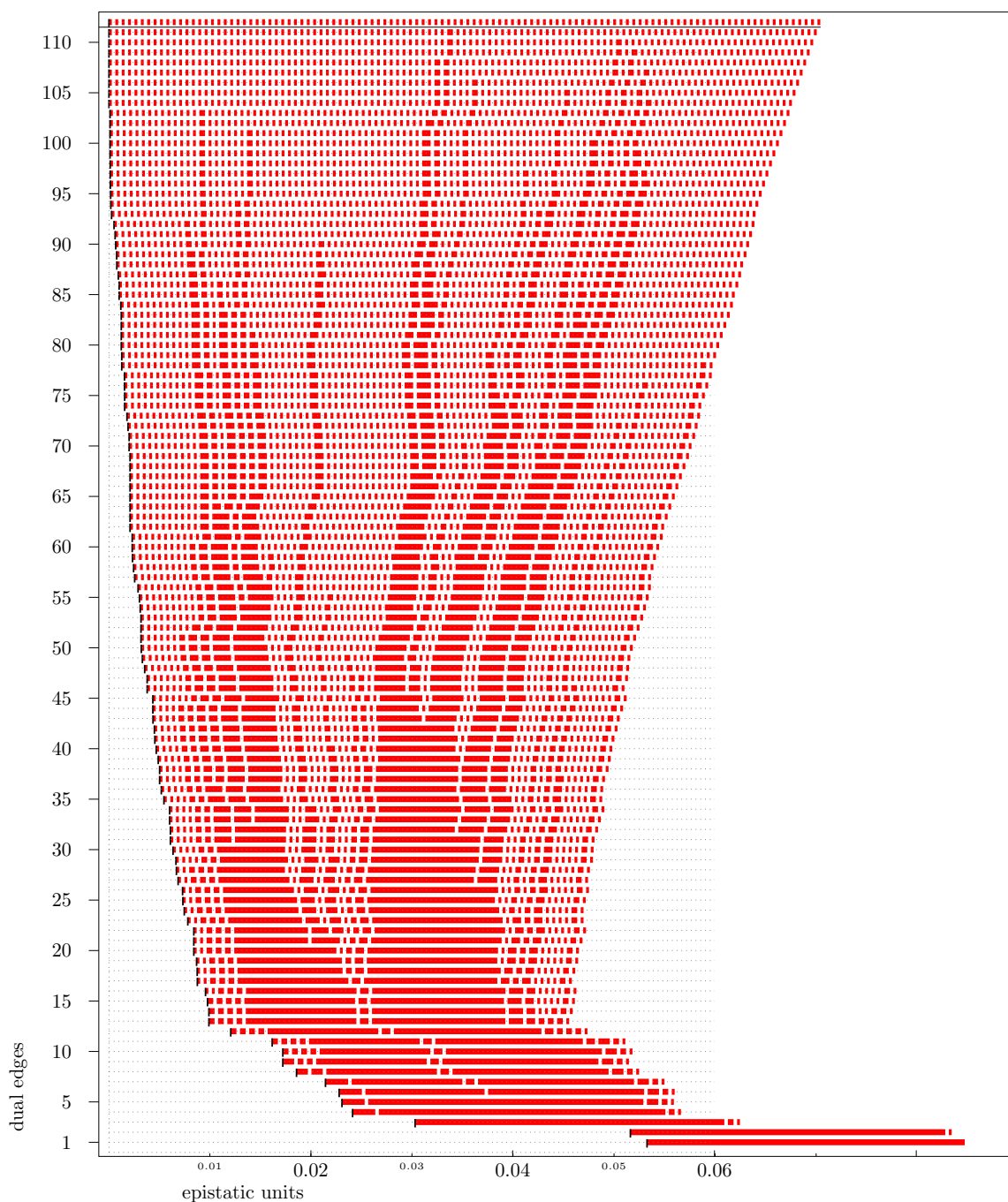
Figure S14: Non-generic product model associated to the parallel transport $**0** \to **1**$ within the Tan data. Its unique non-simplicial maximal cell has 7 vertices and is split into a bipyramid by a slight perturbation of its height values, cf. Theorem 8 of (*38*). The corresponding artificial dual edge has edge label 111 and is indicated by a horizontal line.
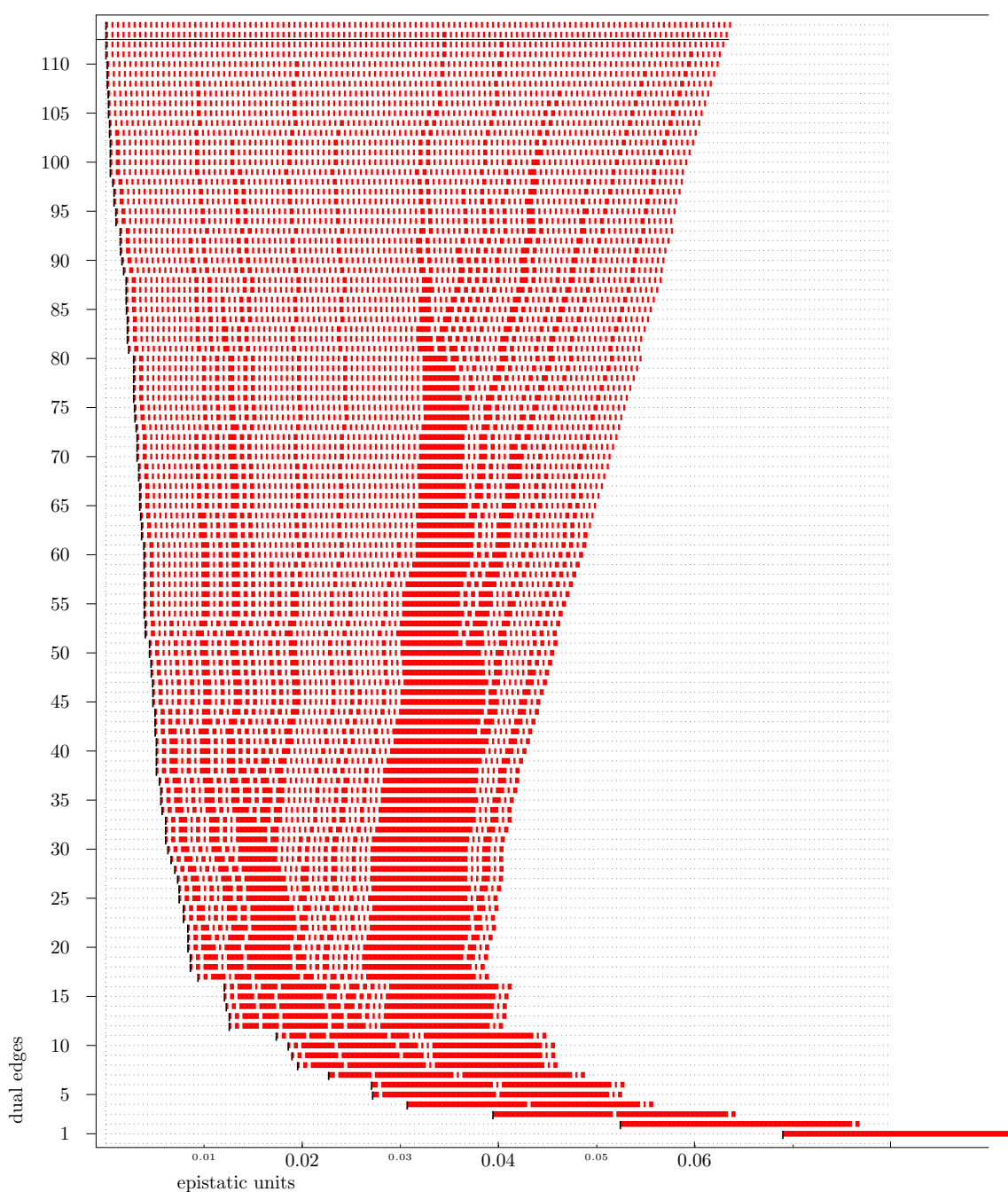
Figure S15: Non-generic product model associated to the parallel transport $0**** \rightarrow 1****$ within the Tan data. There are two non-simplicial maximal cells, both of cardinality 7. As in (Fig. S14) they are split into a bipyramid each at the beginning of the filtration process.
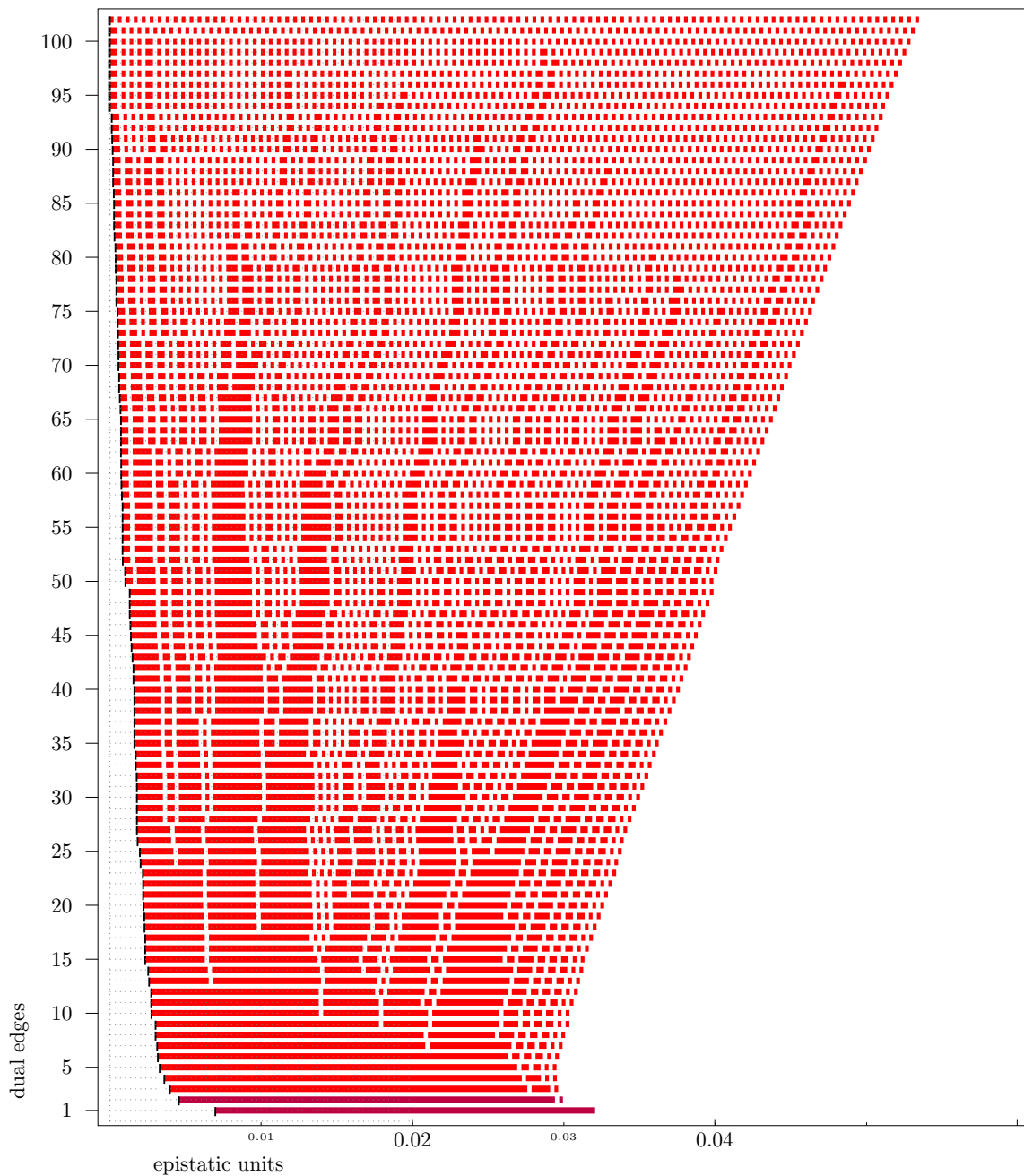
Figure S16: Product model for the parallel transport Khan $***0* \rightarrow ***1*$. The semisignificant bipyramid labeled 2 reads $\{(1000)_o\} + \{(0000)_o, (1010)_0, (0110)_o, (1010)_p, (0011)_p\} + \{(0010)_o\}$ and the semisignificant bipyramid labeled 1 reads $\{(1100)_o\} + \{(1011)_o, (1010)_p, (1001)_p, (0011)_p, (1111)_p\} + \{(1011)_p\}$.
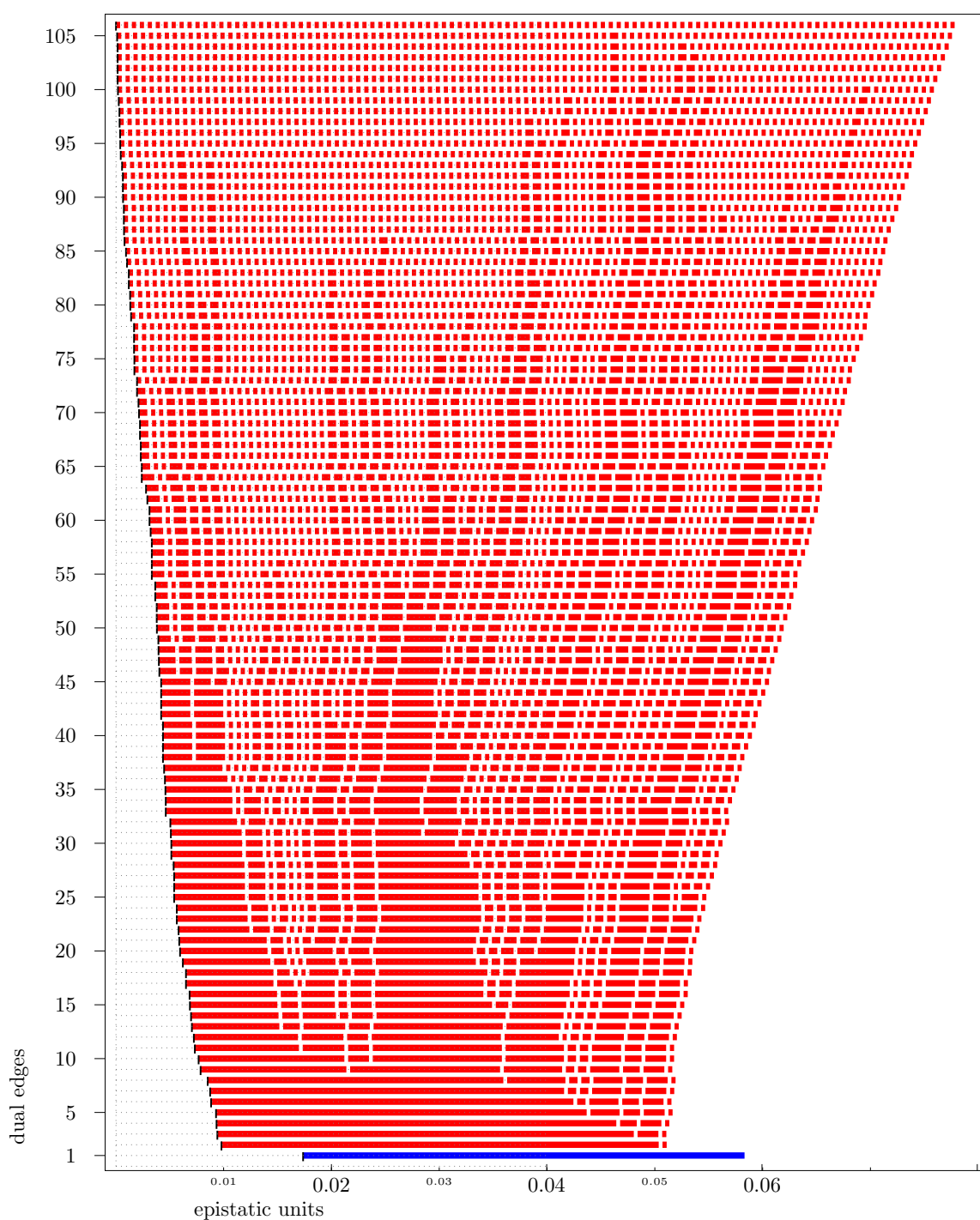
Figure S17: Product model for the parallel transport Eble $0**** \to 1****$. The unique significant bipyramid reads $\{(0001)_o\} + \{(0000)_o, (1001)_o, (0101)_o, (0011)_o, (0001)_p\} + \{(0101)_p\}$.
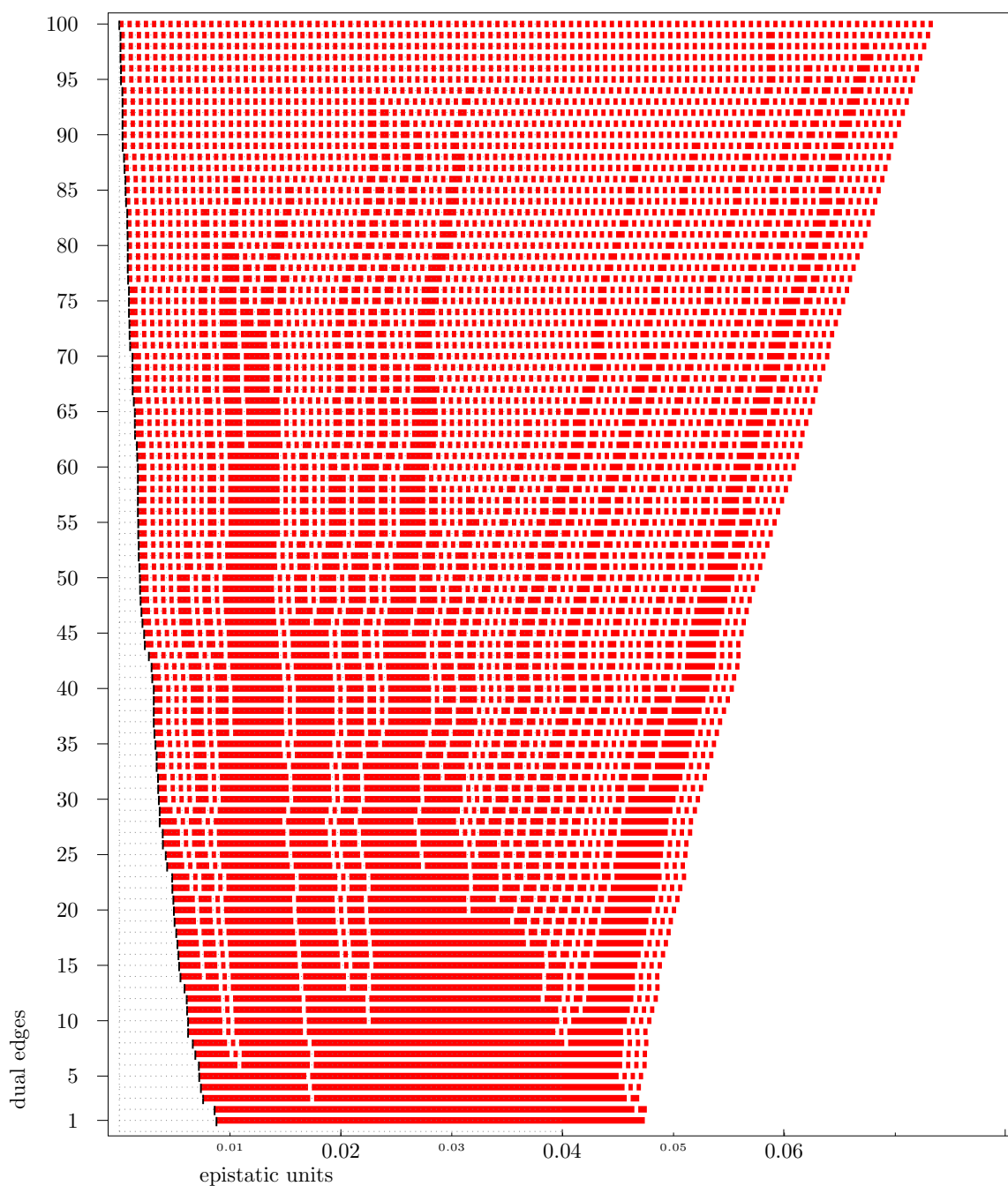
Figure S18: Product model for the parallel transport Eble ∗0∗∗∗ → ∗1∗∗∗.

(a)

(b)
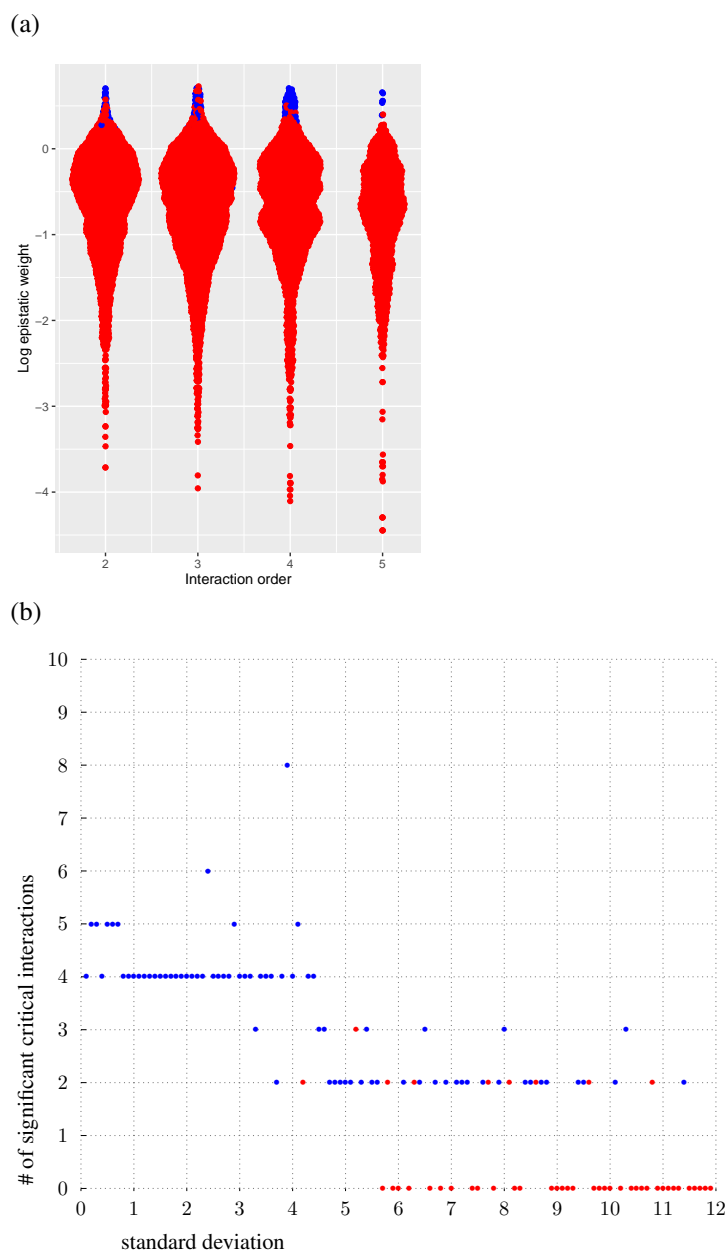
Figure S19: **Synthetic data demonstrate method performance.** Synthetic height functions over the 4-dimensional cube are generated with 100 replicates each and standard deviation as indicated. The heights of the wild type 0000 and 0001 are sampled with mean 53, all the other vertices with mean 50. (a) The distribution of $\log_{10}$-transformed epistatic weights is roughly constant as a function of interaction order, indicating the dimensional normalization is effective. (b) The number of significant interactions decreases as the standard deviation of the input data for each genotype increases. A blue dot is drawn if the interaction is significant and a red dot is drawn otherwise.

58

(a)

(b)

(c)
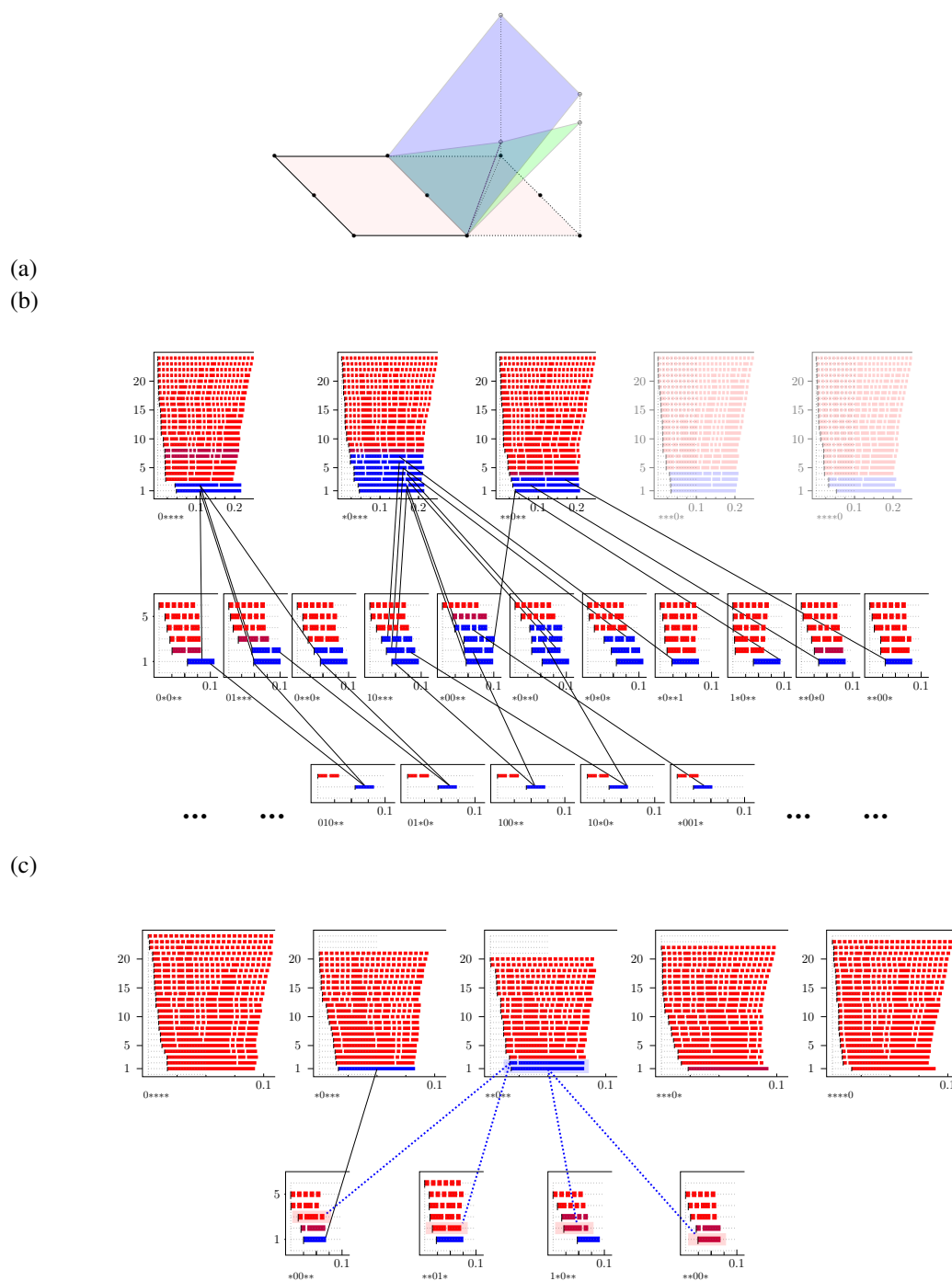
Figure S20: **Meta-epistatic charts illustrate whether or not higher-order interactions arise from lower-order interactions.** (a) Cartoon of the principle underlying meta-epistatic charts. The important loci in the interaction are depicted as black dots in a hyperplane through the genotypes, where the true dimensions of the genotypes are flattened onto the cartoon plane (pink). Higher-order interactions that derive from lower-order interactions occur in a new hyperplane (blue), which magnifies the weights of a subset of the landscape. In contrast, novel higher-order interactions that only arise in higher dimensions do not lie in a single additional hyperplane but instead require at least two additional hyperplanes (green). In (b) and (c) two meta-epistatic charts are represented. In each chart we identify the source of a higher-order interaction for the Eble and Gould data respectively. The results are compiled in Table S6.
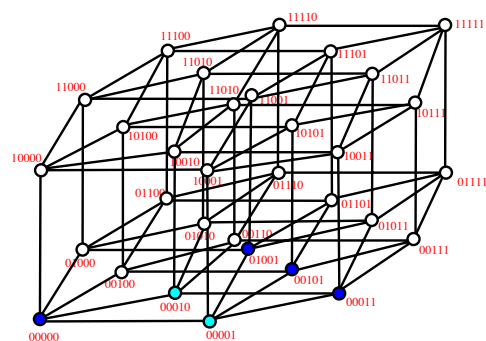
Figure S21: **Vertices of the bipyramid** $\{00001\} + \{00000, 01001, 00101, 00011\} + \{00010\}$ **arising for the Khan data set (37) restricted to** $n = 4$ **loci.** Dark blue dots correspond to common face $s \cap t$ of the bipyramid and light blue dots correspond to the satellite vertices of s and t.

Figure S22: **Complete filtration of the Eble fitness landscape over the whole 5-cube.**

Table S1: Number of circuits of $[0, 1]^n$ and bipyramids among these.

dimensions	circuits	bipyramids	percentage
2	1	1	100.00%
3	20	8	40.00%
4	1348	1088	80.71%
5	353616	309056	87.40%

Table S2: Parallel analysis GouldCFU $0****\rightarrow$ Gould $0****$, non-critical red/red-case omitted.

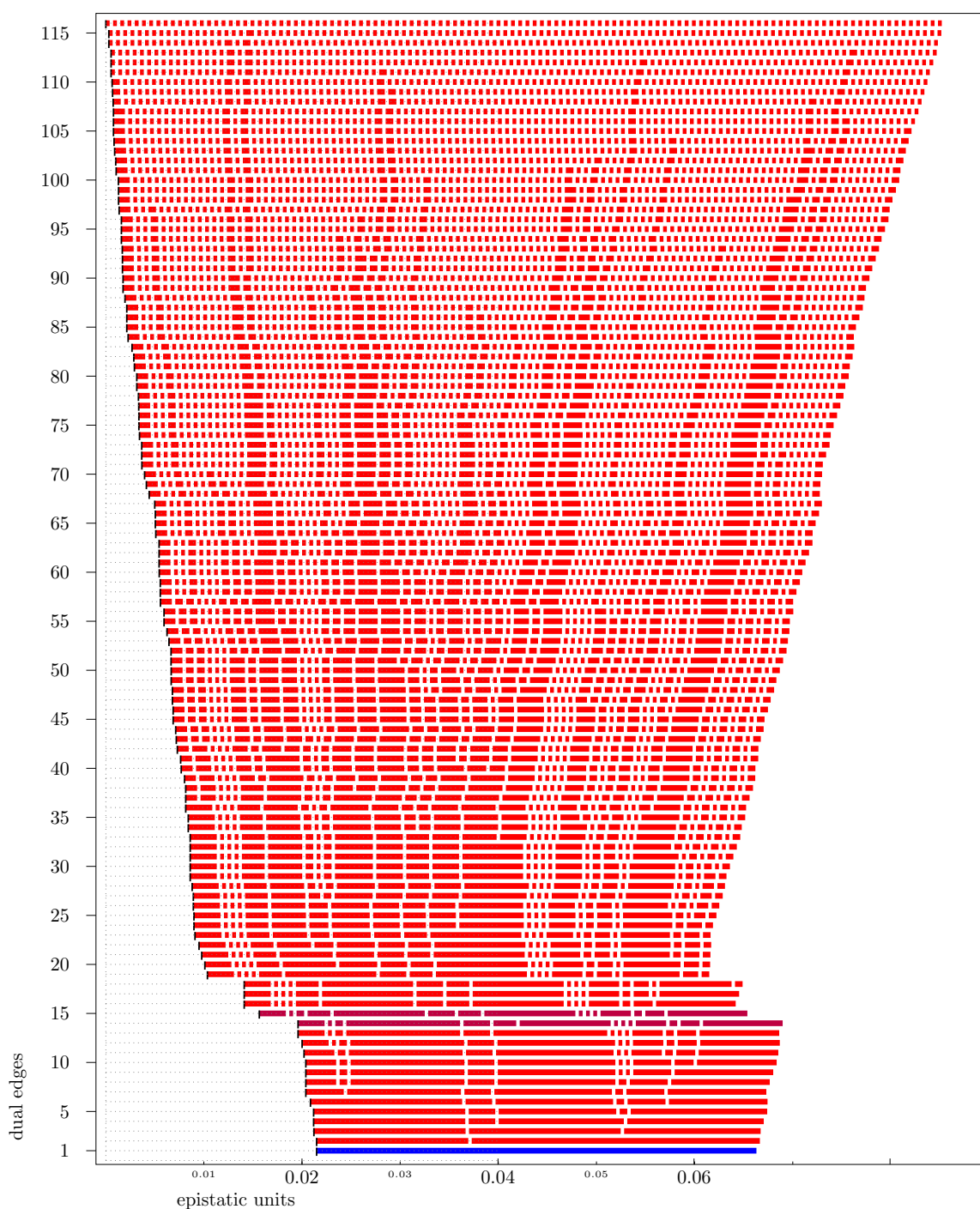| No.	bipyramid	type	e_o	e_p	e_o/e_p	p_o	p_p	p_o/p_p
22	{01001}+{01000,01100,01010,00111}+{00110}	red/blue	0.010	0.027	0.357	0.978	0.038	25.873
21	{01001}+{01000,00100,01100,00111}+{00110}	red/blue	0.010	0.027	0.357	0.978	0.038	25.873
20	{01001}+{01000,00010,01010,00111}+{00110}	red/blue	0.010	0.027	0.357	0.978	0.038	25.873
19	{01001}+{01000,00100,00010,00111}+{00110}	red/blue	0.014	0.039	0.357	0.978	0.038	25.873
18	{01100}+{01001,01110,01101,00111}+{01111}	red/red	0.017	0.006	2.747	0.815	0.677	1.204
17	{01000}+{01100,01010,00110,00111}+{01110}	red/red	0.021	0.013	1.584	0.783	0.433	1.808
16	{00100}+{01100,01001,00101,00111}+{01101}	red/red	0.026	0.017	1.514	0.807	0.302	2.672
15	{01001}+{01100,01010,01110,00111}+{00110}	red/red	0.027	0.017	1.619	0.941	0.231	4.074
14	{00001}+{00010,01001,00011,00111}+{01011}	red/red	0.031	0.012	2.630	0.905	0.312	2.901
13	{01000}+{00100,00010,00001,01001}+{00111}	red/red	0.057	0.011	5.217	0.869	0.479	1.814
12	{00010}+{01000,01010,00110,00111}+{01100}	red/red	0.057	0.019	2.943	0.531	0.148	3.588
11	{00010}+{01000,00100,00110,00111}+{01100}	red/red	0.057	0.019	2.943	0.531	0.148	3.588
	{00010}+{01000,01010,01001,00111}+{01100}	red/blue	0.067	0.047	1.431	0.853	0.032	27.079
	{00010}+{01000,00100,01001,00111}+{01100}	red/blue	0.067	0.047	1.431	0.853	0.032	27.079
10	{01010}+{01001,01110,01011,00111}+{01111}	red/red	0.067	0.018	3.722	0.323	0.186	1.737
9	{00000}+{01000,00100,00010,00001}+{01001}	red/red	0.068	0.035	1.911	0.851	0.086	9.872
8	{00100}+{01000,01100,00110,00111}+{01010}	red/red	0.085	0.025	3.408	0.317	0.083	3.819
7	{01000}+{00100,00010,01001,00111}+{00001}	red/red	0.087	0.017	5.217	0.869	0.479	1.814
	{00100}+{01000,01100,01001,00111}+{01010}	red/blue	0.095	0.052	1.816	0.791	0.019	40.984
	{00100}+{01000,00010,01001,00111}+{01010}	red/blue	0.095	0.052	1.816	0.791	0.019	40.984
6	{01101}+{01001,01110,00111,01111}+{01011}	red/red	0.157	0.010	15.097	0.533	0.362	1.472
5	{00001}+{00100,01001,00101,00111}+{01100}	red/blue	0.159	0.029	5.516	0.541	0.028	19.049
4	{01010}+{00010,01001,01011,00111}+{00011}	blue/blue	0.192	0.028	6.871	0.032	0.042	0.758
3	{00010}+{00100,00001,01001,00111}+{00101}	red/red	0.197	0.014	13.654	0.262	0.211	1.242
2	{01100}+{01010,01001,01110,00111}+{01011}	red/red	0.209	0.019	11.109	0.502	0.175	2.869
1	{01000}+{00010,01010,01001,00111}+{01011}	blue/blue	0.229	0.032	7.188	0.458	0.049	9.271
	{00100}+{00010,00001,01001,00111}+{00011}	blue/red	0.365	0.007	53.243	0.026	0.526	0.049

Table S3: Parallel analysis GouldCFU $1****\rightarrow$ Gould $1****$, non-critical red/red-case omitted.

No.	bipyramid	type	e_o	e_p	e_o/e_p	p_o	p_p	p_o/p_p
23	{11001}+{11000,10101,11101,11011}+{11111}	red/blue	0.002	0.051	0.033	0.962	0.001	1286.096
22	{11100}+{11000,10101,11110,11101}+{11111}	red/red	0.017	0.006	2.799	0.773	0.689	1.122
21	{10000}+{11010,10101,10011,10111}+{11111}	red/red	0.023	0.002	10.615	0.967	0.875	1.105
20	{10000}+{11010,10101,10011,11011}+{11111}	red/red	0.023	0.002	10.615	0.967	0.875	1.105
19	{10000}+{11000,11010,10101,11011}+{11111}	red/red	0.023	0.002	10.615	0.967	0.875	1.105
18	{10000}+{11010,10110,10101,10111}+{11111}	red/red	0.023	0.002	10.615	0.967	0.875	1.105
17	{10000}+{11010,10110,10101,11110}+{11111}	red/red	0.023	0.002	10.615	0.967	0.875	1.105
16	{10000}+{11000,11010,10101,11110}+{11111}	red/red	0.023	0.002	10.615	0.967	0.875	1.105
15	{11011}+{11010,10101,10011,11111}+{10111}	red/blue	0.027	0.039	0.695	0.580	0.012	47.154
14	{10110}+{10000,10010,11010,10111}+{10011}	red/red	0.031	0.012	2.513	0.693	0.277	2.502
13	{11001}+{10000,10001,10101,11011}+{10011}	red/blue	0.033	0.031	1.066	0.388	0.007	54.190
12	{11010}+{11000,10101,11011,11111}+{11101}	red/red	0.059	0.017	3.428	0.905	0.318	2.846
11	{11010}+{11000,10101,11110,11111}+{11101}	red/red	0.059	0.017	3.428	0.905	0.318	2.846
10	{11010}+{10000,11000,10101,11011}+{11001}	red/blue	0.060	0.068	0.881	0.902	0.000	∞
9	{10000}+{11000,11100,10101,11110}+{11101}	red/red	0.070	0.012	5.959	0.897	0.426	2.106
8	{10100}+{10000,11100,10110,10101}+{11110}	red/red	0.080	0.021	3.820	0.430	0.274	1.569
7	{11110}+{11010,10110,10101,11111}+{10111}	red/red	0.134	0.021	6.534	0.130	0.227	0.573
6	{11000}+{10000,11010,10101,11110}+{10110}	red/blue	0.163	0.035	4.659	0.737	0.019	38.586
5	{10010}+{10000,11010,10110,10111}+{10101}	red/red	0.163	0.028	5.788	0.776	0.075	10.402
4	{10010}+{10000,11010,10011,10111}+{10101}	red/red	0.163	0.028	5.788	0.776	0.075	10.402
	{11000}+{11010,10101,11110,11111}+{10110}	red/blue	0.186	0.037	5.000	0.695	0.026	26.834
3	{11000}+{10000,11100,10101,11110}+{10110}	red/blue	0.200	0.043	4.659	0.737	0.019	38.586
2	{11000}+{10000,11010,10101,11011}+{10011}	red/red	0.239	0.007	35.102	0.621	0.628	0.989
1	{11000}+{10000,11001,10101,11011}+{10001}	red/red	0.253	0.030	8.530	0.671	0.104	6.452
	{11110}+{11000,11010,10101,11111}+{11011}	red/blue	0.301	0.026	11.785	0.288	0.035	8.348
	{10001}+{10000,10101,10011,11011}+{11010}	red/blue	0.313	0.039	8.062	0.598	0.014	43.650

Table S4: Parallel analysis GouldCFU $*0***$ $\to$ Gould $*0***$, non-critical red/red-case omitted.

No.	bipyramid	type	e_o	e_p	e_o/e_p	p_o	p_p	p_o/p_p
21	{10001}+{10000,00001,10101,10011}+{00111}	red/blue	0.012	0.024	0.481	0.963	0.026	36.756
20	{00010}+{10000,10010,00011,00111}+{10011}	red/red	0.021	0.012	1.714	0.797	0.270	2.952
19	{10100}+{10000,00100,10110,10101}+{00110}	red/red	0.022	0.007	3.155	0.869	0.717	1.212
18	{10110}+{10000,10010,00111,10111}+{10011}	red/red	0.031	0.012	2.513	0.693	0.277	2.502
17	{00100}+{10000,00110,10101,00111}+{10110}	red/red	0.040	0.018	2.266	0.915	0.290	3.155
16	{00100}+{10000,00110,10110,10101}+{00111}	red/red	0.049	0.022	2.266	0.915	0.290	3.155
15	{00001}+{10000,00100,00010,00111}+{00110}	red/blue	0.079	0.026	3.047	0.698	0.023	30.749
14	{00010}+{10000,10010,00110,00111}+{10110}	red/red	0.113	0.008	13.352	0.295	0.461	0.640
13	{00000}+{10000,00100,00010,00001}+{00111}	red/blue	0.133	0.048	2.775	0.476	0.001	707.281
12	{10010}+{10000,10011,00111,10111}+{10101}	red/red	0.133	0.023	5.788	0.776	0.075	10.402
11	{10010}+{10000,10110,00111,10111}+{10101}	red/red	0.133	0.023	5.788	0.776	0.075	10.402
10	{00011}+{10000,00001,10011,00111}+{10101}	red/red	0.208	0.001	275.689	0.413	0.949	0.435
9	{00010}+{10000,00100,00001,00111}+{00101}	red/red	0.227	0.017	13.654	0.262	0.211	1.242
8	{00100}+{10000,00001,00101,00111}+{10101}	red/red	0.269	0.033	8.193	0.579	0.101	5.733
7	{00001}+{10000,00100,00101,00111}+{10101}	red/red	0.269	0.033	8.193	0.579	0.101	5.733
6	{00001}+{10000,00010,00011,00111}+{10010}	red/red	0.275	0.004	67.167	0.493	0.755	0.653
5	{00001}+{10000,00011,10011,00111}+{10010}	red/red	0.275	0.004	67.167	0.493	0.755	0.653
4	{00110}+{10000,00100,10101,00111}+{00101}	red/red	0.306	0.009	32.813	0.413	0.610	0.677
3	{00110}+{10000,10010,10110,00111}+{10111}	red/blue	0.344	0.024	14.403	0.186	0.035	5.345
2	{00110}+{10000,00010,10010,00111}+{00011}	red/blue	0.354	0.024	14.760	0.175	0.030	5.853
	{00001}+{10000,10101,10011,00111}+{10111}	red/blue	0.408	0.028	14.815	0.108	0.013	8.308
1	{00100}+{10000,00010,00001,00111}+{00011}	blue/red	0.421	0.008	53.243	0.026	0.526	0.049
	{00110}+{10000,10110,10101,00111}+{10111}	red/blue	0.486	0.034	14.403	0.186	0.035	5.345

65

Table S5: Parallel analysis GouldCFU $*1***$ → Gould $*1***$, non-critical red/red-case omitted.

No.	bipyramid	type	e_o	e_p	e_o/e_p	p_o	p_p	p_o/p_p
23	{01100}+{11000,01110,01101,11110}+{11010}	red/red	0.001	0.018	0.054	0.998	0.292	3.418
22	{01100}+{11000,01010,01001,01110}+{11010}	red/red	0.001	0.018	0.054	0.998	0.292	3.418
21	{01100}+{11000,01001,01110,01101}+{11010}	red/red	0.001	0.018	0.054	0.998	0.292	3.418
20	{11001}+{11000,01001,11101,11011}+{11111}	red/blue	0.002	0.059	0.033	0.962	0.001	1286.096
19	{11110}+{11010,01110,01101,11111}+{01111}	red/red	0.005	0.009	0.488	0.945	0.576	1.641
18	{11100}+{11000,01100,11110,11101}+{01101}	red/red	0.005	0.023	0.218	0.952	0.193	4.933
17	{11000}+{11010,01110,01101,11110}+{11111}	red/red	0.010	0.026	0.369	0.981	0.106	9.255
16	{01110}+{11000,11010,01101,11110}+{11111}	red/red	0.010	0.026	0.369	0.981	0.106	9.255
15	{01100}+{11000,01101,11110,11101}+{11111}	red/red	0.013	0.014	0.905	0.866	0.346	2.503
14	{11000}+{01001,11010,01110,01101}+{01111}	red/red	0.013	0.020	0.656	0.974	0.160	6.087
13	{11000}+{01001,11010,01101,11111}+{01111}	red/red	0.013	0.020	0.656	0.974	0.160	6.087
12	{01000}+{11000,01100,01010,01001}+{01110}	red/red	0.024	0.015	1.584	0.783	0.433	1.808
11	{11010}+{11000,01001,01101,11111}+{11101}	red/red	0.059	0.017	3.428	0.905	0.318	2.846
10	{11010}+{11000,01001,11011,11111}+{11101}	red/red	0.059	0.017	3.428	0.905	0.318	2.846
9	{11010}+{11000,01101,11110,11111}+{11101}	red/red	0.059	0.017	3.428	0.905	0.318	2.846
8	{01010}+{01001,11010,01110,01011}+{01111}	red/red	0.067	0.018	3.722	0.323	0.186	1.737
7	{01110}+{01001,11010,01101,01111}+{11111}	red/red	0.075	0.022	3.483	0.841	0.136	6.184
6	{01001}+{11010,01110,01101,01111}+{11111}	red/red	0.075	0.022	3.483	0.841	0.136	6.184
5	{11011}+{01001,11010,01011,11111}+{01111}	red/red	0.081	0.000	1235.241	0.126	0.996	0.127
4	{11000}+{01010,01001,11010,01110}+{01011}	red/blue	0.170	0.030	5.666	0.718	0.026	27.722
3	{11000}+{01001,11010,11011,11111}+{01011}	red/blue	0.170	0.030	5.666	0.718	0.026	27.722
2	{01101}+{01001,11010,01110,01111}+{01011}	red/red	0.192	0.013	15.097	0.533	0.362	1.472
1	{01101}+{01001,11010,01111,11111}+{01011}	red/red	0.192	0.013	15.097	0.533	0.362	1.472

66

Table S6: Significant 4-dimensional interactions, which cannot be seen in lower dimensions, cf. (Fig. S20). The value $p \uparrow$ refers to the p-value of the 4-dimensional bipyramid in question whereas $p \downarrow$ is the p-value of its ridge intersected with the $\cap$-face, cf. (Fig. S20c) for the Gould data.

Data	significant bipyramid	$\cap$-face	$p \uparrow$	$p \downarrow$
Eble	-	-	-	-
Gould				
$**0**$	$\{00010\} + \{00000, 10010, 00011, 11011\} + \{10001\}$	$**01*$	0.041	0.270
		$*00**$	0.041	0.149
	$\{10010\} + \{00000, 11000, 10001, 11011\} + \{01001\}$	$1*0**$	0.041	0.076
		$**00*$	0.041	0.063
Khan				
$0****$	$\{00010\} + \{00000, 01001, 00101, 00011\} + \{00001\}$	$0***1$	0.009	0.052

Table S7: Bacterial species considered in the two microbiome data sets.

	Gould data set	Eble data set
Species 1	*L. plantarum*	*L. plantarum*
Species 2	*L. brevis*	*L. brevis*
Species 3	*A. pasteurianus*	*A. cerevisiae*
Species 4	*A. tropicalis*	*A. malorum*
Species 5	*A. orientalis*	*A. orientalis*

Table S8: Parallel analysis Eble $0{**}{**} \to 1{**}{**}$, non-critical red/red-case omitted.

No.	bipyramid$_s$	type	e_o	e_p	e_o/e_p	p_o	p_p	p_o/p_p
23	{00001}+{00000,01001,01011,00111}+{01111}	red/red	0.001	0.012	0.066	0.953	0.390	2.444
22	{00001}+{00000,01001,01101,00111}+{01111}	red/red	0.001	0.012	0.066	0.953	0.390	2.444
21	{01110}+{00000,00110,01011,01111}+{00111}	red/blue	0.001	0.025	0.041	0.923	0.038	24.226
20	{01110}+{00000,01100,00110,01111}+{00111}	red/blue	0.001	0.035	0.041	0.923	0.038	24.226
19	{00110}+{00000,01100,00111,01111}+{01101}	red/red	0.002	0.012	0.201	0.827	0.303	2.729
18	{00110}+{00000,01100,00101,00111}+{01101}	red/red	0.003	0.014	0.201	0.827	0.303	2.729
17	{01110}+{00000,01000,01100,01111}+{01101}	red/red	0.003	0.013	0.264	0.742	0.251	2.956
16	{00110}+{00000,00010,01011,00111}+{00011}	red/red	0.004	0.003	1.568	0.755	0.843	0.896
15	{00010}+{00000,01010,00110,01011}+{01110}	red/red	0.007	0.010	0.748	0.606	0.488	1.242
14	{01010}+{00000,00010,00110,01011}+{00111}	red/red	0.008	0.005	1.583	0.443	0.639	0.693
13	{01010}+{00000,00110,01110,01011}+{01111}	red/red	0.009	0.024	0.359	0.475	0.062	7.686
12	{01010}+{00000,01000,01110,01011}+{01111}	red/red	0.009	0.024	0.359	0.475	0.062	7.686
11	{00100}+{00000,01100,00110,00101}+{00111}	red/red	0.009	0.018	0.498	0.533	0.269	1.981
10	{01001}+{00000,00001,01101,00111}+{00101}	red/red	0.014	0.014	1.018	0.288	0.313	0.920
9	{00101}+{00000,01100,01101,00111}+{01111}	red/red	0.015	0.026	0.584	0.228	0.062	3.695
	{00101}+{00000,01100,00110,00111}+{01111}	red/blue	0.018	0.040	0.446	0.321	0.035	9.119
8	{01101}+{00000,01001,00111,01111}+{01011}	red/red	0.019	0.003	6.623	0.068	0.800	0.085
7	{01101}+{00000,01000,01001,01111}+{01011}	red/red	0.019	0.003	6.623	0.068	0.800	0.085
6	{01001}+{00000,00001,01011,00111}+{00011}	red/red	0.019	0.005	3.571	0.153	0.689	0.222
5	{01000}+{00000,01010,01110,01011}+{00110}	red/red	0.020	0.011	1.750	0.169	0.443	0.381
4	{01000}+{00000,01100,01110,01111}+{00110}	red/red	0.020	0.011	1.750	0.169	0.443	0.381
3	{01000}+{00000,01001,01011,01111}+{00111}	red/red	0.021	0.013	1.535	0.140	0.339	0.413
2	{01100}+{00000,01000,01101,01111}+{01001}	blue/blue	0.045	0.037	1.215	0.000	0.003	0.176
1	{01001}+{00000,01000,01011,01111}+{01110}	blue/red	0.048	0.024	1.993	0.000	0.056	0.002
	{01100}+{00000,01000,01110,01111}+{01011}	blue/blue	0.064	0.034	1.855	0.000	0.005	0.000
	{00010}+{00000,00011,01011,00111}+{00001}	blue/blue	0.065	0.043	1.518	0.000	0.001	0.001
	{01100}+{00000,01101,00111,01111}+{01001}	blue/red	0.066	0.024	2.775	0.000	0.105	0.000
	{00001}+{00000,00101,01101,00111}+{01100}	blue/blue	0.066	0.033	1.989	0.000	0.009	0.000
	{01001}+{00000,01011,00111,01111}+{00110}	blue/blue	0.068	0.036	1.917	0.000	0.007	0.000
	{01100}+{00000,00110,01110,01111}+{01011}	blue/blue	0.083	0.045	1.829	0.000	0.002	0.000
	{01100}+{00000,00110,00111,01111}+{01011}	blue/red	0.084	0.021	4.035	0.000	0.210	0.000

Table S9: Regressions over {0001}+{0000,1001,1011,0111}+{1111} for normalized lifespan data for Eble 0∗∗∗∗ and Eble 1∗∗∗∗.

	Coefficient	Std. error	t-statistic	p-value
β_0	0	0	nan	nan
x_1	-0.0270	0.009	-2.987	0.003
x_2	-0.0149	0.012	-1.246	0.213
x_3	-0.0156	0.012	-1.306	0.192
x_4	0.2039	0.008	26.022	0.000
β_0	0.2320	0.005	44.642	0.000
x_1	0.0310	0.005	5.957	0.000
x_2	0.0610	0.007	8.874	0.000
x_3	-0.0185	0.007	-2.692	0.007
x_4	-0.0861	0.007	-12.518	0.000

70