High-dimensional microbiome interactions shape host fitness

Authors:

Alison Gould,[1*] Vivian Zhang,[1*] Lisa Lamberti,[2,3] Eric Jones,[4] Benjamin Obadia,[1] Alex Gavryushkin,[2,3] Jean Carlson,[4] Niko Beerenwinkel,[2,3] Will Ludington[1]

1. UC Berkeley, Dept Molecular & Cell Biology
2. ETH Zurich, Dept Biosystems Science & Engineering
3. SIB Swiss Institute of Bioinformatics, Basel, Switzerland
4. UC Santa Barbara, Complex Systems Group, Dept Physics

* authors contributed equally

Abstract:

With hundreds of species interacting with each other as well as with specific proteins and cells in the body, the gut microbiome is a complex ecosystem embedded within a complex organism. Microbiome impacts on host health can shape key aspects of fitness, such as development,[1] fecundity,[2] and lifespan,[3,4] while the host in turn can shape the microbiome.[5] However, complex interactions between microbes can make downstream effects unpredictable, such as when toxin-producing *Clostridium* species cause pathogenesis after antibiotics reduce gut diversity.[6] A pressing need exists to deconstruct the effects of gut diversity on host health, and new mathematical approaches are needed to quantify the complex patterns of interactions. Central to the microbiome-host relationship are questions of how bacterial diversity is maintained in the gut[7] and how this diversity impacts host fitness.[8] Here we show that interactions between bacteria are major determinants of host physiology and the maintenance of bacterial diversity. We performed a complete combinatorial dissection of the naturally low-diversity *Drosophila* gut microbiome using germ-free flies colonized with each possible combination of five core species of bacteria, forming a five-dimensional landscape in ecological state space. For each species combination, we measured the resulting bacterial community abundances and fly fitness traits including (i) development, (ii) reproduction, and (iii) lifespan. Notably, we found that the fly gut environment promotes bacterial diversity, which in turn accelerates development, reproduction, and aging. From these measurements we calculated the impact of bacterial interactions on fly fitness by adapting a combinatorial geometry approach,[9] to the microbiome.[10] We found that host phenotypes (e.g. lifespan) from single bacterial species are not predictive of host phenotypes in diverse communities. By contrast, higher-order interactions (involving 3, 4, and 5 species) are widely prevalent and impact both host physiology and the maintenance of bacterial diversity, as recently predicted by ecologists.[11] With regard to evolution, the impacts of bacterial interactions on gut community composition parallel the impacts on host fitness traits, providing a feedback that, propagated over time, may poise a population for emergence of co-evolving microbiome-host units.

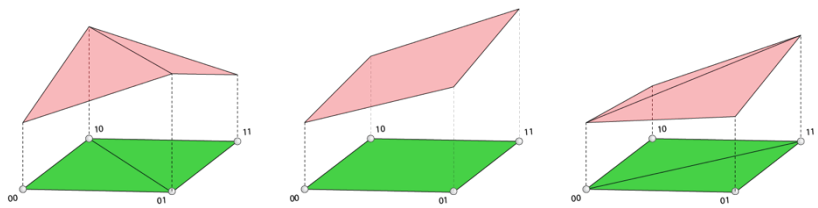
Introduction:

Gut bacteria impact host fitness in beneficial and detrimental ways, for example by improving a poor diet[12-14] or by pathogenesis.[6,15] Identifying specific bacteria responsible for these effects has been difficult, in part due to high gut diversity but also because interactions between bacteria can depend on context, both in terms of neighbor species and the gut environment.[16] For example, a bacterium may produce a specific B-vitamin in response to its neighbors.[17,18] Each microbial response may impact the host, and host feedbacks can mitigate or exacerbate changes in the microbial community.[19] However, vignette examples may be misleading, as the true complexity of a gut microbiome has never been exhaustively quantified. Thus, it remains an outstanding challenge to reverse engineer the interaction networks that characterize community microbiome-host effects relative to host interactions with a single bacterial species. Quantifying the set of all possible interactions of n species is a combinatorial problem involving 2^n distinct bacterial communities. As n approaches the diversity of the mammalian gut with hundreds of species, this problem becomes experimentally intractable.

The gut microbiome of the fruit fly *Drosophila melanogaster* is an effective combinatorial model because as few as five species of bacteria consistently inhabit the gut of wild and laboratory flies,[20-22] yielding 2^5 possible combinations of species. Here, we isolated the five core fly gut bacteria in culture, constructed germ-free flies by bleaching the embryos, and reinoculated the newly emerged adult flies via continuous feeding with defined flora using established protocols.[23,24] We made the 32 possible combinations of the five bacterial species commonly found in the fruit fly gut: *Lactobacillus plantarum* (Lp), *L. brevis* (Lb), *Acetobacter pasteurianus* (Ap), *A. tropicalis* (At), and *A. orientalis* (Ao). We quantified the microbiome composition and resultant host phenotypes to determine the relationship between gut microbe interactions and host fitness.

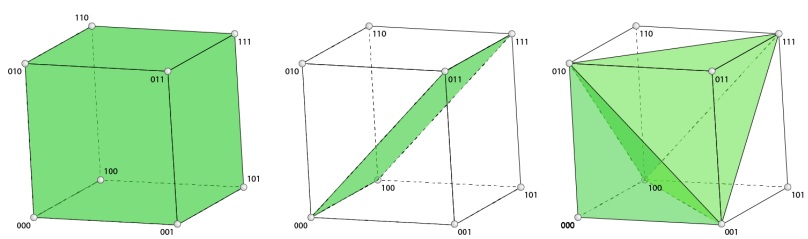
To quantify microbiome-host interactions, we employed mathematical methods originally developed for higher-order epistatic gene interactions[9] (Box; Math Supplement). More precisely, we provide a mathematical framework to deconstruct microbiome-host complexity by making a conceptual analogy between the presence of bacterial species and the occurrence of genetic mutations.[10] Our approach is combinatorial in nature with a distinguishing feature that it considers not only the 26 'standard' interaction tests among five species, but also an additional 910 more fine-grained, 'non-standard' interaction tests (Box; Math Supplement).

Box: Geometric interpretation of species interactions. Like epistatic gene interactions,[9] bacterial species interactions in the fly gut can be described geometrically. We use binary strings to denote the species composition. For two species A and B, 00 denotes the absence of both species, 01 and 10 denote the presence of only species A or B, respectively, and 11 denotes the presence of both species. To each such string one associates a phenotype measurement w, for example, fitness. Interaction between the two species can then quantitatively be described by the formula $u_{11} = w_{00} + w_{11} - w_{01} - w_{10}$ measuring non-additivity of the two-species combination. The interaction is positive if the sum of the phenotypes w_{11} and w_{00} is bigger than the sum of the phenotypes w_{01} and w_{10}, e.g. F 1 right panel.



F 1. Geometric interpretation of the interaction between two species.
The vertices 00, 10, 01, and 11 of the green rectangle represent the four microbiome compositions for two bacterial species. The heights above the points 00, 10, 01, and 11 represent the corresponding phenotypes. A single flat red plane connects the four phenotype points if there is no interaction (**center**)—if the phenotype of both species, w_{11}, can be deduced from the phenotypes of the two single species, w_{01} and w_{10}. The figures on the left and right represent cases in which the interaction is negative (**left**) and positive (**right**). In these cases, the red surfaces connecting the four phenotypes are divided into two triangular regions, indicating the curvatures of the surfaces.

This geometric approach for describing interactions generalizes to higher dimensions and yields many quantitative interaction measurements, including standard tests like u_{11} as well as non-standard tests, which use fewer than the complete number of vertices on the cube (F 2; see BPS[9] and Math Supplement for complete description). Together, these expressions can be used, for instance, to analyze how non-additivity among a subset of species combinations depends on other bystander species (Fig. S14-S17).



F 2. Geometric interpretation of the standard three-species interaction compared with two non-standard interactions.
(**Left**) Standard three-way interaction, $u_{111} = w_{111} - (w_{110} + w_{101} + w_{011}) + (w_{100} + w_{010} + w_{001}) - w_{000}$, comparing the phenotypes (not drawn) of all eight bacterial combinations, represented as vertices of the cube. (**Center**) By contrast, the vertices of the green rectangular region describe a non-standard test, which yields the interaction $g = w_{000} - w_{011} - w_{100} + w_{111}$ involving the phenotypes (not drawn) of the four bacterial combinations 000, 100, 011, and 111. (**Right**) The five vertices of the green solid bipyramid delineate a non-standard test, $m = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000}$, derived from a linear combination of other interactions. This particular bipyramid compares the phenotypes (not drawn) of three single bacterial combinations to the combination with all three species and the germ-free case.

Results & Discussion:

Microbial diversity increases in the intra-host environment

Gut microbiomes support high diversity, which facilitates complexity in the microbiome-host relationship. A classic problem in ecology is how this species diversity persists despite positive feedbacks that tend to destabilize the community.[7] Ecological models suggest that specific patterns of negative species interactions can dampen positive feedbacks to produce stable communities.[7,11] Two related patterns of stabilizing interactions have been proposed: (i) sufficient competition between pairs of bacteria,[7] and (ii) higher-order (more than pairwise) negative interactions that emerge when higher diversity is present.[11]

We first measured the stability of bacterial diversity *in vitro*. In agreement with previous reports,[25] we found that *in vitro* culture supported only low diversity (Fig 1a), with a maximum of two species coexisting in three rich media types that have a similar composition to fly food. This observation suggests that simple bacterial interactions do not support diversity (Fig. 1a). However, when we examined the same combinations in the guts of flies *in vivo*, high diversity was maintained, up to the complete five species community (Fig. 1a, Fig. S1), indicating that the host environment stabilizes diversity. Diversity significantly increased the total bacterial load ($r=0.63$, $p=0.0001$, $n=31$ bacterial combinations, Pearson correlation). Median total bacterial load ranged from 49,000 colony forming units (CFUs) per fly for Ap alone to 737,000 CFUs per fly for Lb+At+Ao. However, on a species-by-species basis, abundance stayed constant or decreased as species diversity increased (Lp: $r=-0.07$, $p=0.8$; Lb: $r=-0.37$, $p=0.2$; Ap: $r=-0.50$, $p=0.06$; At: $r=-0.59$, $p=0.02$; Ao: $r=-0.55$, $p=0.03$; Spearman correlations).

We first quantified bacterial species interactions *in vivo* by calculating pairwise correlations in species abundances as a function of the total number of species present in the gut (Fig. 1b). Correlations became more negative for individual species pairs as diversity increased ($p=0.03$, $n=10$ species pairs, Kendall's Tau and Wilcoxon signed rank, see Methods). We used two methods to calculate the directional interactions (A→B vs. B→A). First, we fit a generalized Lotka-Volterra model and found that bacterial interactions are generally positive when only two species are present (Fig. 1c), suggesting unstable diversity.[26] However, as diversity is increased, interactions become more negative (Fig. 1c). We directly calculated the same interactions at high and low diversity using Paine's classic approach[27] and detected equivalent patterns (Fig. S2), suggesting again that stabilizing interactions emerge as diversity increases, consistent with the recent theoretical prediction that higher-order interactions support diversity.[11] To specifically test whether higher-order interactions occur, we used the combinatorial approach developed previously[9] (see Box). Basing our calculations on the average total bacterial load per fly, we uncovered both significant positive and negative high-order interactions by both standard and non-standard tests (Fig. S3; Math Supplement), indicating that the biological interactions involve more than just pairs of species[11] and that they change as diversity increases.

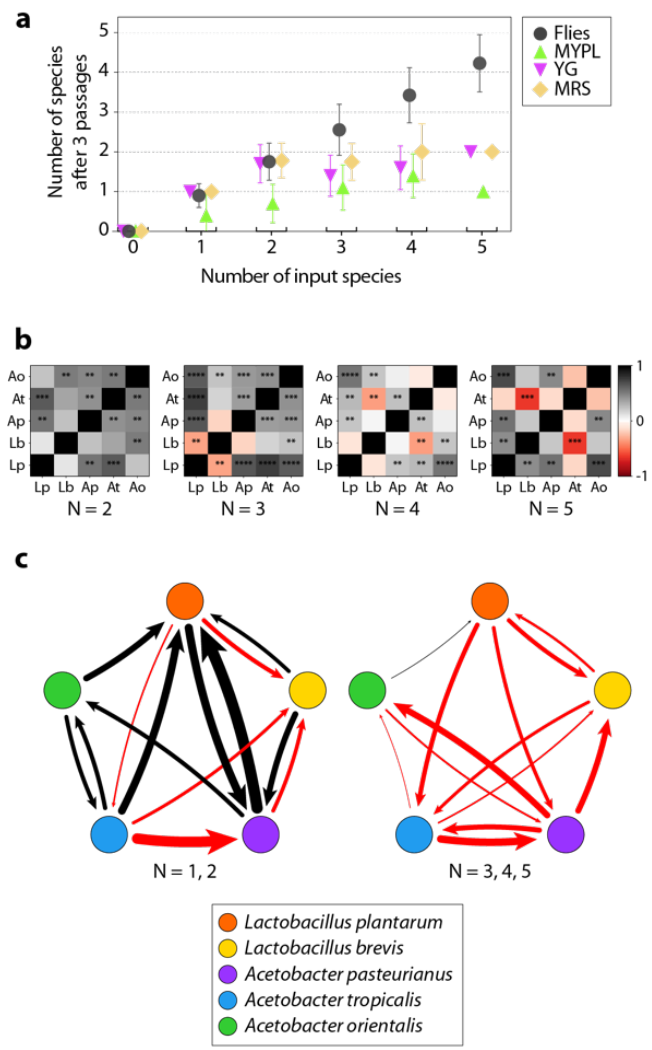


Figure 1. Microbiome interactions stabilize diversity in the fly gut. (**a**) Microbial diversity is maintained inside the fly gut to a greater degree than in liquid coculture. 24 flies per bacterial treatment were continuously fed their bacterial treatment (32 total treatments) for 10 days before individual flies were crushed, plated and enumerated as CFUs (Fig. S1). The same bacterial treatments were inoculated (3-6 replicates) into three rich media types (MRS, MYPL, and YG [see Methods]) that are similar in composition to fly food, passaged three times at 48-h intervals, and assayed by plating after the third passage. For all treatments the bacterial diversity displayed is the number of species detected versus the number inoculated. (**b**) Pairwise correlations in abundance for the 5 species of bacteria in fly guts with totals of 2, 3, 4, and 5 species present. More positive correlations are apparent at low diversity, whereas more negative correlations occur as diversity increases (p=0.03; see Methods). (**c**) Data were fit to a generalized Lotka-Volterra model to determine the directed interaction matrix (see Methods). Consistent with the correlations in **b**, more negative interactions occur in more diverse guts. Direct calculation of interaction strength[27] produced a similar pattern (Fig. S2).

Microbial diversity preferentially inhibits weak colonizers

The intra-host environment additionally drives probabilistic bacterial colonization in the fly gut,[24] whereas isotropic media lead to deterministic bacterial colonization (e.g. in well-mixed soil[25]). To test how diversity impacts the probability of colonization, we calculated the proportion of flies colonized by each species as a function of the number of species fed (Fig. S4a). Lp and Ao colonized at nearly 100% independent of background diversity, whereas Ap and At (and to a lesser extent Lb) colonized at a lower rate as diversity increases (Fig. S4a). This pattern changed when the input of new bacteria was terminated by transferring flies continuously to fresh food (Fig. S4b). Ap in particular colonized at a higher rate in more diverse guts (Fig. S4b), suggesting that bacterial interactions increase Ap populations. The diversity-dependent colonization load is consistent with niche processes of competitive exclusion:[24,28] as diversity increases, available niches are more likely to become occupied by a strong colonizer (e.g. Lp or Ao), subsequently inhibiting weaker colonizers (e.g. Lb or At).

We note that the limit of detection (~1000 CFUs) may mask low-abundance colonization. While the limit of detection did not impact our analysis of bacterial interactions (Fig. 1b,c; Math Supplement), additional CFU enumeration of individual species in individual flies indicated that most flies, which appeared uncolonized by one or more of the inoculated bacterial species, were in fact colonized at levels below the limit of detection (47 of 48 flies tested), which we have observed previously.[24]

Microbial diversity accelerates development

We next investigated the effect of gut diversity on host fitness traits. The time an egg takes to develop into an adult influences fitness because faster development reduces generation time.[29] On our replete food, both germ-free and flies colonized with all 5 bacterial species had a ~10-day development time (Fig. 2a). We hypothesized that fly development would be relatively robust to changes in bacterial composition. However, as we dissected the 5-member bacterial community, we observed faster development with increasing bacterial diversity (Fig. 2a; Fig. S5; n=32 bacterial combinations, r=–0.43, p=0.01, Spearman correlation coefficient). The overall decreased development time demonstrates that higher gut diversity can increase fly fitness.

Bacterial food conversion drives development time

We hypothesized that changes in fly development are primarily nutritional and therefore due to changes in total bacterial load rather than presence/absence of individual species.[30] Because gut bacterial abundance correlates with food bacterial abundance in flies,[31] we compared the adult bacterial load with each of the physiology phenotypes (Fig. S6). We found a significant negative correlation between bacterial load and development time (indicating faster development; r=–0.44, p<0.005, n=32 bacterial combinations, Spearman correlation), suggesting that bacterial interactions influence some aspects of host physiology through total bacterial load.

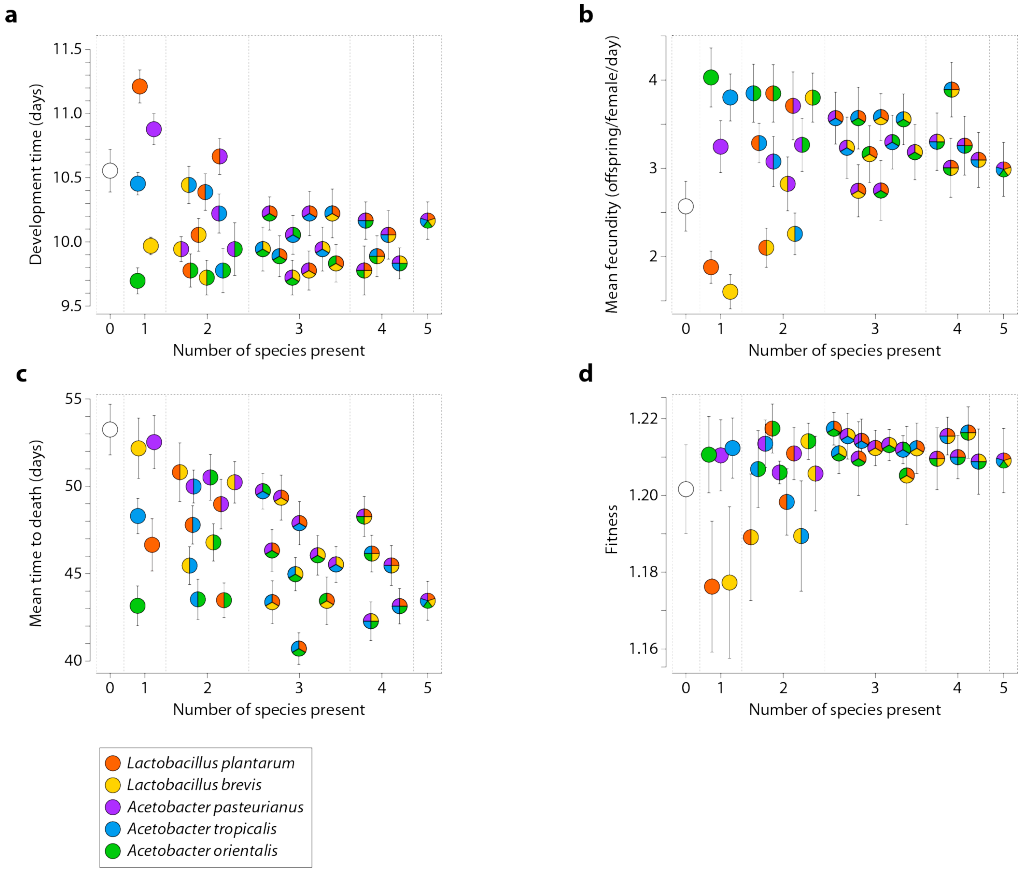


Figure 2. Microbiome diversity impacts host physiology. (**a**) The number of days to adulthood was measured as the first pupa to emerge from an individual fly vial during the lifespan experiment. The development time increases as gut diversity increases. Median n=24 per bacterial treatment (Fig. S5). (**b**) Mean fecundity per female per day was measured concomitantly with development time over the flies' lifespans. Median n=65 vials measured per bacterial treatment. Variation in fecundity decreases as gut diversity increases (Fig. S5, S9). (**c**) Lifespan decreases as gut diversity increases. Median n=100 flies per bacterial treatment (Fig. S5, S10). (**d**) Fitness calculations using a Leslie matrix populated with data from **a-c** reveals an increase in fitness and a decrease in fitness variation between different microbiomes as gut diversity increases.

The link between juvenile development time and adult bacterial composition could be influenced both by bacterial populations in the food and by parental effects. To differentiate these hypotheses, we performed two experiments. First, we quantified bacterial load in the fly food for 16 bacterial combinations (Fig. S7). Significantly fewer bacteria occurred per mg in fly food than in the gut of adult flies ($p=1.3 \times 10^{-8}$, n=16, paired sample t-test), but there was a significant correlation between total bacterial abundance in fly food and in fly guts (r=0.69, p=0.003, n=16, Spearman correlation) as previously reported.[31] However, in contradiction to our prediction that bacterial load drives development time, there was no correlation between food bacterial abundance and development time (r=0.018, p=0.95, n=16 total CFUs, Spearman correlation). On an individual species basis, Ao food abundance was correlated with development time (r=–0.95, p=0.003, n=7 bacterial combinations with Ao, Spearman correlation; Fig. 2a, S7), and fecundity (r=–0.86, p=0.02, n=7 bacterial combinations with Ao, Spearman correlation; Fig. 2b, S7) but there was no correlation of food bacterial loads with the other fly physiology parameters measured (Fig. 2c), suggesting that a single species, namely Ao, can speed up development in the larval stage and increase fecundity in adults.

We next tested whether there was a maternal effect on development time by removing the maternal bacterial association. We harvested eggs from germ-free flies and associated them with all 32 bacterial combinations. This experiment uncovered insignificant differences in development time compared with the experiment with gnotobiotic females, indicating that maternal effects do not set development time (Fig. S8). In contrast, flies inoculated with heat-killed duplicates of the experiment developed significantly more slowly (p<0.005, n=16, paired sample t-test), suggesting that active bacterial metabolism[32] is important for fly development (Fig. S8). From a fitness perspective, these results indicate that female flies gain an advantage for their offspring by associating with these diverse bacterial consortia.

Variation in fly fecundity decreases with increasing diversity

Fitness is closely tied to fecundity, the number of viable offspring produced by an individual. We counted the number of adult progeny produced per fly over its lifespan and determined that overall, average fecundity did not vary as a function of increasing biodiversity (r=0.07, p=0.7, n=32, Spearman correlation) or bacterial load (r=–0.043, p>0.05, n=32, Spearman correlation; Fig. S5, S9). However, the variance in fecundity between bacterial combinations with the same number of species decreased as species were added (Fig. 2b, S9; r=–0.91, p=0.08, n=4 diversities, Pearson correlation; p=0.001, t-test comparing deviations with 1 or 2 species vs. 3 or 4 species), indicating that diversity decreases the fecundity rate.

Germ-free flies live longer than flies colonized by bacteria

In 1927, Steinfeld[3] reported that germ-free flies live longer than their microbially colonized counterparts. This observation—that the microbiome can impact aging—has been replicated in flies and vertebrates.[4,33] Consistent with these previous studies, our germ-free flies survived longer than flies colonized with all 5 species of bacteria (mean lifespan ± standard error of the mean, 53.5 ± 1.5 germ-free vs. 43.5 ± 1.1 for 5-species gnotobiotics). Examining both male and female fly survival as a function of increasing gut diversity, we found a decrease in survival over many bacterial associations (Fig. 2c, S5, S10; r=–0.54, p=0.002, n=32, Spearman correlation) and as a function of bacterial load (r=–0.48, p=0.006, n=32, Spearman correlation). The trend of decreased lifespan corresponded to an increase in fecundity for female flies (average daily fecundity vs. lifespan: r=–0.50, p=0.003, n=32, Spearman correlation; Fig.

S11a) and is not explained by differences in fly activity (Fig. S12). Such life history tradeoffs are well-documented in the literature and are believed to constitute a differential allocation of resources between long-term body maintenance and reproduction.[29,34] As tired parents, we hypothesized that the tradeoff may be inherent and that high reproduction and long life are mutually exclusive.

To examine this hypothesis, we used antibiotics to remove the microbiome of high-fecundity female flies and measured the resulting change in lifespan. We first allowed female flies with high-fecundity microbiomes to reproduce for 21 days (to a level greater than the total lifetime fecundity of germ-free flies; Fig. S9), and we subsequently eliminated the microbiome using antibiotics. In general, the midlife elimination of gut flora lengthened the female fly lifespan by roughly 15% compared to flies continuously fed live bacteria (Fig. S11b; $p=0.007$, $n=7$, paired sample t-test), without decreasing total fecundity (Fig. S11b; $p=0.2$, $n=7$, paired sample t-test). This result demonstrates that the life history tradeoff is not necessarily fixed and suggests that flies respond with rapid and high fecundity to a perceived risk of death due to microbes. However, two specific bacterial combinations yielded no increase in lifespan when their host was given antibiotics: Ao and Lp+Lb+Ao, suggesting a fixed life history tradeoff in only these two combinations. Very similar microbiome compositions significantly expanded the lifespan of the host (Fig. S11b; e.g. Lp+Ao and Lp+At+Ao), demonstrating the complexity of the microbiome-host relationship. When we performed the reverse experiment and associated midlife germ-free flies with high-fecundity bacteria, there was a significant increase in fecundity and a significant decrease in lifespan (Fig. S11b; $p=0.007$, $n=4$, paired sample t-test for fecundity; $p=0.006$, $n=4$, paired sample t-test for lifespan), consistent with our initial experiments (Fig. 2), and indicating that the life history tradeoff is dynamic throughout life.

Microbial diversity increases host fitness

We wondered how the lifespan-fecundity tradeoff impacted organismal fitness, which is defined as the rate of population growth. To address this question, we combined our data for development, fecundity, and lifespan in the Leslie matrix,[35] a classical model of discrete population growth, to calculate organismal fitness under each bacterial association. Overall, fitness was relatively consistent across distinct bacterial associations (Fig. 2d). However, the variation in fitness between low-diversity microbiomes was higher than that between high-diversity microbiomes (Fig. 2d), indicating that the life history tradeoff is more effective at high diversity.

Microbial interactions impact host fitness

Previous studies of complex microbiomes have suggested that interactions among species impact the host. For example, the induction of *Clostridium difficile* pathogenicity is repressed by the presence of other species.[6] Alternatively, interactions may be mediated by the host: gut colonization by a specific *Escherichia coli* strain can induce tolerance to *Burkholderia* infection in the lungs mediated by IGF-1.[15] Our data suggest complex interactions between bacteria with impacts on host physiology. For instance, flies mono-colonized with Lp, At, or Ao have mean survival times of 47, 48, and 43 days respectively (Fig. 2c). No significant changes occurred with pairs of species, but fly survival dropped to 41 days when all three bacteria were inoculated together (Fig. 2c). To determine whether the interaction is of higher order or whether pairwise interactions explain the three-species phenotype, we applied a combinatorial approach (Box),[9] originally developed as a framework to calculate epistatic interactions between genes in high dimensions. Here we generalize the method to characterize high-dimensional interactions in the microbiome.

We used this framework to calculate the interactions between gut bacterial species and their impacts on host physiology (Math Supplement). Notably, the effects on host physiology could not be attributed to simple changes in individual species abundances, aside from the development time of 7 samples containing Ao, as noted previously. Instead, we detected significant effects of species interactions on host physiology and total bacterial load (Fig. 3a-d, S3), with wide prevalence of non-additive pairwise and high-order interactions. Interestingly, development-rate interactions tended to be negative (Fig. 3b). Negative epistasis in genetics suggests that two loci are in the same pathway (they are redundant). By analogy, negative microbiome epistasis in development time suggests that redundant mechanisms, such as nutrition,[30] facilitate microbiome effects on the host. In contrast, time to death interactions tended to be positive (Fig. 3d), suggesting that bacterial interactions synergistically modulate various pathways. CFUs and fecundity (Fig. 3e-h) both had significant positive and negative interactions for specific bacterial combinations (Fig. 3a,c), suggesting both synergy and redundancy. The magnitudes of the interactions, when normalized to the number of bacterial species present, were often as large as the effects of individual species introductions (Math Supplement), indicating that the species interactions are as important as the presence of the species themselves (Math Supplement). Thus, changes in gut microbiome composition impact host physiology in complex, non-additive ways.

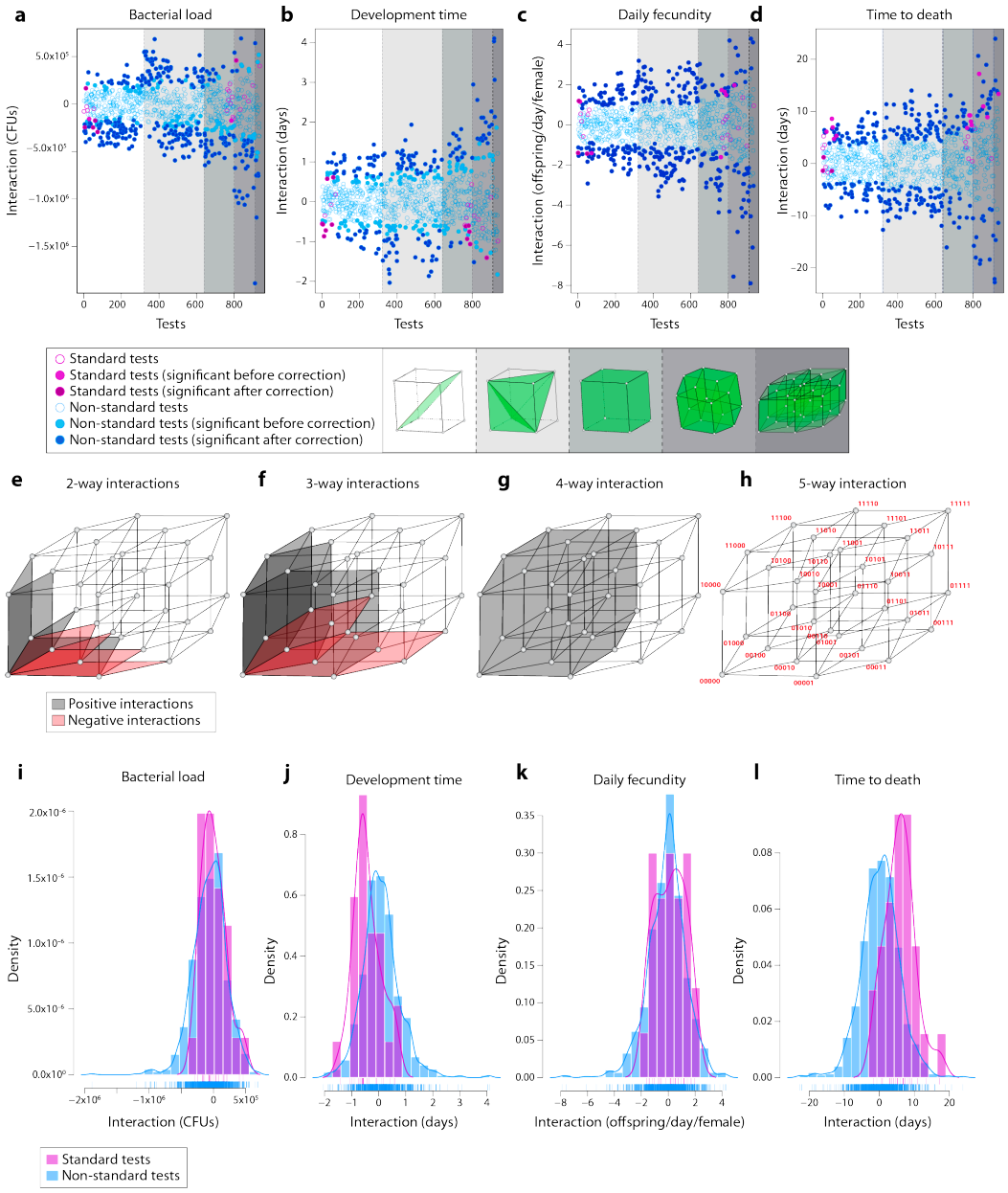


Figure 3. Microbial interactions drive the microbiome-host relationship. (**a-d**) Interactions become stronger as diversity increases. For (**a**) bacterial load, (**b**) development, (**c**) fecundity, and (**d**) time to death, we calculated interactions for standard tests (pink dots) and non-standard tests (blue dots) for phenotype data (Fig. 1,2). See legend for significance at $p<0.05$. The Benjamini-Hochberg correction was applied for multiple comparison (dark fill color); open circles: non-significant interactions. See Fig. S13 for just the standard tests with species identities. (**e-g**) Standard interactions for daily fecundity explained geometrically. The depicted regions inside the projection of the 5-cube correspond to specific interactions. Positive (grey) and negative (red) statistically significant interactions after Benjamini-Hochberg correction ($p<0.05$; compare with panel **3c** and Fig. S13). (**h**) The vertices the 5-cube correspond to bacterial combinations. For daily fecundity, the single standard 5-way interaction is not statistically significant and therefore not shaded. (**i-l**) Comparisons between the distributions of standard and non-standard interactions reveal significant shifts, indicating context-dependence of the interactions due to bystander species.

Complexity in the microbiome could arise from context-dependent effects, where two species interact differently depending on the presence of a third species. We detected a large dependence on context by calculating non-standard tests (see Box; Math Supplement), which are examples of marginal and conditional epistasis in genetics.[9] Comparing these non-standard tests to the standard ones provides a quantitative measure of the context-dependence of species interactions. We found significant context-dependence (Fig. 3i-l; Fig. S14-S17; Math Supplement), indicating that low-dimensional interactions are significantly changed by additional species. Thus, the complexity of the microbiome is high, and the impacts of individual species vary greatly according to their co-inhabitants of the gut. However, in some specific cases where a standard test indicated a non-additive interaction, certain related, non-standard tests for the same species composition indicated an additive interaction (Fig. S14-S17 compare the different tests in Box F 2). For instance, the standard test for lifespan interactions indicates a significant three-way standard interaction between Lp, Lb, and Ao. However, the non-standard test equation $i = w_{000} - w_{010} - w_{101} + w_{111}$ (Math Supplement) indicates additivity: by considering Lp and Ao together as a single locus, the Lp+Lb+Ao lifespan is predictable as the sum of the lifespans of Lb-colonized flies and (Lb+Ao)-colonized flies relative to germ-free flies. Therefore, our methods suggest that lower-dimensional routes may exist to additively traverse the high-dimensional microbiome landscape (Fig. 3a-d; Math Supplement Fig. S5-8) if we consider certain subsets of species together as single entities. While we discovered no consistent patterns enabling the inference of these groups *a priori* (Fig. S14-S17), the presence of additive low-dimensional routes suggests a path to predictability. Thus, our results suggest we should consider bacterial species interactions rather than the presence/absence of individual species in determining the structure of the microbiome. A major challenge for the future will be to discover the rules and mechanisms by which these low-dimensional interactions scale under increasing diversity.

From an evolutionary perspective, many previous reports documented host-specialized microbiomes with a phylogenetic signature and correlated fitness consequences of this co-adaptation.[36-39] We observed no relationship between host fecundity and bacterial abundances (Fig. S6a), suggesting that the microbiome-host unit may not necessarily select more fit individuals nor favor cohesive group selection.[10,40] Rather, we see a significant relationship between the bacterial interactions generated by bacterial abundance and host fecundity phenotypes (Fig. S6b-d), indicating that microbiome interactions shape community structure and that this structure similarly shapes host fitness. The large number of interactions in the community landscape and correlated interactions in host fitness traits suggest that differences in microbiome diversity between hosts may be stable[24,41] and have phenotypic consequences (Fig. 2), which increases standing phenotypic diversity in a population of hosts, priming hosts and their microbiomes for divergent selection but not causing it. We propose correlated interactions between microbiome ecology and host fitness as a mechanism that poises a microbiome-host system for evolutionary divergence and the formation of specialized gut communities.

Materials and Methods

Fly stock maintenance: *Wolbachia*-free *Drosophila melanogaster* Canton-S flies were reared on cornmeal-based medium (6.67% cornmeal, 2.7% active dry yeast, 1.6% sucrose, 0.75% sodium tartrate, 0.73% ethanol, 0.68% agar, 0.46% propionic acid, 0.09% methylparaben, 0.06% calcium chloride, and 0.01% molasses). Fly stocks were maintained at 25 ˚C, 60% humidity, and 12:12 h light:dark cycles. Fly stocks were tested for the presence of known RNA viruses by RT-PCR and were virus-free.[24] Germ-free fly stocks were kept in sterile conditions over multiple generations to reduce heterogeneity due to parental nutrition derived from microbiome variability.

Germ-free fly preparation: *Wolbachia*-free and virus-free *Drosophila melanogaster* Canton-S flies reared on a cornmeal-based medium were transferred to embryo collection cages and allowed to acclimate in the cage for at least one day before egg collection. On the morning of egg collection, a yeast paste was added on a grape juice agar plate. Flies were left to lay eggs on this grape juice agar plate for 5-6 h. Eggs were then collected into a 0.4 μm cell strainer. In a biosafety cabinet, fly eggs were rinsed twice in 10% bleach (0.6% sodium hypochlorite) for 2.5 min each, once in 70% ethanol for 30 s, and three times in sterile dH_2O for 10 s each. Approximately 50 eggs were transferred to sterile fly medium (10% glucose, 5% active dry yeast, 1.2% agar, 0.42% propionic acid) with a sterile cotton swab.

Bacteria strains: *Lactobacillus plantarum*, *L. brevis*, *Acetobacter pasteurianus*, *A. tropicalis*, and *A. orientalis* bacteria were isolated from *D. melanogaster* lab flies. Bacteria were grown overnight in MRS medium in a shaker set at 30 ˚C. The bacteria were resuspended at a concentration of 10^8 cells/mL in sterile phosphate-buffered saline (PBS) for fly gnotobiotic preparations[23] so that constant numbers of CFUs were inoculated per fly vial. The 32 combinations of the 5 bacterial strains were mixed using a Beckman Coulter Biomek NXP workstation to standardize the inoculum. Vials were swabbed to ensure correct bacterial species were present without contaminants.

Adult gnotobiotic fly preparation: Germ-free flies aged 5-7 days were sorted into 10% glucose, 5% active dry yeast medium inoculated with a defined mixture of bacteria. Each vial contained a total of $5x10^6$ CFUs (50 μL of 10^8 bacteria/mL). Ten female and ten male flies were transferred into each vial. Gnotobiotic flies were flipped into freshly inoculated medium every 3 days.

Concurrent lifespan assay, fecundity (pupae counts), fly development (Fig 2a-c): We measured all host fitness phenotypes concurrently in mixed sex populations in order to mimic more natural conditions. To measure the lifespan of flies on each combination of bacteria, we recorded the number of flies living and number of flies dead daily until the entire population was dead. Dead flies were removed when the vials were flipped. Average daily female fecundity was assessed by counting the total number of pupae in each vial after the adults were flipped to a fresh vial. Due to the variable development times involved, vials were monitored daily for 14 days after removing the adults. To determine development times, we counted the day when the first adult emerged from each vial. We chose this metric because adults were housed in the same vial for 3 days and therefore the start of development was not synchronized.

Development Assays (Fig. S8): In the experiments presented in Fig. S8, development times were assessed for each egg introduced to the vial. Eggs were first dechorionated and sterilized as described in *Germ-Free Fly Preparation* above. Eggs were then suspended in 1x PBS with 0.1% TritonX to allow pipetting of the eggs. Roughly 30 eggs (and always >20 eggs) were pipetted into the recipient vial. Timing of pupation and eclosion in vials in which flies had previously developed were assayed at 1-day

intervals for non-heat-killed (blue dots) and heat-killed (red dots) preparations. For the germ-free eggs inoculated with fresh bacteria (Fig. S8 black points), development timing was assessed at ~3-h intervals.

Bacterial load counts (Fig. 1 & S1): To assess the number of bacterial CFUs per fly, flies were washed in 70% ethanol for 5 s, rinsed in ddH_2O for 5 s, and put into the well of a 96-well plate containing 100 μL PBS and 80 μL 0.5 mm glass beads (Biospec). Plates were heat sealed with aluminum sealing film (E&K Scientific), then bead beaten for 60 s at maximum speed in a MiniBeadBeater-8 (Biospec) converted to hold a 96-well plate using a custom-built attachment. Plates were then pinned with a 96-pin replicator (Boekel) in three technical replicates per fly onto selective media that allowed us to visually distinguish each bacterial species. Selective media were: MRS (Difco) with X-gal, which grows only Lp (yellowish-white colonies) and Lb (blue colonies); MYPL with 5 mg/L tetracycline, which grows only Ap (rounder, thicker, browner colonies) and Ao (flat colonies with ruffled borders); and MYPL with 50 mg/L gentamycin, which grows only At and Ao. Plates were grown at 30˚C. A standard curve was constructed for each strain to calculate CFUs from the observed bacterial counts (Fig. S18). MYPL media is 1% w/v mannitol, 1% w/v yeast extract, 1% v/v lactic acid, 0.5% w/v bacto-peptone, 0.1% tween 80, pH 6.5. YG media is 5% w/v glucose, 2.5% w/v baker's yeast, 0.42% v/v proprionic acid.

Fly activity assay (Fig. S12): Gnotobiotic flies were prepared as previously described. Ten females and ten male flies were sorted into each vial. Each vial was flipped every 3 days into medium inoculated with the required bacterial mixture. After the 9th day (the third flip), gnotobiotic flies were flipped into a vial containing sterile medium (10% glucose, 5% yeast, 1.2% agar, and 0.42% propionic acid). These vials were placed into the LAM25 (Locomotor Activity Monitor; Trikinetics) kept at 25 ˚C, 60% humidity, and 12:12 h light:dark cycles and monitored for 7 days.

Fitness calculations (Fig 2d): We estimated fitness for each bacterial treatment using a Leslie matrix (1,000x per treatment). For each replicate Leslie matrix, we randomly sampled from the experimental replicates of development time, fecundity and lifespan. Female fecundity was counted as zero until the day of adult emergence. Thereafter, the fecundity was filled from the data by random sampling of the 5 replicate vials for each time point.
The diagonal was filled with 1s corresponding to the development time. After development, the adult survival data were used by randomly sampling the 5 adult survival-probability replicates for each day. The remaining values in the matrix were zeros. We then calculated the dominant eigenvalue of the matrix for each of the 1,000 replicate samplings, yielding a range of fitness estimates. This fitness value, λ, corresponds to the daily fold expansion of the population, $N_{t+1} = \lambda N_t$, under ideal conditions.

Statistical analyses: All statistics were calculated using R (v.3.3.3)[42] unless otherwise noted. Survival data average curves were calculated as the cumulative proportion of the population that died over time. A 3-parameter Gompertz function with an upper limit of 1 was selected using the 'drc' package[43] in R, $N(t) = N_0 e^{-e^{-a(t-b)}}$, where $N(t)$ is the proportion of the population surviving as a function of time (Fig. S10). Model selection using the Akaike information criterion was applied to pick the best function. The same approach was applied to fit the fecundity data (Fig. S9), resulting in a 3-parameter Gompertz, $f(t) = f_0 e^{-e^{-a(t-b)}}$.

The Shapiro-Wilk test was used to determine if data were consistent with a normal distribution. For correlation tests, Pearson correlations were used where the data were consistent with a normal distribution. Spearman correlations were used where the data were inconsistent with a normal distribution.

References

1 Smith, M. I. *et al.* Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science* **339**, 548-554, doi:10.1126/Science.1229000 (2013).

2 Leitão-Gonçalves, R. *et al.* Commensal bacteria and essential amino acids control food choice behavior and reproduction. *PLoS Biology* **15**, e2000862, doi:10.1371/journal.pbio.2000862.s013 (2017).

3 Steinfeld, H. M. Length of life of Drosophila melanogaster under aseptic conditions. *Dissertation thesis*, 1-23 (1927).

4 Smith, P. *et al.* Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. *eLife* **6**, doi:10.7554/eLife.27014 (2017).

5 Martinson, V. G., Douglas, A. E. & Jaenike, J. Community structure of the gut microbiota in sympatric species of wild Drosophila. *Ecology Letters* **26**, 32, doi:10.1111/ele.12761 (2017).

6 Ng, K. M. *et al.* Microbiota-liberated host sugars facilitate post-antibiotic expansion of enteric pathogens. *Nature* **502**, 96-99, doi:10.1038/nature12503 (2014).

7 Coyte, K. Z., Schluter, J. & Foster, K. R. The ecology of the microbiome: Networks, competition, and stability. *Science (New York, NY)* **350**, 663-666, doi:10.1126/science.aad2602 (2015).

8 Sze, M. A. & Schloss, P. D. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. *mBio* **7**, doi:10.1128/mBio.01018-16 (2016).

9 Beerenwinkel, N., Pachter, L. & Sturmfels, B. Epistasis and shapes of fitness landscapes. *Statistica Sinica* (2007).

10 Bordenstein, S. R. & Theis, K. R. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLoS Biology* **13**, e1002226, doi:10.1371/journal.pbio.1002226 (2015).

11 Grilli, J., Barabás, G., Michalska-Smith, M. J. & Allesina, S. Higher-order interactions stabilize dynamics in competitive network models. *Nature*, 1-5, doi:10.1038/nature23273 (2017).

12 Ceja-Navarro, J. A. *et al.* Gut microbiota mediate caffeine detoxification in the primary insect pest of coffee. *Nature Communications* **6**, 1-9, doi:10.1038/ncomms8618 (2015).

13 Brune, A. Symbiotic digestion of lignocellulose in termite guts. *Nature reviews Microbiology* **12**, 168-180, doi:10.1038/nrmicro3182 (2014).

14 Dill-McFarland, K. A., Breaker, J. D. & Suen, G. Microbial succession in the gastrointestinal tract of dairy cows from 2 weeks to first lactation. *Scientific Reports*, 1-12, doi:10.1038/srep40864 (2017).

15 Schieber, A. M. P. *et al.* Disease tolerance mediated by microbiome E. coli involves inflammasome and IGF-1 signaling. *Science (New York, NY)* **350**, 558-563, doi:10.1126/science.aac6468 (2015).

16 Cho, I. & Blaser, M. J. Applications of Next-Generation Sequencing: The human microbiome: at the interface of health and disease. *Nature reviews Genetics* **13**, 260-270, doi:10.1038/nrg3182 (2012).

17 Degnan, P. H., Taga, M. E. & Goodman, A. L. Vitamin B12 as a Modulator of Gut Microbial Ecology. *Cell metabolism* **20**, 769-778, doi:10.1016/j.cmet.2014.10.002 (2014).

18 Seth, E. C. & Taga, M. E. Nutrient cross-feeding in the microbial world. *Frontiers in microbiology* **5**, 350, doi:10.3389/fmicb.2014.00350 (2014).

19 Rolig, A. S., Parthasarathy, R., Burns, A. R., Bohannan, B. J. M. & Guillemin, K. Individual Members of the Microbiota Disproportionately Modulate Host Innate Immune Responses. *CHOM* **18**, 613-620, doi:10.1016/j.chom.2015.10.009 (2015).

20 Ryu, J. H. *et al.* Innate immune homeostasis by the homeobox gene caudal and commensal-gut mutualism in Drosophila. *Science* **319**, 777-782, doi:10.1126/science.1149357 (2008).

21 Wong, C. N. A., Ng, P. & Douglas, A. E. Low-diversity bacterial community in the gut of the fruitfly Drosophila melanogaster. *Environmental Microbiology* **13**, 1889-1900, doi:10.1111/j.1462-2920.2011.02511.x (2011).

22 Chandler, J. A., Lang, J. M., Bhatnagar, S., Eisen, J. A. & Kopp, A. Bacterial communities of diverse Drosophila species: ecological context of a host-microbe model system. *PLoS genetics* **7**, e1002272, doi:10.1371/journal.pgen.1002272 (2011).

23 Newell, P. D. & Douglas, A. E. Interspecies Interactions Determine the Impact of the Gut Microbiota on Nutrient Allocation in Drosophila melanogaster. *Applied and Environmental Microbiology* **80**, 788-796, doi:10.1128/AEM.02742-13 (2013).

24 Obadia, B. *et al.* Probabilistic Invasion Underlies Natural Gut Microbiome Stability. *Current biology : CB* **27**, 1999-2006.e1998, doi:10.1016/j.cub.2017.05.034 (2017).

25 Friedman, J., Higgins, L. M. & Gore, J. Community structure follows simple assembly rules in microbial microcosms. *Nature Publishing Group* **1**, 1-7, doi:10.1038/s41559-017-0109 (2017).

26 May, R. M. Will a large complex system be stable? *Nature* **238**, 413-414 (1972).

27 Paine, R. T. Food-web analysis through field measurement of per capita interaction strength. *Nature*, 1-3 (1992).

28 Vega, N. M. & Gore, J. Stochastic assembly produces heterogeneous communities in the Caenorhabditis elegans intestine. *PLoS Biology* **15**, e2000633, doi:10.1371/journal.pbio.2000633.s012 (2017).

29 Travers, L. M., Garcia-Gonzalez, F. & Simmons, L. W. Live fast die young life history in females: evolutionary trade-off between early life mating and lifespan in female Drosophila melanogaster. *Scientific Reports*, 1-7, doi:10.1038/srep15469 (2015).

30 Yamada, R., Deshpande, S. A., Bruce, K. D., Mak, E. M. & Ja, W. W. Microbes Promote Amino Acid Harvest to Rescue Undernutrition in Drosophila. *Cell reports* **10**, 865-872, doi:10.1016/j.celrep.2015.01.018 (2015).

31 Blum, J. E., Fischer, C. N., Miles, J. & Handelsman, J. Frequent Replenishment Sustains the Beneficial Microbiome of Drosophila melanogaster. *mBio* **4**, e00860-00813-e00860-00813, doi:10.1128/mBio.00860-13 (2013).

32 Fischer, C. *et al.* Metabolite exchange between microbiome members produces compounds that influence Drosophila behavior. *eLife* **6**, doi:10.7554/eLife.18855 (2017).

33 Clark, R. I. *et al.* Distinct Shifts in Microbiota Composition during Drosophila Aging Impair Intestinal Function and Drive Mortality. *CellReports*, 1-13, doi:10.1016/j.celrep.2015.08.004 (2015).

34 Chippindale, A. K., Leroi, A. M., Kim, S. B. & Rose, M. R. Phenotypic plasticity and selection in Drosophila life-history evolution. I. Nutrition and the cost of reproduction *Journal of evolutionary biology*, 1-23 (1993).

35 Leslie, P. H. On the use of matrices in certain population mathematics. *Biometrika* **33**, 183-212 (1945).

36 Ramalho, M. O., Bueno, O. C. & Moreau, C. S. Species-specific signatures of the microbiome from Camponotus and Colobopsis ants across developmental stages. *PLOS ONE* **12**, e0187461, doi:10.1371/journal.pone.0187461.s010 (2017).

37 Kwong, W. K. *et al.* Dynamic microbiome evolution in social bees. *Science Advances* **3**, e1600513, doi:10.1126/sciadv.1600513 (2017).

38 Brucker, R. M. & Bordenstein, S. R. The Hologenomic Basis of Speciation: Gut Bacteria Cause Hybrid Lethality in the Genus Nasonia. *Science (New York, NY)* **341**, 667-669, doi:10.1126/science.1240659 (2013).

39 Moeller, A. H. *et al.* Cospeciation of gut microbiota with hominids. *Science (New York, NY)* **353**, 380-382, doi:10.1126/science.aaf3951 (2016).

40 Moran, N. A. & Sloan, D. B. The Hologenome Concept: Helpful or Hollow? *PLoS Biology* **13**, e1002311, doi:10.1371/journal.pbio.1002311.g001 (2015).

41 Faith, J. J. *et al.* The long-term stability of the human gut microbiota. *Science (New York, NY)* **341**, 1237439, doi:10.1126/science.1237439 (2013).

42 R: A language and environment for statistical computing. (R Foundation for Statistical Computing, Vienna, Austria, 2017).

43 Ritz, C., Baty, F., Streibig, J. C. & Gerhard, D. Dose-Response Analysis Using R. *PLoS One* **10**, e0146021 (2015).

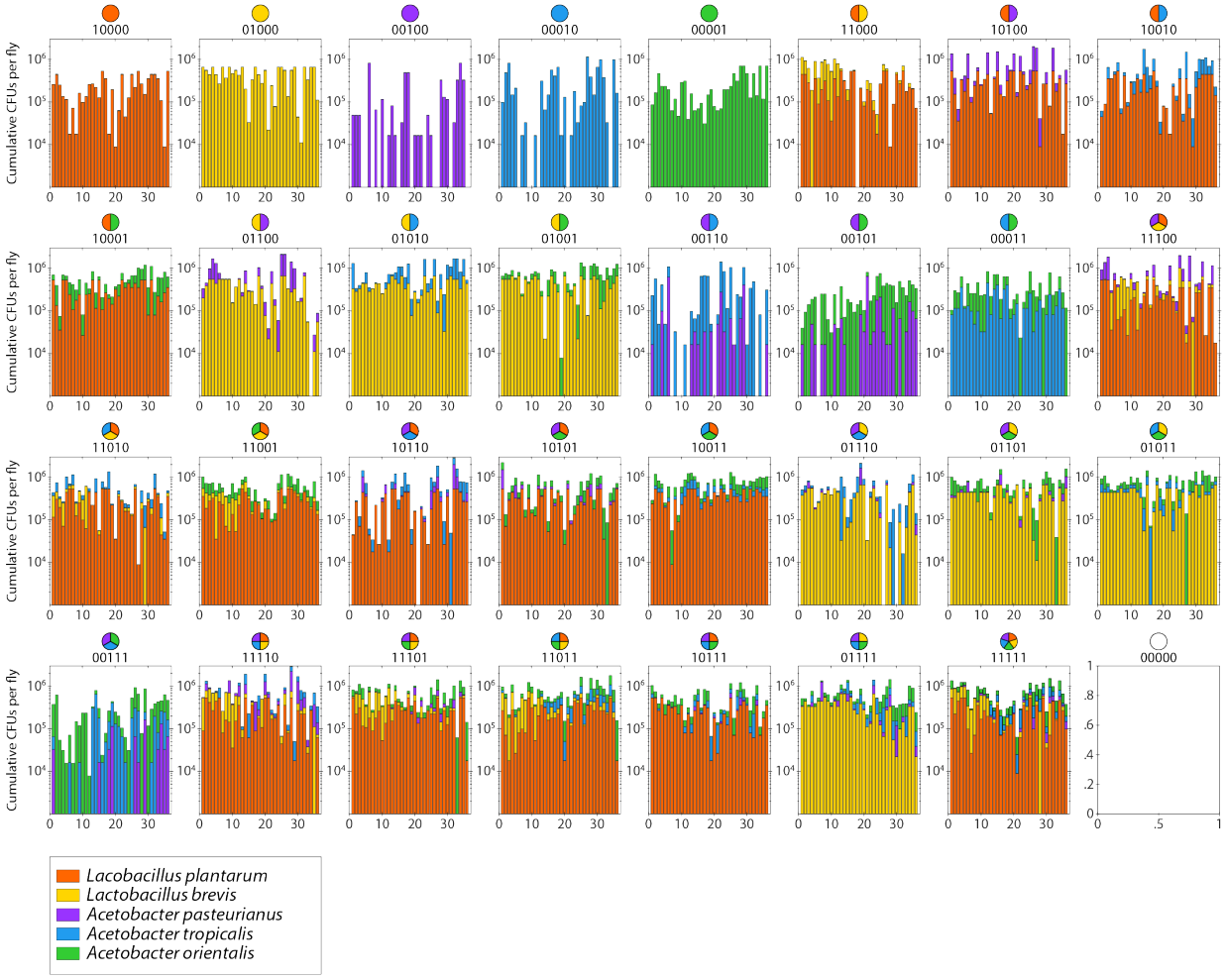


Figure S1. Representative raw bacterial abundance counts (CFUs) for each fly with each bacterial combination. X-axes indicate individual flies. Y-axes are CFUs on a $\log_{10}$ scale. For each bacterial combination, the first 12 flies are from a treatment where flies were fed defined bacteria continuously for 10 days and then transferred daily to fresh food for 5 days. The last 24 flies for each bacterial combination are for the treatment where fly development, fecundity, and time to death were measured (Fig. 2). Note that there are subtle differences between the first 12 and last 24 flies for some bacterial combinations (*e.g.*, Ap+At+Ao). These differences are quantified in Fig. S3a versus S3b, indicating that continuous bacterial ingestion represses the gut abundances of specific species, such as Ap.

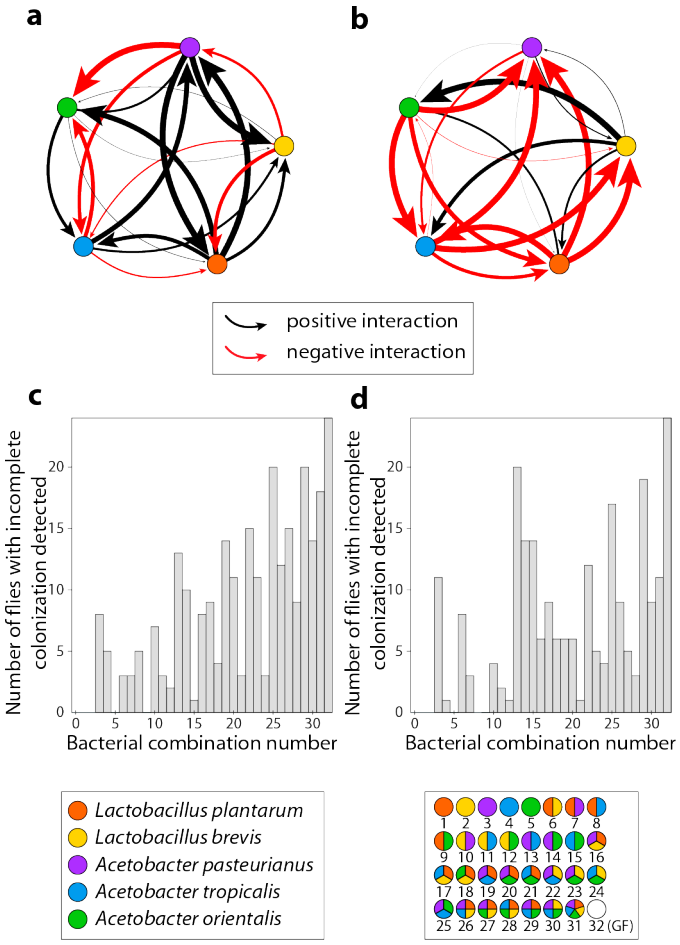


Figure S2. Pairwise bacterial interactions in the fly gut transition from positive to negative as diversity increases. Direct interaction strength was calculated per Paine[26] (Math Supplement). Mean microbial abundances across replicates were used. (**a**) Interactions calculated by comparing treatments with two species to treatments with one species. (**b**) Interactions calculated by comparing the treatment with five species to the treatments with four species. (**c**) The number of flies where not all inoculated bacterial species were detected (1,000 CFUs limit of detection) increases as the total species diversity increases when flies were continuously fed bacteria. (**d**) The number of flies where not all inoculated bacterial species were detected (1,000 CFUs limit of detection) increases as the total species diversity increases when flies were daily transitioned to germ-free food for 5 days following an initial 10-day continuous inoculation period.

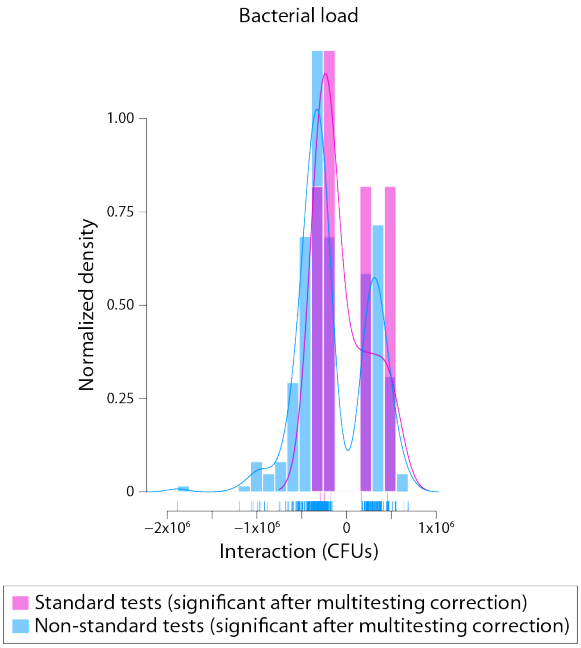


Figure S3. Microbial interactions drive the per-fly bacterial load. Standard interactions (pink) and non-standard interactions (blue) both show significant densities of non-zero scores. Mean total bacterial abundance from each bacterial combination (n=24 flies continuously fed bacteria; same raw data used in Fig. 1b) was calculated and converted to a z-score. Interactions were calculated for all 26 standard tests and 910 non-standard tests (Math Supplement). Histograms of interaction scores (large vertical bars) were converted to a probability density function (smooth curve) using a Gaussian kernel. Actual scores are shown below the histograms (short vertical lines).

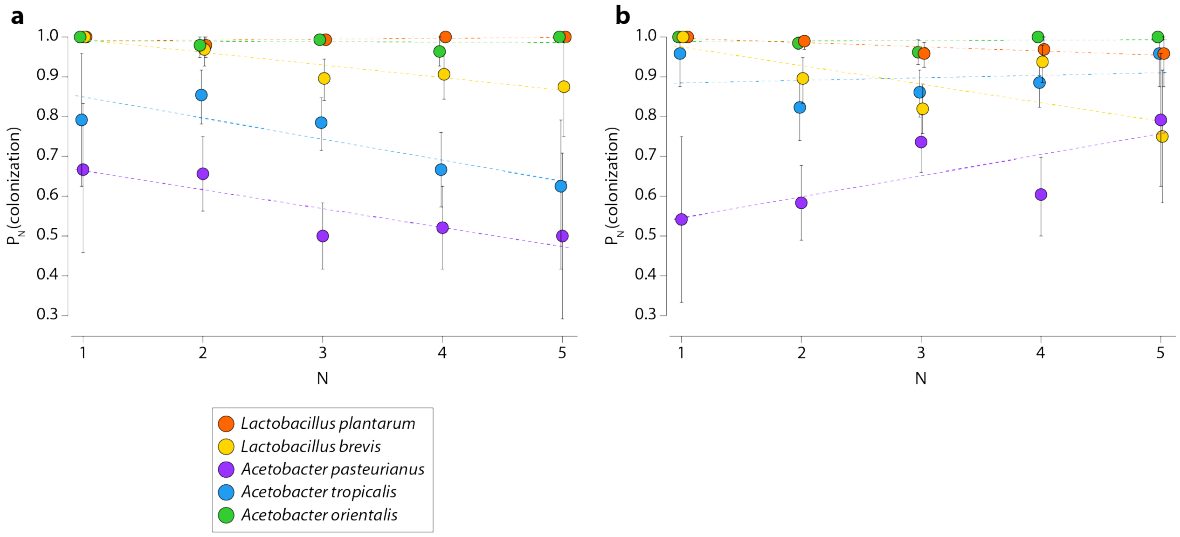


Figure S4. Higher gut diversity decreases the proportion flies colonized by weak colonizers. (**a**) Flies continuously fed food colonized with a high bacterial load are less likely to contain certain species (1,000 CFUs limit of detection). Ap, At, and to a lesser extent Lb were less likely to be detected as total diversity increased. Lp and Ao underwent complete colonization across a gradient of species diversity. (**b**) This change in colonization likelihood shifted when we flipped flies daily to fresh food for 5 days after an initial 10-day colonization by continuous inoculation. Notably, the Ap colonization likelihood increased with increasing diversity when we removed the bacterial supply, suggesting that Ap is a better competitor than it is a colonizer. n=24 individual flies per bacterial combination. Error bars are standard deviations.

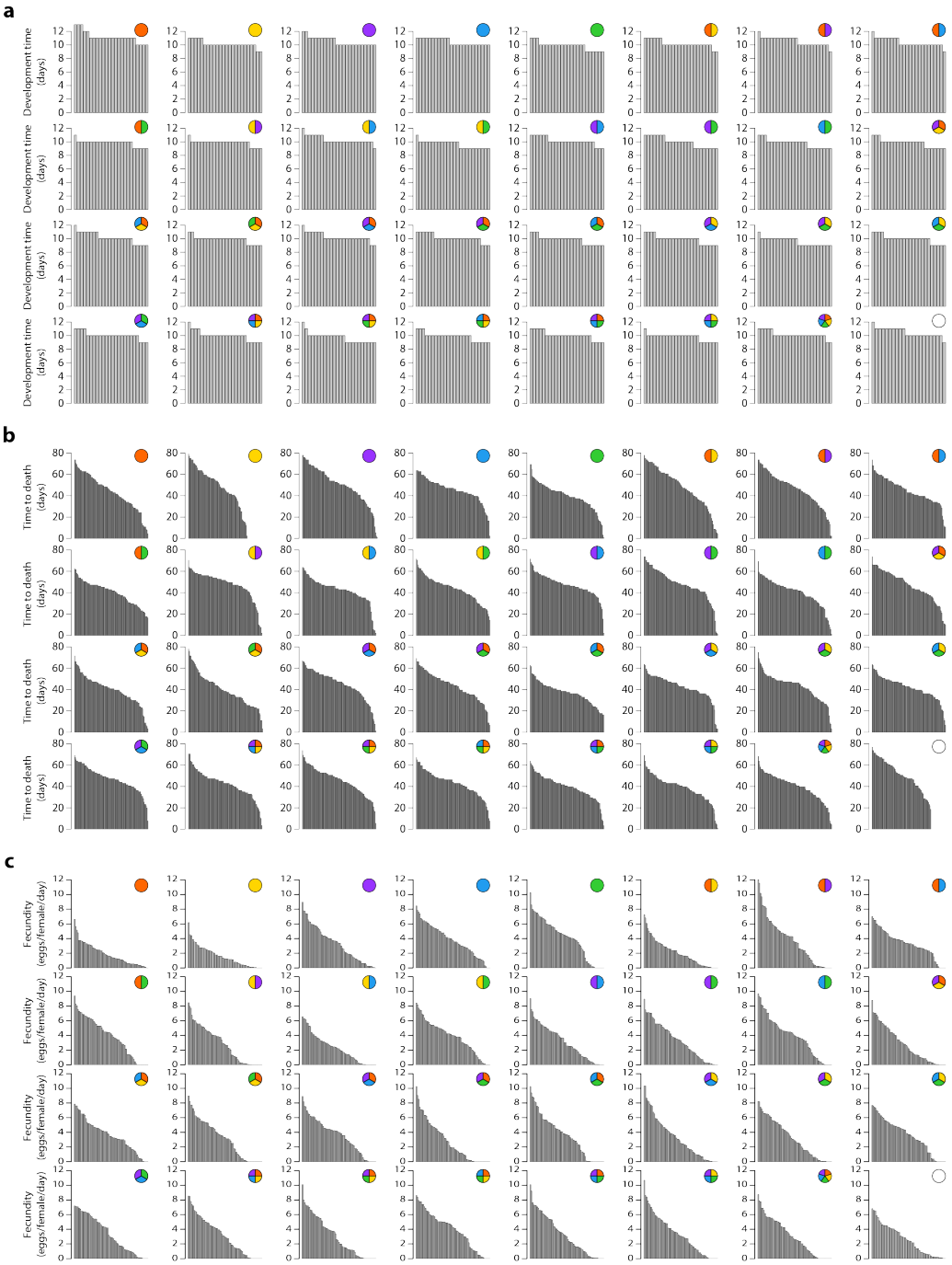


Figure S5. Raw data from development rate, fecundity, and time to death. (**a**) Raw data for development time by microbial treatment. Each bar within a treatment is the fastest-developing fly within a vial. (**b**) Raw data for time to death by microbial treatment. Each bar represents the lifespan of an individual fly. Male and female flies are aggregated because no statistically significant difference was detected between male and female lifespans in these mixed-sex experiments. (**c**) Raw data for fecundity per day per female by treatment. Each bar represents the total fecundity measured from a single fly vial normalized to the number of adult female flies.

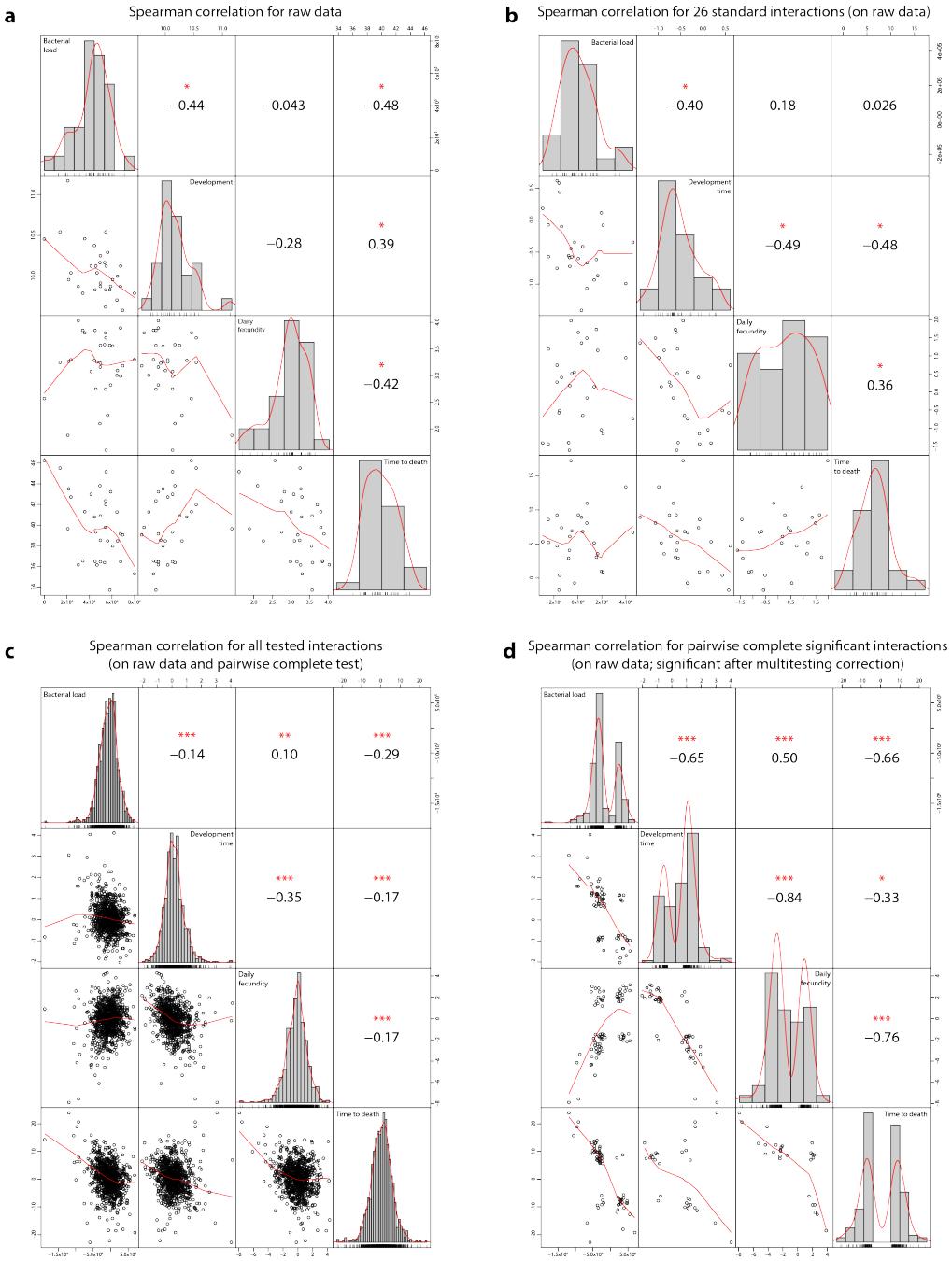


Figure S6. Correlations of raw phenotypes and interactions. (**a**) Raw data correlations between measured host and bacterial phenotypes indicate significant relationships between phenotypes except between fecundity and development and between fecundity and bacterial load. (**b**) Correlations of all standard interactions between measured host and bacterial phenotypes reveal different relationships than the raw phenotype data (compare with **a**). (**c**) Correlations of interaction strengths for all standard and non-standard tests (26+910=936) highlight significant relationships between all phenotypes. (**d**) Significant interactions have opposite sign between phenotypes except between bacterial load and fecundity, which have the same sign. We observed significant relationships between all phenotypes. Correlations of only significant interactions (where the interaction was significant for both phenotypes) after multiple comparison correction (standard and non-standard tests) between measured host and bacterial phenotypes are shown. Scatter plots appear below the diagonal and histograms are on the diagonal (gray bars = histogram; red line = kernel density function with Gaussian kernel). Correlations are all Spearman coefficients (data were not all normally distributed; Shapiro-Wilk tests). Significance values (*p<0.005; **p<0.001; ***p<0.0001) appear above the diagonal.

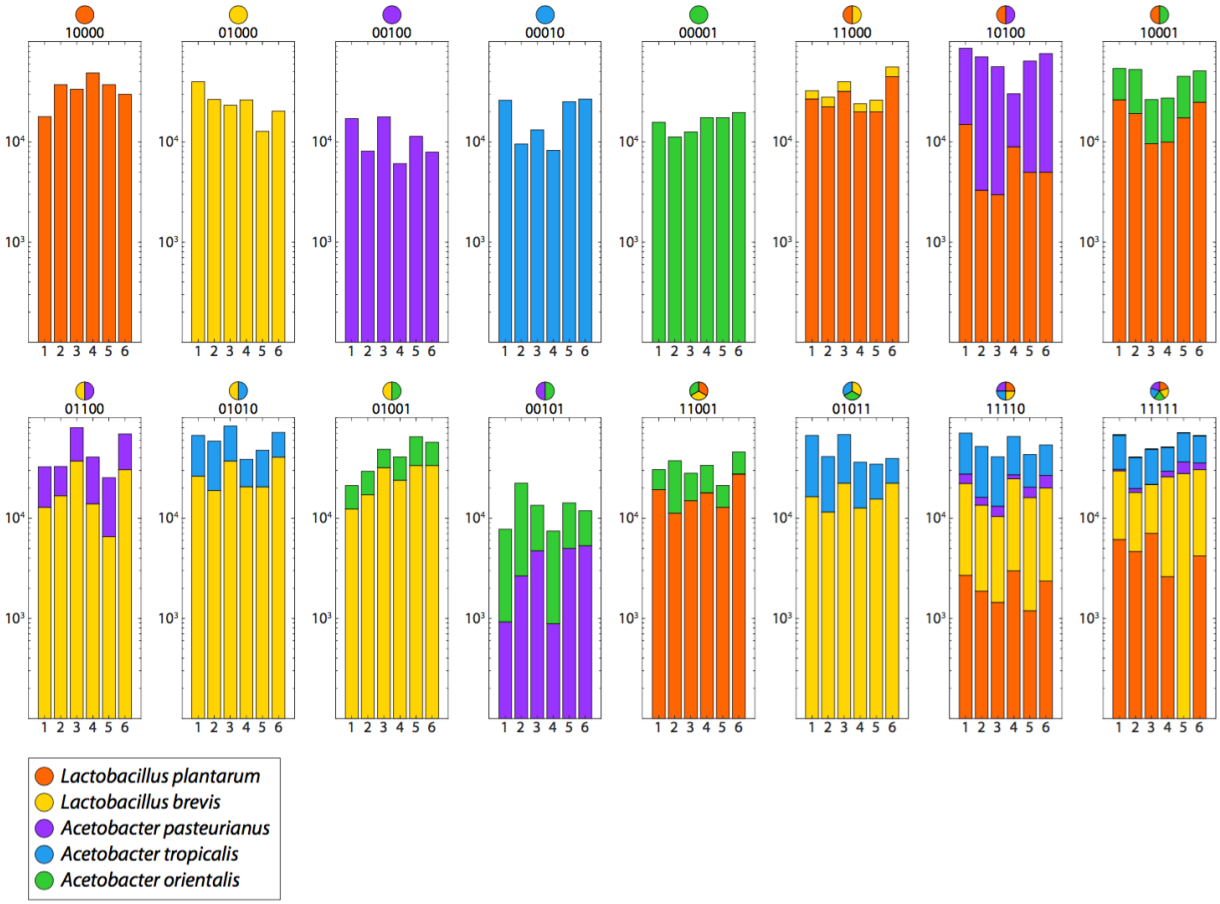


Figure S7. Raw bacterial abundance counts (CFUs) for fly food treatments with 16 selected bacterial combinations. X-axes indicate individual food samples 1 to 6. For each combination, the first three samples are from the first biological replicate, and the last three are from the second biological replicate. Y-axes are CFUs on a $\log_{10}$ scale.

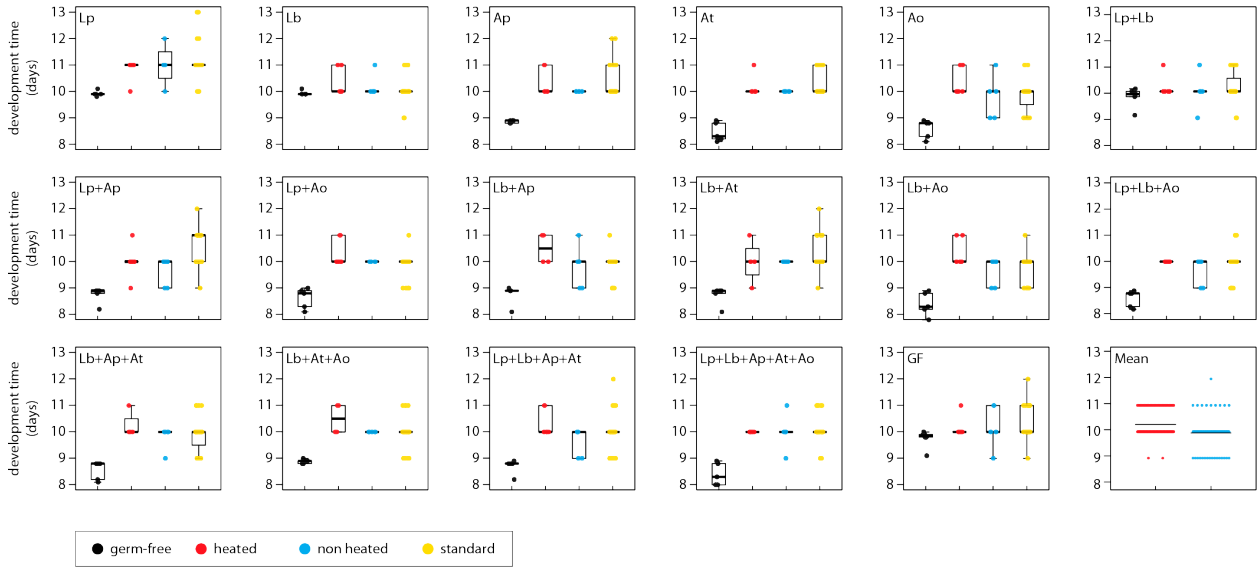


Figure S8. Parental effects and live bacteria influence offspring developmental pace. Sixteen microbial combinations and germ-free flies (as in Fig. S7) were tested for their impacts on developmental pace (number of days from egg laid to adult emergence from pupal case). In the original experiments, the developmental time was measured as eggs were directly laid by females continuously inoculated with their bacterial combination. To test the role of parental effects, we experimentally varied the source of the eggs as well as the bacterial treatment. Black points show data for n=20 eggs taken from germ-free mothers. Yellow points show the standard data (Fig. 2a) as a reference. Blue points show data where colonized vials (with flies and bacteria) were emptied of all their flies (and larvae) and then n=20 germ-free eggs were introduced. No significant differences were detected between yellow and blue treatments. Red points depict data where the treatment was identical to the blue points except that the vials were heat-killed of bacteria at 60˚C for 1 hour (and tested for sterility) prior to the introduction of germ-free eggs. This treatment significantly increased the development time (see main text). The fastest development times were for eggs introduced to fresh vials inoculated with bacteria but without previous fly occupation. In this final treatment (black dots), there was very little variation between treatments except that flies lacking all *Acetobacter* species were delayed by 1 to 2 days with respect to their cohort. Box and whisker plots: box shows 25th to 75th percentile. Thick bar shows median. Whiskers extend to 1.5x the interquartile range. All data points are also shown with each box.

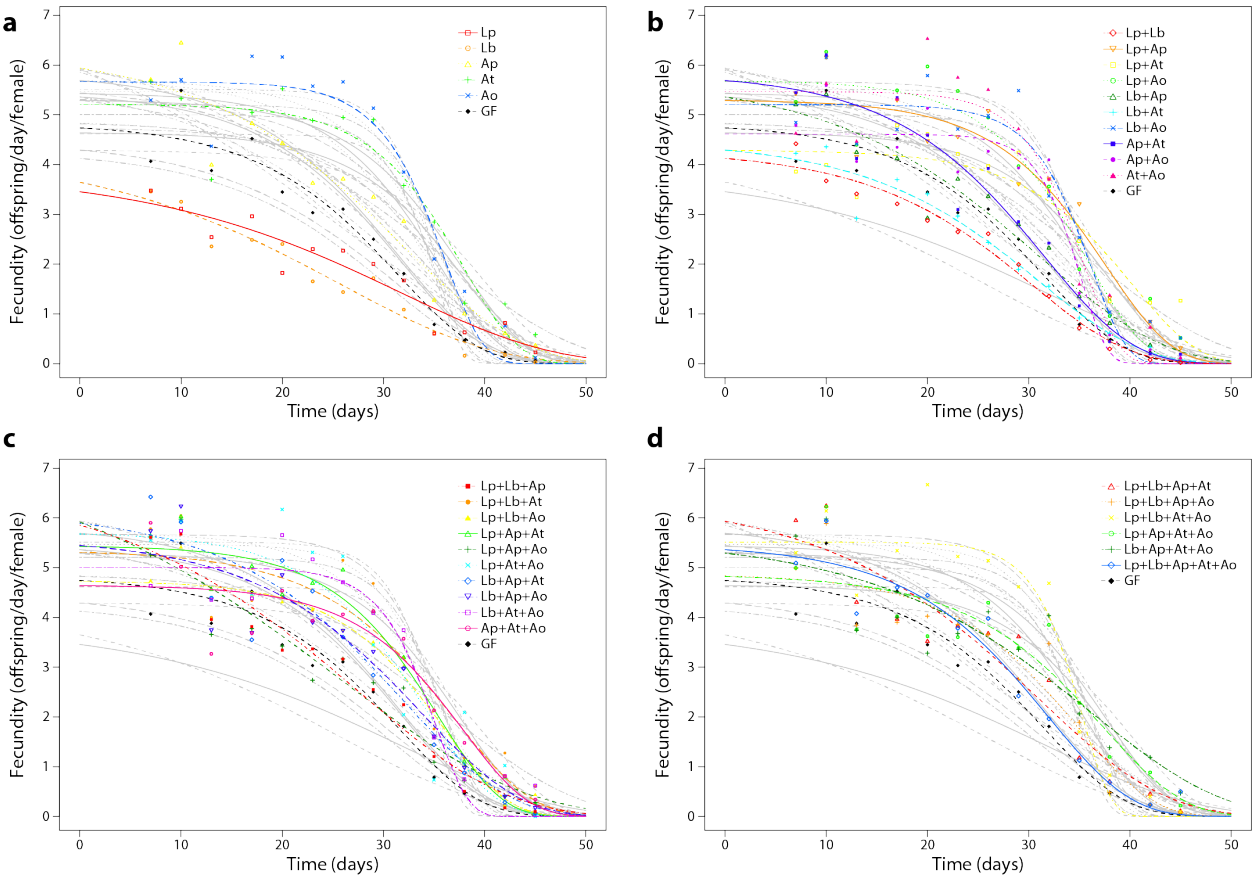


Figure S9. Curve fits to raw fecundity data aggregated from all 5 experimental replicates for each bacterial combination. Curve fits to a 3-parameter Gompertz distribution are depicted (see Methods). Bacterial combinations are grouped by the number of species. (**a**) Single species and germ-free flies. (**b**) Species pairs and germ-free flies. (**c**) Species trios and germ-free flies. (**d**) Species 4-way combinations, 5-way combination, and germ-free flies. All grayscale curves are kept as a reference.

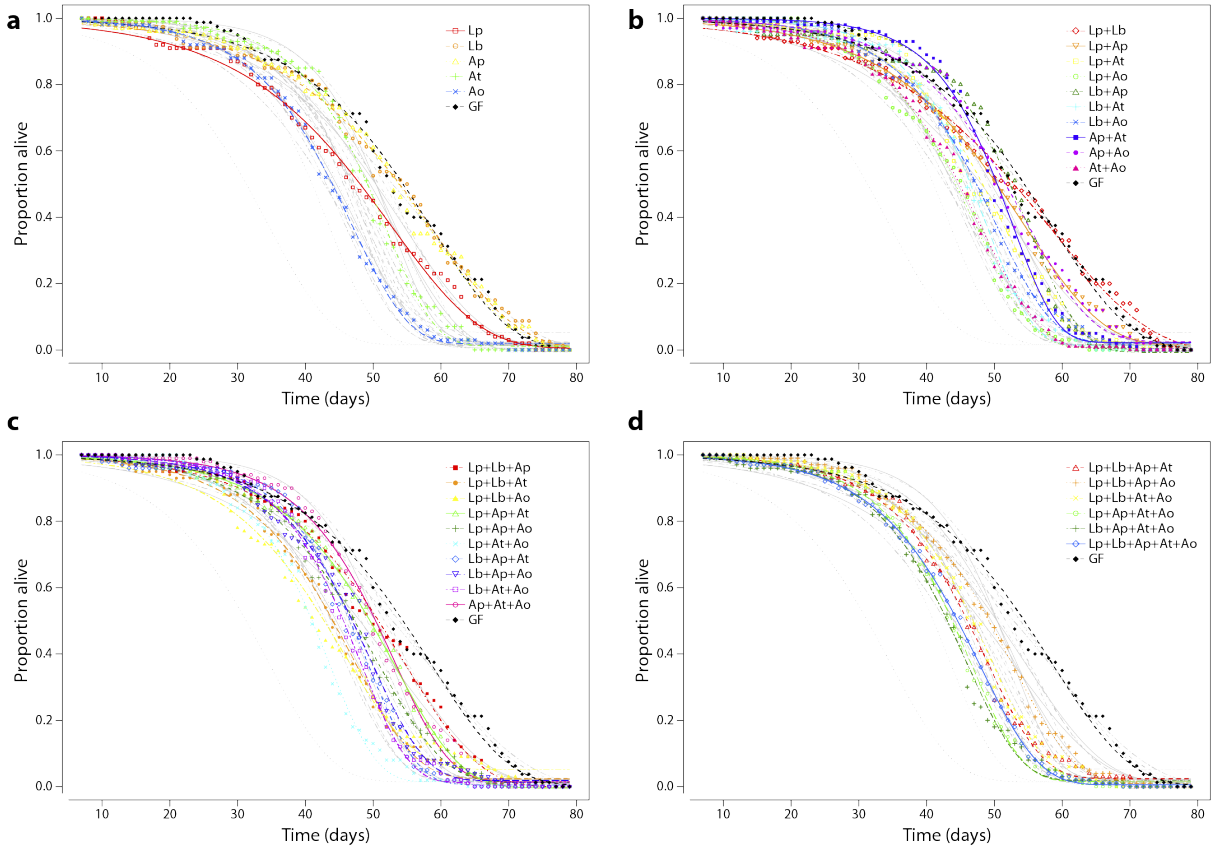


Figure S10. Curve fits to raw lifespan data aggregated from all 5 experimental replicates for each bacterial combination. Curve fits to a 3-parameter Gompertz distribution with an upper limit at one are depicted (see Methods). Bacterial combinations are grouped by the number of species. (**a**) Single species and germ-free flies. (**b**) Species pairs and germ-free flies. (**c**) Species trios and germ-free flies. (**d**) Species 4-way combinations, 5-way combination, and germ-free flies. All grayscale curves are kept as a reference.

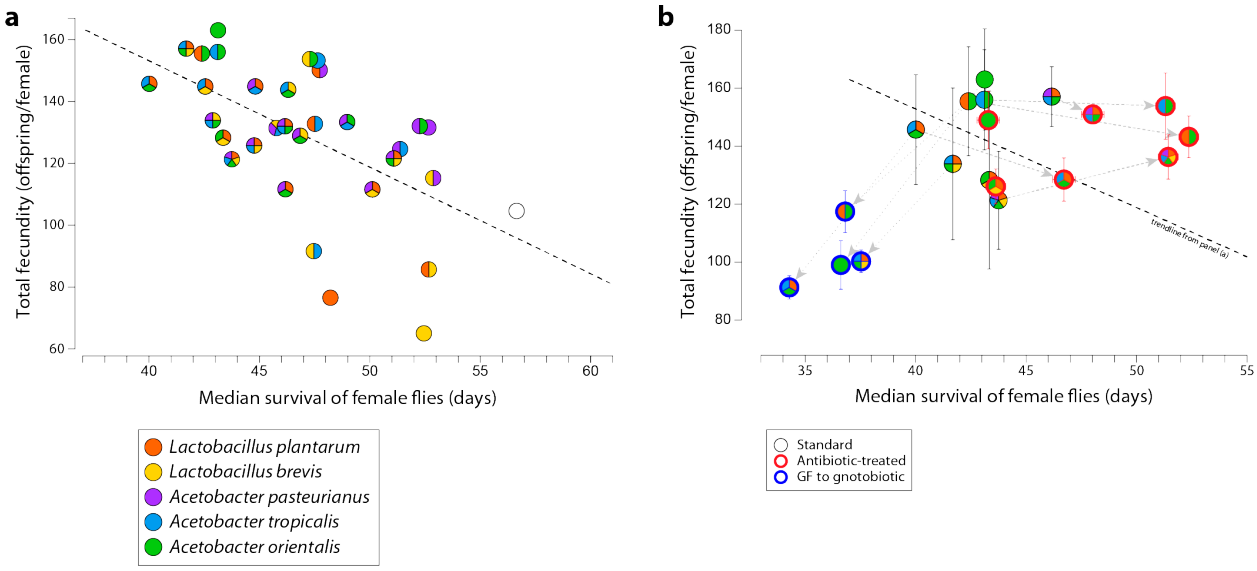


Figure S11. A fitness tradeoff exists between lifespan and fecundity. (**a**) In agreement with prior reports, higher fecundity is associated with shorter lifespan. This tradeoff is apparent for average daily fecundity as well as total fecundity per female. The scatter plot is based on the data from Fig. 2b-c for only female flies. (**b**) The lifespan fecundity tradeoff can be broken by putting flies on antibiotics after their peak reproduction (dots = data from Fig. 2b-c; red circles = gnotobiotic flies treated with antibiotics; see Methods) after 21 days, which encompasses the natural peak fecundity (Fig. S9). Note the shifts in lifespan between the regular treatment, the antibiotic treatment, and the late-life bacterial inoculation treatment. Lifespan was significantly extended, whereas total fecundity stayed high. Shifting germ-free flies to gnotobiotic treatment after 21 days of life decreased lifespan without increasing reproduction. n=100 flies per treatment for the standard and antibiotic-treated experiments n=60 flies per treatment for the germ-free switched to gnotobiotic experiment. Error bars show standard error of the mean.

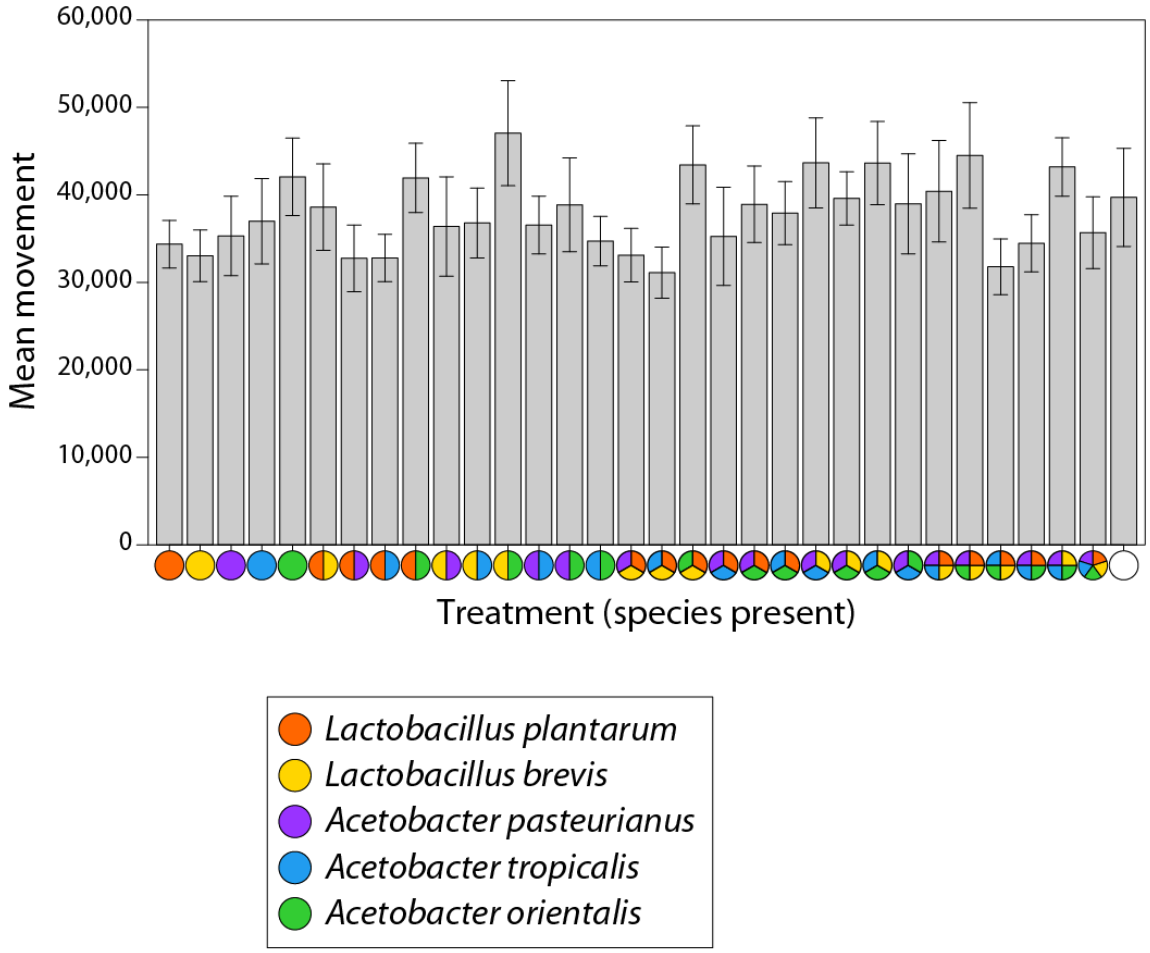


Figure S12. Average fly activity is unrelated to the fitness phenotypes. Fly movement is associated with overall metabolism, including food intake and energy expenditure. To search for behavior changes underlying the physiological differences in our bacterial treatments, we examined changes in fly motility for each bacterial treatment (n=32) in 5 replicate trials (n=20 flies per trial) using the LAMS (Trikinetics) population-based motility assay. Error bars show standard error of the mean. Trials were carried out for 7 days. Flies were flipped into fresh vials and placed in the activity-monitoring device. The first 24 h of data were removed to allow for fly acclimation to the new vial. Overall, we found no significant differences between bacterial combinations nor were there any correlations in the mean values with the other physiological data.

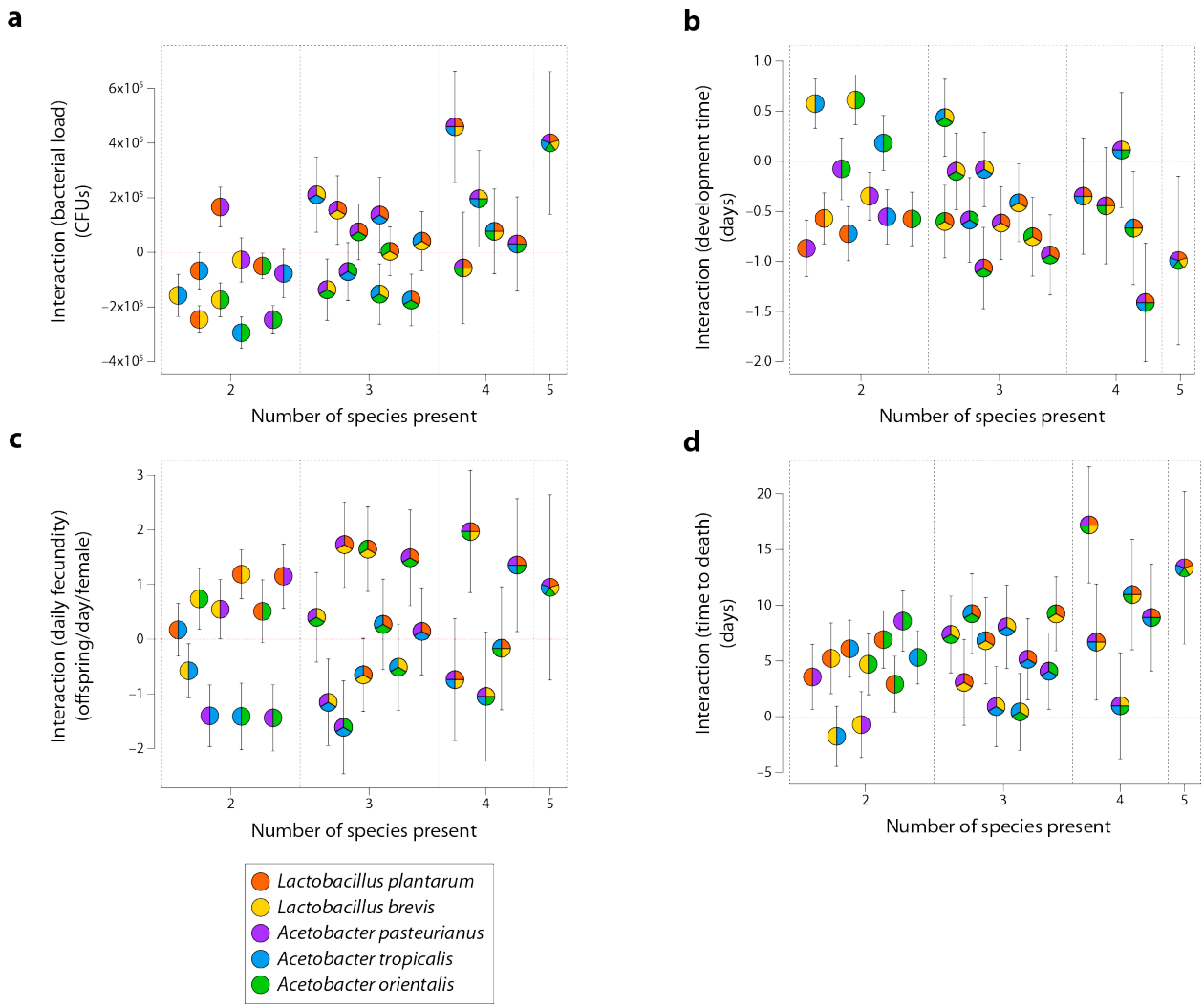


Figure S13. Standard interactions calculated for each phenotype in Fig. 2 and Fig. 3a-d. For all four phenotypes we computed the same 26 standard interactions. In the plots we separated interactions by the number of bacterial species involved (x-axis). The different bacterial combinations are expressed in colors. As general trends, we observe that the same interactions decrease for development time while they increase for time to death, when the diversity of the bacterial species in the microbiome increases. See T-polynomials (Math Supplement). For daily fecundity, we see that the magnitude of the interactions more specifically depend on the combination of present bacterial species. Error bars are the propagated error from the raw phenotypes.

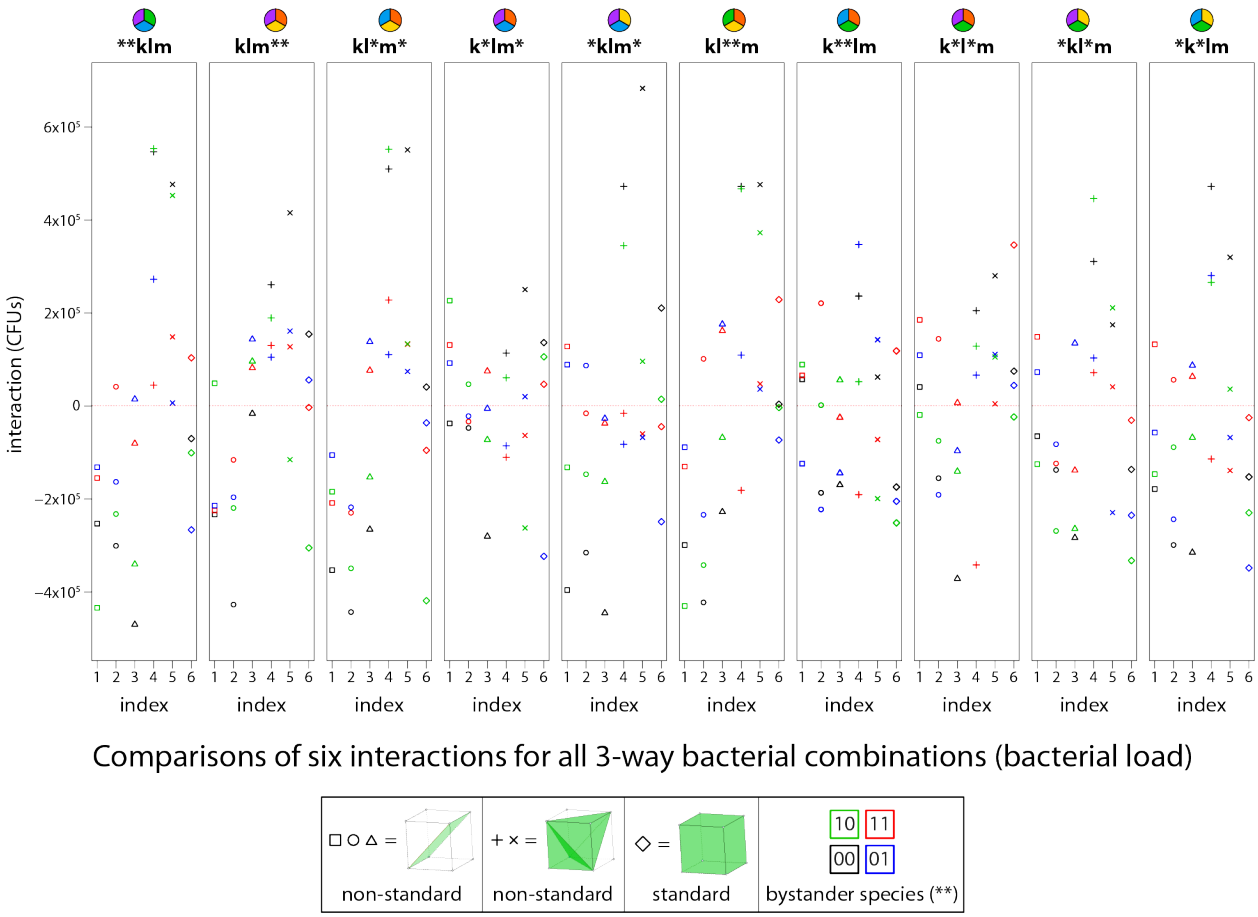


Figure S14. Detailed comparisons of the standard three-way interaction versus the related non-standard tests that incorporate w_{111} and w_{000} for total CFU counts. Interactions on the total bacterial load in flies between sets of three species (equations *g*=square, *i*=circle, *k*=triangle, *m*=plus, *n*=ex ('x'), and u_{111}=diamond) in Math Supplement) are compared to determine (i) whether additive non-standard tests can describe cases of non-additive standard tests and (ii) whether context of other species changes interactions. Each of the 10 combinations of three species (denoted in panel titles as "k", "l", and "m") is compared along with the four variants of bystander species (denoted in the panel titles as "*", and shown by the different colored symbols). The lack of correlation between the colors indicates that bystanders change interactions. As an example of an additive non-standard test explaining a non-additive standard test, consider the second panel from the left, klm**). Here, k=Lp, l=Lb, and m=Ap. Note that the black diamond (standard test) is positive, but the black triangle (equation *i*) is roughly zero, indicating that if we consider the Ap bacterial load and the load of the combination with Lp+Lb together, then they are equal to the load in flies with the bacterial combination of Lp+Lb+Ap.

Relevant equations (from Math Supplement):

$g = w_{000} - w_{100} - w_{011} + w_{111}$ (squares) non-standard test;

$i = w_{000} - w_{010} - w_{101} + w_{111}$ (circles) non-standard test;

$k = w_{000} - w_{001} - w_{110} + w_{111}$ (triangles) non-standard test;

$m = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000}$ (pluses) non-standard test;

$n = w_{110} + w_{101} + w_{011} - w_{000} - 2w_{111}$ (exes) non-standard test;

$u_{111} = w_{111} - (w_{110} + w_{101} + w_{011}) + (w_{100} + w_{010} + w_{001}) - w_{000}$ (diamonds), the standard test.

Figure S15. Detailed comparisons of the standard three-way interaction versus the related non-standard tests that incorporate w_{111} and w_{000} for development time. Interactions on the development time in flies between sets of three species (equations *g*=square, *i*=circle, *k*=triangle, *m*=plus, *n*=ex ('x'), and u_{111}=diamond) in Math Supplement) are compared to determine (i) whether additive non-standard tests can describe cases of non-additive standard tests and (ii) whether context of other species changes interactions. Each of the 10 combinations of three species (denoted in panel titles as "k", "l", and "m") is compared along with the four variants of bystander species (denoted in the panel titles as "*", and shown by the different colored symbols). The lack of correlation between the colors indicates that bystanders change interactions. As an example of an additive non-standard test explaining a non-additive standard test consider the first panel on the left, **klm). Here, k=Ap, l=At, and m=Ao. Note that the black diamond (standard test) is negative, but the black square (equation *k*) is roughly zero, indicating that if we consider the Ap development time and the At+Ao development time together, then they predict the development time in flies with the bacterial combination of Ap+At+Ao.

Relevant equations (from Math Supplement):

$g = w_{000} - w_{100} - w_{011} + w_{111}$ (squares) non-standard test;

$i = w_{000} - w_{010} - w_{101} + w_{111}$ (circles) non-standard test;

$k = w_{000} - w_{001} - w_{110} + w_{111}$ (triangles) non-standard test;

$m = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000}$ (pluses) non-standard test;

$n = w_{110} + w_{101} + w_{011} - w_{000} - 2w_{111}$ (exes) non-standard test;

$u_{111} = w_{111} - (w_{110} + w_{101} + w_{011}) + (w_{100} + w_{010} + w_{001}) - w_{000}$ (diamonds), the standard test.

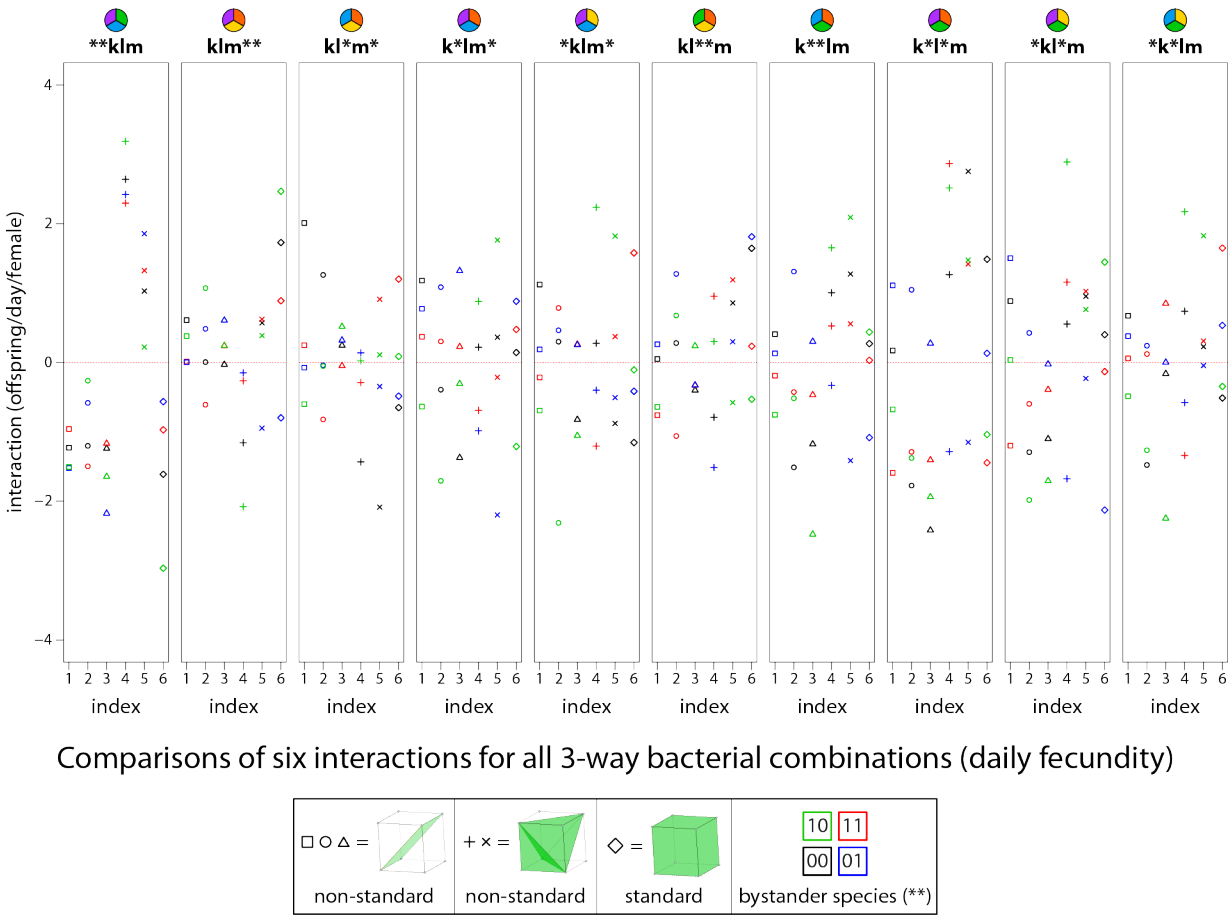


Figure S16. Detailed comparisons of the standard three-way interaction versus the related non-standard tests that incorporate w_{111} and w_{000} for daily fly fecundity. Interactions on the daily fecundity in flies between sets of three species (equations *g*=square, *i*=circle, *k*=triangle, *m*=plus, *n*=ex ('x'), and u_{111}=diamond) in Math Supplement) are compared to determine (i) whether additive non-standard tests can describe cases of non-additive standard tests and (ii) whether context of other species changes interactions. Each of the 10 combinations of three species (denoted in panel titles as "k", "l", and "m") is compared along with the four variants of bystander species (denoted in the panel titles as "*", and shown by the different colored symbols). The lack of correlation between the colors indicates that bystanders change interactions. As an example of an additive non-standard test explaining a non-additive standard test consider, the sixth panel from the left, kl**m. Here, k=Lp, l=Lb, and m=Ao. Note that the black diamond (standard test) is positive, but the black square (equation *g*) is roughly zero, indicating that if we consider the Lp fecundity and the Lb+Ao fecundity together, then they predict the fecundity in flies with the bacterial combination of Lp+Lb+Ao.

Relevant equations (from Math Supplement):

$g = w_{000} - w_{100} - w_{011} + w_{111}$ (squares) non-standard test;

$i = w_{000} - w_{010} - w_{101} + w_{111}$ (circles) non-standard test;

$k = w_{000} - w_{001} - w_{110} + w_{111}$ (triangles) non-standard test;

$m = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000}$ (pluses) non-standard test;

$n = w_{110} + w_{101} + w_{011} - w_{000} - 2w_{111}$ (exes) non-standard test;

$u_{111} = w_{111} - (w_{110} + w_{101} + w_{011}) + (w_{100} + w_{010} + w_{001}) - w_{000}$ (diamonds), the standard test.

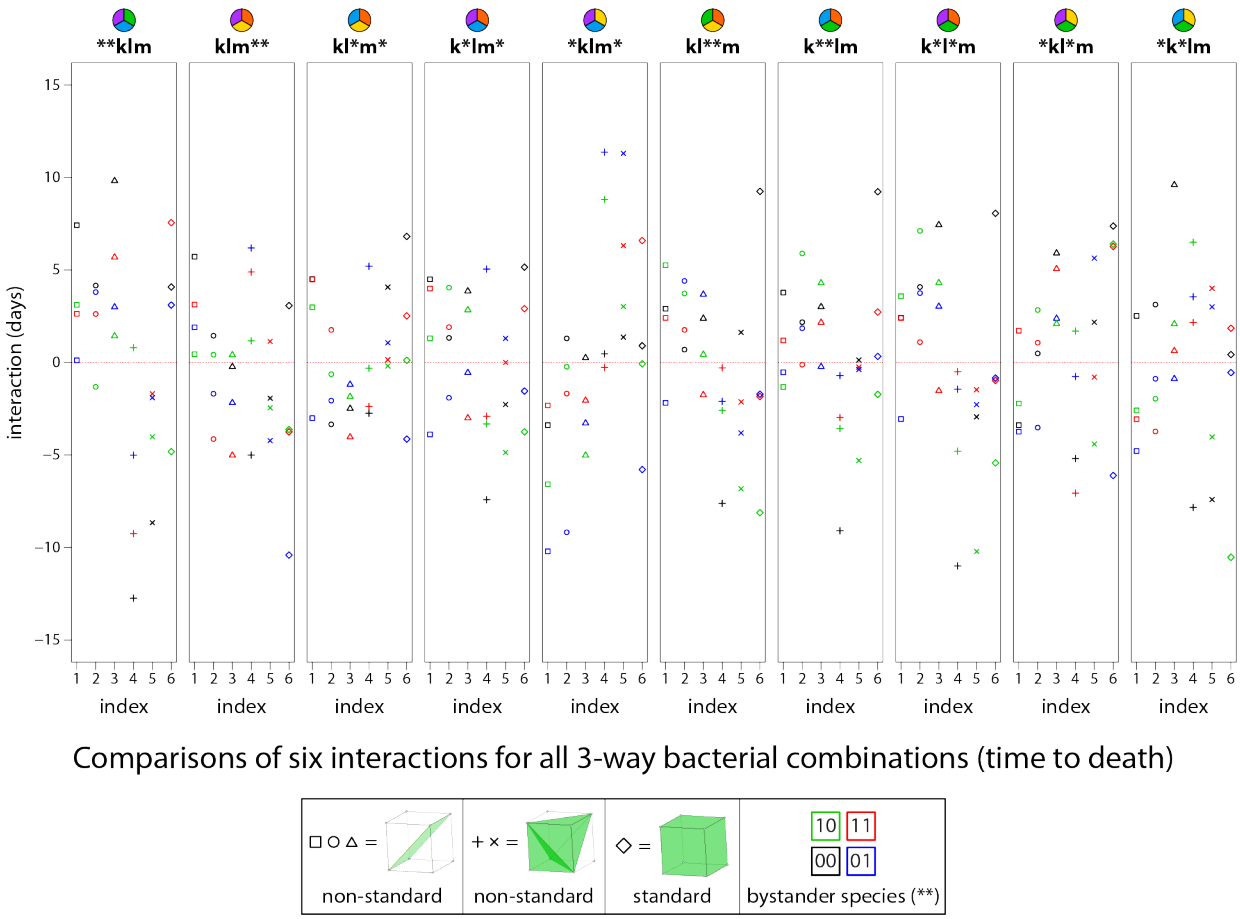


Figure S17. Detailed comparisons of the standard three-way interaction versus the related non-standard tests that incorporate w_{111} and w_{000} for time to death. Interactions on the total bacterial load in flies between sets of three species (equations *g*=square, *i*=circle, *k*=triangle, *m*=plus, *n*=ex ('x'), and u_{111}=diamond) in Math Supplement) are compared to determine (i) whether additive non-standard tests can describe cases of non-additive standard tests and (ii) whether context of other species changes interactions. Each of the 10 combinations of three species (denoted in panel titles as "k", "l", and "m") is compared along with the four variants of bystander species (denoted in the panel titles as "*", and shown by the different colored symbols). The lack of correlation between the colors indicates that bystanders change interactions. As an example of an additive non-standard test explaining a non-additive standard test consider the sixth panel from the left, kl**m). Here, k=Lp, l=Lb, and m=Ao. Note that the black diamond (standard test) is positive, but the black circle (equation *i*) is roughly zero, indicating that if we consider the Lb time to death and the Lp+Ao time to death together, then they predict the time to death in flies with the bacterial combination of Lp+Lb+Ao.

Relevant equations (from Math Supplement):

$g = w_{000} - w_{100} - w_{011} + w_{111}$ (squares) non-standard test;

$i = w_{000} - w_{010} - w_{101} + w_{111}$ (circles) non-standard test;

$k = w_{000} - w_{001} - w_{110} + w_{111}$ (triangles) non-standard test;

$m = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000}$ (pluses) non-standard test;

$n = w_{110} + w_{101} + w_{011} - w_{000} - 2w_{111}$ (exes) non-standard test;

$u_{111} = w_{111} - (w_{110} + w_{101} + w_{011}) + (w_{100} + w_{010} + w_{001}) - w_{000}$ (diamonds), the standard test.

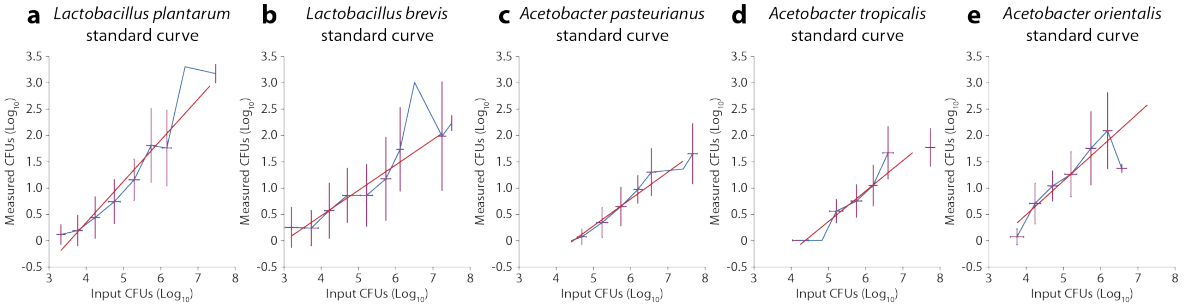


Figure S18. Standard curves used to calculate total CFUs from colony counts (Fig. S1, S7). A standard curve was constructed by plating known concentrations of bacteria on selective medium with a 96-pin replicator as in *Bacterial load counts* (Methods). Counts were fit to a power law curve using custom scripts in MATLAB v2017a.

Math Supplement: Mathematical Framework for Detecting Microbiome Interactions in the Host

December 18, 2017

Contents

1 Pairwise Species Interactions (Figure 1) 1

2 Introduction to Geometric Interactions (Figure 3) 3

3 Glossary for interactions - Mathematical Terminology 3

4 Interaction space 4
4.1 Interaction coordinates . . . 4
4.2 Circuits . . . 4
4.3 Recursively constructing higher order interactions . . . 5

5 Two disjoint families of interactions 6
5.1 Standard interactions . . . 7
5.2 Non-standard interactions . . . 9
5.3 Significance testing of interactions . . . 9

6 Supplemental Results of Interaction Testing 10
6.1 Results for standard interactions . . . 10
6.2 Results for all computed interactions . . . 13
6.3 Comparisons for the triple mutants . . . 15

7 Discussion of Interaction Coordinates 15

8 Supplements 16
8.1 Data transformation . . . 16

1 Pairwise Species Interactions (Figure 1)

Data Collection For each of the 32 fully combinatorial experiments in which flies are constantly fed defined bacteria on their food, we collect the mean CFU abundance for each of the 24 individual fly samples by averaging the three technical replicates for each sample. For each experiment, the median of these *sample CFU counts* may be taken to obtain *median CFU counts* (we use the median rather than the mean because the CFU count data is quite variable). The raw data are displayed in Figure S1.

Pairwise correlations (Figure 1b) The diversity of an experiment is the number of bacterial species that are provided on the food. For the set of experiments that have a given diversity N, we compute the pairwise correlation between bacteria i and j in the following way. For the subset of experiments of diversity N with microbes i and j, the sample CFU counts for i and j for each experiment are aggregated. Then, we calculate the Spearman's rank correlation coefficient of these pairs of data, and arrive at a pairwise correlation value between species i and j that is between -1 (ordinal CFU counts between the species are perfectly anticorrelated) and 1 (perfectly correlated). This process is repeated for each species pairing to build a pairwise correlation matrix for each diversity.

Pairwise interactions (Figure 1c) We infer the species interactions by assuming that the system obeys the generalized Lotka-Volterra equations,

$$\frac{d}{dt}x_i = x_i\left(\mu_i + \sum_{j=1}^{N} M_{ij}x_j\right),$$

and by assuming that the median CFU counts of each experiment are at equilibrium. At equilibrium, the time derivative on the left hand side vanishes, and we assume that the steady state is non-trivial (i.e. $x_i \neq 0$); if we call the steady state of each microbe for a given experiment $\tilde{x}_i$, then an experiment of diversity N will correspond to N algebraic equations of the form

$$0 = \mu_i + \sum_{j=1}^{N} M_{ij}\tilde{x}_j.$$

If there are M experiments of diversity N, then there are MN equations that must be simultaneously satisfied. We consider a "low-diversity" group that groups diversities 1 and 2 together (which consists of $1*5+2*10 = 25$ equations), and a "high-diversity" group that groups diversities 3, 4, and 5 together (which consists of $3*10+4*5+5*1 = 55$ equations). For each of these groups, we can rewrite these linear equations in matrix form as $0 = A\vec{x}$, where A is made up of the $\tilde{x}_i$ and $\vec{x} = [M_{11}, M_{12}, \ldots, M_{55}, \mu_1, \ldots, \mu_5]^T$. We assume $\mu_i = 1$ to obtain a nonzero result for M, which effectively absorbs the growth rates μ_i into the interaction values M_{ij}. We solve this system of equations for $\vec{x}$ with linear least-squares, since the linear system using the high-diversity group is overdetermined. The solution to this least-squares problem is the interaction matrix M, which we plot as a food web in Figure 1c.

Pairwise interactions (Figure S2a,b) Paine [7] presented a direct calculation of interaction strength, which we used to validate the overall trends we observed as a function of diversity. To measure the effect of a single species on the other species in a community, Paine removed the single species of interest and measured the resulting changes in abundance of the remaining species. The interaction strength is calculated as

$$\frac{treatmentCFUs - controlCFUs}{controlCFUs},$$

where $treatmentCFUs$ is the median species abundance with the species of interest absent and $controlCFUs$ is the median abundance when it is present [7]. To be clear, the CFU abundances in the equation are both (*treatment* and *control*) for the same species, while the removal is for a second species. Thus, an asymmetric matrix of interactions is calculated. Using the combinatorial experiments where each species has been removed and, in each removal, each remaining species has been quantified, allows us to measure the complete set of pairwise interactions for the cases where diversity goes from two to one species (Figure S2a) and from five to four species (Figure S2b) by removing one species at a time. We display the asymmetric matrices for the N=1,2 and N=4,5 cases as directed graphs.

Determining statistical trends For the four correlation matrices and the two interaction matrices, we show that both the correlations and the interactions become more negative at higher diversities. To achieve this, we compare the same matrix element across different diversities using the non-parametric Kendall rank correlation coefficient. Each matrix element is ranked according to its size over increasing diversities,

resulting in an ordinal vector of length 4 for the correlation matrix and of length 2 for the interaction matrix. Through the Kendall rank correlation coefficient, the ranking for each matrix element is compared to a strictly increasing vector ([1, 2, 3, 4] or [1, 2] for the two respective cases), resulting in a τ coefficient that is between -1 and 1. For the correlation matrices there are 10 such τ coefficients with a mean of -0.4 (corresponding to all possible pairings of 5 bacteria), and for the interaction matrices there are 25 τ coefficients with a mean of -0.44 (since interactions need not be symmetric, and self-interactions occur as well). The matrices becoming more negative at higher diversities would correspond to negative τ coefficients. To determine whether the distribution of τ coefficients is significantly more negative than a distribution centered at 0, we apply the Wilcoxon signed-rank test. The resulting one-sided p-values for the correlation matrices and interaction matrices are 0.0323 and 0.0139, indicating that there is a significant trend in the values of both the correlation and interaction matrices to decrease. When we assume that undetected species have an abundance of 1000 CFUs, corresponding to the limit of detection, the resulting one-sided p-values were 0.0323 (unchanged) for correlation matrices and 0.0359 for interaction matrices, indicating the results are robust to changes in CFUs below the limit of detection.

2 Introduction to Geometric Interactions (Figure 3)

For an n-genotype system, higher order genetic interactions, epistasis, are commonly described in terms of interaction coordinates and related forms see [2]. For example, marginal and conditional epistasis are formulas obtained with this approach, where one tests the effect of one set of loci on the interactions between other sets of loci. Interaction might also include circuits, see [2] for details. Besides a graspable biological interpretation, circuits also have a geometric interpretation, as these are formulas whose terms can be parametrized by certain sets of vertices on an n-cube and where n is the number of loci considered. This geometric interpretation is particularly useful, as it facilitates the parametrization of these types of formulas. Different sets of vertices in an n-cube then yield different circuits. In general, however, it is an open problem to determine all circuits in an n-cube, resp. in an n-genotype system, as the number increases dramatically with n, see [5] where the 4-dimensional case has been considered. Moreover, there is not a clear biological interpretation of all possible circuits of the n-cube.

In this work, we exploit the geometry of the 5-cube in terms of lower dimensional cubes and lift well known lower rank interactions to the higher dimensional space. For instance, we test whether the positive interaction between Ap and Lp stays positive when At is introduced. Clearly, with this approach we only examine a subset of all possible interactions, but the ones we find are interpretable and comparable with other studies (e.g. Newell and Douglas 2014 [6]). Moreover, we believe that our approach is simple enough to be scalable to higher dimensions. In the following, we explain how this approach generalizes to n-dimensional systems and to all lower rank interactions inside this system, and we give examples of how the mathematical approach applies to other biological systems specifically to bacterial interactions in the gut microbiome.

3 Glossary for interactions - Mathematical Terminology

The terminology we use in this work is an adaptation of the wording used in genetics to the study of interactions among bacterial species in fruit flies. For convenience we include the following intuitive definitions of terms we will repeatedly use later on:

- **n-species system**: a system of n types of bacterial species in the microbiome (present or absent).
- **n-cube (n dimensional unit hypercube)**: is an n-dimensional generalization of a square (2-cube) and three dimensional cube whose sides have unit lengths. When n=5 the 5-cube has 32 vertices, 80 edges, 80 square faces, 40 3-cubes and 10 4-cubes, see [3].
- **Interaction coordinates**: are basis elements for the interaction spaces associated to a system consisting of n-bacterial species (or an n-cube). Interaction coordinates are given by linear combinations of phenotypes associated to all combinations of species in the n-species system.
- **Circuits**: are certain linear combinations of interaction coordinates.

- **Triangulation**: is the local shape of the fitness landscape imposed by the phenotypes of neighboring genotypes (see Box 1a for an example).
- **Standard interactions (or standard tests)**: are 2, 3, 4 and 5-way interactions on all pairs, triples, quadruples and five tuples of species leaving the remaining species absent.
- **Non-standard interaction (or non-standard tests)**: higher-order interactions arising as a generalization of the standard test, interaction coordinates for the 3 and 4 and 5 species systems as well as circuits for the 3 species case, and by allowing the species not present to be occupied by bystanders, whose presence/absence is constant across the species considered in the standard test.

In this work, we will focus on a 5 species system consisting of 32 bacterial combinations. We encode the different bacterial combinations by a binary string S of lengths 5. Each entry of such a string S represents a bacterial species isolated inside a number of flies guts. For instance, S=00000 describes the germ-free fly, S=11111 describes the fly colonized with all 5 species of bacteria. With this binary notion, each bacterial combination also defines a vertex of a 5-dimensional cube G. Together with the 5-cube G, we also consider the following four phenotypes associated with the bacterial combinations: bacterial load (CFUs), development rate, fecundity, and time to death. These phenotypes associated with each bacterial combination in binary notation are denoted simply by w_S. The order of the bacterial species and the specific bacterial loads are summarized in Table 1.

Table 1: **The 32 different bacterial loads.**

4 Interaction space

In this section, we recall the notion of interaction coordinates and circuits (for the three cube) first defined in [2]. We first describe these spaces abstractly for arbitrary values of n. In a second part, we describe how one can recursively extend these formulas to define new interaction formulas in higher dimensional settings.

4.1 Interaction coordinates

Let $i = i_1, i_2, \ldots, i_n$ be a vertex of an n-dimensional cube with at least two coordinates i_j, i_k being 1. The interaction coordinates u_i can be defined (up to a scalar) in the following way:

$$u_{i_1,i_2,\ldots,i_n} := \frac{1}{2^n - 1} \sum_{j_1=0}^{1} \sum_{j_2=0}^{1} \cdots \sum_{j_n=0}^{1} (-1)^{i_1 j_1 + i_2 j_2 + \cdots + i_n j_n} w_{j_1 j_2 \ldots j_n} \tag{1}$$

where w_- are values of a corresponding phenotypes and indexed by the vertices of the n-dimensional cube. It then follows, that there are $2^n - n - 1$ interaction coordinates. Moreover, these coordinates are linearly independent and form a vector space basis of the interaction space.

In genetics, interaction coordinates are also called marginal epistasis in [2].

4.2 Circuits

Linear combinations of interaction coordinates might give rise to circuits. In the following, the only circuits we will consider are the ones previously presented in the 3-dimensional setting in [2]. In formulas, these

circuits can be presented as follows:

$$
\begin{aligned}
a &:= u_{110} + u_{111} = w_{000} - w_{010} - w_{100} + w_{110} \\
b &:= u_{110} - u_{111} = w_{001} - w_{011} - w_{101} + w_{111} \\
c &:= u_{101} + u_{111} = w_{000} - w_{001} - w_{100} + w_{101} \\
d &:= u_{101} - u_{111} = w_{010} - w_{011} - w_{110} + w_{111} \\
e &:= u_{011} + u_{111} = w_{000} - w_{001} - w_{010} + w_{011} \\
f &:= u_{011} - u_{111} = w_{100} - w_{101} - w_{110} + w_{111} \\
g &:= u_{110} + u_{101} = w_{000} - w_{011} - w_{100} + w_{111} \\
h &:= u_{110} - u_{101} = w_{001} - w_{010} - w_{101} + w_{110} \\
i &:= u_{110} + u_{011} = w_{000} - w_{010} - w_{101} + w_{111} \\
j &:= u_{110} - u_{011} = w_{001} - w_{011} - w_{100} + w_{110} \\
k &:= u_{101} + u_{011} = w_{000} - w_{001} - w_{110} + w_{111} \\
l &:= u_{101} - u_{011} = w_{010} - w_{011} - w_{100} + w_{101}
\end{aligned}
$$

$$
\begin{aligned}
m &:= -u_{011} - u_{101} - u_{110} - u_{111} = w_{001} + w_{010} + w_{100} - w_{111} - 2w_{000} \\
n &:= -u_{011} - u_{101} - u_{110} + u_{111} = w_{011} + w_{101} + w_{110} - w_{000} - 2w_{111} \\
o &:= u_{011} + u_{101} - u_{110} - u_{111} = w_{010} + w_{100} + w_{111} - w_{001} - 2w_{110} \\
p &:= u_{011} + u_{101} - u_{110} + u_{111} = w_{000} + w_{011} + w_{101} - w_{110} - 2w_{001} \\
q &:= u_{011} - u_{101} + u_{110} - u_{111} = w_{001} + w_{100} + w_{111} - w_{010} - 2w_{101} \\
r &:= u_{011} - u_{101} + u_{110} + u_{111} = w_{000} + w_{011} + w_{110} - w_{101} - 2w_{010} \\
s &:= -u_{011} + u_{101} + u_{110} + u_{111} = w_{000} + w_{101} + w_{110} - w_{011} - 2w_{100} \\
t &:= -u_{011} + u_{101} + u_{110} - u_{111} = w_{001} + w_{010} + w_{111} - w_{100} - 2w_{011}.
\end{aligned}
$$

Biological and geometric interpretations of these circuits follow from the descriptions given in [2, §.3] for the genetic setting.

4.3 Recursively constructing higher order interactions

We now describe how interaction coordinates in lower dimensional cubes and the circuits a-t extend to interaction formulas in higher dimensional cubes.

For instance, the two-way interaction

$$u_{11} = w_{00} + w_{11} - w_{01} - w_{10}$$

extends to the following $80 = 10 \cdot 2^3$ different interaction formulas in a five species system:

$$
\begin{aligned}
\alpha_{**klm} &= w_{00klm} + w_{11klm} - w_{01klm} - w_{10klm} \\
\alpha_{*k*lm} &= w_{0k0lm} + w_{1k1lm} - w_{0k1lm} - w_{1k0lm} \\
\alpha_{*kl*m} &= w_{0kl0m} + w_{1kl1m} - w_{0kl1m} - w_{1kl0m} \\
\alpha_{*klm*} &= w_{0klm0} + w_{1klm1} - w_{0klm1} - w_{1klm0} \\
\alpha_{k**lm} &= w_{k00lm} + w_{k11lm} - w_{k01lm} - w_{k10lm} \\
\alpha_{k*l*m} &= w_{k0l0m} + w_{k1l1m} - w_{k0l1m} - w_{k1l0m} \\
\alpha_{k*lm*} &= w_{k0lm0} + w_{k1lm1} - w_{l0lm1} - w_{k1lm0} \\
\alpha_{kl**m} &= w_{kl00m} + w_{kl11m} - w_{kl01m} - w_{kl10m} \\
\alpha_{kl*m*} &= w_{kl0m0} + w_{kl1m1} - w_{kl0m1} - w_{kl1m0} \\
\alpha_{klm**} &= w_{klm00} + w_{klm11} - w_{klm01} - w_{klm10},
\end{aligned}
\tag{2}
$$

where $**$ indicates two species out of five. The remaining indices k, l, m are then either 0 or 1, and all possible 2^3 combinations are allowed. In genetics, two way interactions of this type where considered for the five locus case in [2, §.7].

Similarly, one extends to the five species system the four interaction coordinates $u_{111},\ u_{110}, u_{011}, u_{101}$ defined by (1) above in the following fashion:

$$\beta_{***kl} = \sum_{j_1=0}^{1}\sum_{j_2=0}^{1}\sum_{j_3=0}^{1}(-1)^{*j_1+*j_2+*j_3}w_{j_1j_2j_3kl}$$

$$\beta_{**k*l} = \sum_{j_1=0}^{1}\sum_{j_2=0}^{1}\sum_{j_4=0}^{1}(-1)^{*j_1+*j_2+*j_4}w_{j_1j_2kj_4l}$$

$$\beta_{**kl*} = \sum_{j_1=0}^{1}\sum_{j_2=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_1+*j_2+*j_5}w_{j_1j_2klj_5}$$

$$\beta_{*k*l*} = \sum_{j_1=0}^{1}\sum_{j_3=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_1+*j_3+*j_5}w_{j_1kj_3lj_5}$$

$$\beta_{*kl**} = \sum_{j_1=0}^{1}\sum_{j_4=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_1+*j_4+*j_5}w_{j_1klj_4j_5}$$

$$\beta_{k*l**} = \sum_{j_2=0}^{1}\sum_{j_4=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_2+*j_4+*j_5}w_{kj_2lj_4j_5}$$

$$\beta_{kl***} = \sum_{j_3=0}^{1}\sum_{j_4=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_3+*j_4+*j_5}w_{klj_3j_4j_5}$$

$$\beta_{*k**l} = \sum_{j_1=0}^{1}\sum_{j_3=0}^{1}\sum_{j_4=0}^{1}(-1)^{*j_1+*j_3+*j_4}w_{j_1kj_3j_4l}$$

$$\beta_{k**l*} = \sum_{j_2=0}^{1}\sum_{j_3=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_2+*j_3+*j_5}w_{kj_2j_3lj_5}$$

$$\beta_{k***l} = \sum_{j_2=0}^{1}\sum_{j_3=0}^{1}\sum_{j_4=0}^{1}(-1)^{*j_2+*j_3+*j_4}w_{kj_2j_3j_4l}.$$

As before, the notation $***$ indicates three species out of five. The remaining indices k, l are assumed to be fixed and either $kl = 00, kl = 01, kl = 11$ or $kl = 10$. Moreover, the sums are taken over integer numbers between 0 and 1.

Given these extended interaction coordinates also allows us to extend the circuits *a*-*t* to the five species setting. To do this, it is enough to consider the circuit formulas *a*-*t* given above, and replace the u_{i_1,i_2,i_3} with the corresponding extended interaction coordinates. The biological and geometric interpretation of these interaction formulas can easily be deduced from the lower dimensional ones (given in [2] for the three species case).

Similarly one also extends the $2^4 - 5 = 11$ interaction coordinates $u_{i_1,\ldots,i_4}$ defined by equation (1) and $n = 4$.

5 Two disjoint families of interactions

In the following, we are interested in comparing the outcomes of two disjoint families of interaction formulas in the five bacterial species setting. The first family is smaller in size and consists of the standard interactions. These involve a variety of different combinations of bacterial species, without involving excessively many computations per test.

The second family (non-standard interactions) consists of a large number of different interaction formulas recursively defined. These interactions are more accurate in describing the interaction of the bacterial combinations under inspections and depend on the presence of the bystanders [2].

These two families of interactions we consider in this work are new and inspired by [1] and [4] (the second reference for standard tests).

5.1 Standard interactions

Consider the two species interaction formula, given by:

$$u_{11} = w_{00} + w_{11} - w_{01} - w_{10}$$

where w_{00} indicates a phenotype, such as daily fecundity, time to death, CFU or development rate, associated to the (empty) bacterial combination 00, and similarly for w_{11}, w_{01} and w_{10}. Biologically, the meaning of this interaction formula is well understood: it compares the phenotype contributions of the wild and double mutant with the phenotype contributions of the single mutants. Geometrically, the summands of u_{11}, w_{00}, w_{11}, w_{01} and w_{10}, are indexed by the four vertices of a unital square.

We then consider the following generalizations of u_{11} for the 3-, 4- and 5-species system:

$$u_{111} = w_{000} + w_{011} + w_{101} + w_{110} - w_{100} - w_{010} - w_{001} - w_{111}$$

$$\begin{aligned} u_{1111} = {} & w_{0000} + w_{1100} + w_{0110} + w_{0011} + w_{1010} + w_{0101} + w_{1001} + w_{1111} \\ & - w_{1000} - w_{0100} - w_{0010} - w_{0001} - w_{1011} - w_{1101} - w_{0111} - w_{1110} \end{aligned}$$

$$\begin{aligned} u_{11111} = {} & w_{00000} - w_{00001} - w_{00010} - w_{00100} - w_{01000} - w_{10000} + w_{11000} \\ & + w_{10100} + w_{10010} + w_{10001} + w_{01100} + w_{01010} + w_{01001} + w_{00110} \\ & + w_{00101} + w_{00011} - w_{11100} - w_{11010} - w_{11001} - w_{10110} - w_{10101} \\ & - w_{10011} - w_{01110} - w_{01101} - w_{01011} - w_{00111} + w_{11110} + w_{11101} \\ & + w_{11011} + w_{10111} + w_{01111} - w_{11111}. \end{aligned}$$

In the following, we simply call u_{11} the 2-way interaction, u_{111} the 3-way interaction, u_{1111} the 4-way interaction, and u_{11111} the 5-way interaction. From the given formulas it is clear that these interactions are defined by 4, 8, 16, and 32 terms, respectively. The signs change according to the definition of the interaction coordinate function given in equation (1). Biologically, one can say that each such 3, 4 and 5-way interaction compares the contribution of phenotypes when an even number of bacterial combinations (equivalent to mutations) were inserted versus contributions of phenotypes measured when an odd numbers of bacterial combinations have been inserted (e.g. w_{00} and w_{11} vs w_{01} and w_{10}).

From the 2-way interaction above one deduces a further 10 two-way interactions in a five species system by considering all combinations of two species and leaving the remaining bacterial species are not present. The three-way interaction gives rise to 10 further three-way interactions of this type, arising by considering three species out of five and letting the remaining bacterial combinations not present. Similarly, the four-way interaction yields 5 similar four-way interactions by letting the remaining species absent. Together, this approach gives:

$$\binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} = 26 \tag{3}$$

different higher order interactions in a five species system, which we call standard interactions. All of these standard interactions compare certain phenotype contributions indexed by an even number of bacteria species present, versus an odd number of bacteria species present. The number of standard interactions we found in Equation (3) matches the results of [4].

To summarize, the standard interactions involving two species of bacteria out of five are:

$$
\begin{aligned}
u_{00**0} &= w_{00000} + w_{00110} - w_{00010} - w_{00100} \\
u_{**000} &= w_{00000} + w_{11000} - w_{01000} - w_{10000} \\
u_{*0*00} &= w_{00000} + w_{10100} - w_{00100} - w_{10000} \\
u_{0**00} &= w_{00000} + w_{01100} - w_{00100} - w_{01000} \\
u_{*00*0} &= w_{00000} + w_{10010} - w_{00010} - w_{10000} \\
u_{0*0*0} &= w_{00000} + w_{01010} - w_{00010} - w_{01000} \\
u_{00*0*} &= w_{00000} + w_{00101} - w_{00001} - w_{00100} \\
u_{*000*} &= w_{00000} + w_{10001} - w_{00001} - w_{10000} \\
u_{0*00*} &= w_{00000} + w_{01001} - w_{00001} - w_{01000} \\
u_{000**} &= w_{00000} + w_{00011} - w_{00001} - w_{00010}
\end{aligned}
$$

These u-interactions arise from the 2-way interaction and always involve two species (indicated with $*$) out of five leaving the other three species absent. Geometrically, these u-interactions involve the four vertices of certain square faces inside a 5-dimensional cube G.

The standard interactions, which involve three species out of five, will be denoted by u and are obtained from the 3-way interaction. These u-interactions involve the eight vertices of the specified cubes inside G:

$$
\begin{aligned}
u_{00***} &= w_{00000} + w_{00011} + w_{00101} + w_{00110} - w_{00100} - w_{00010} - w_{00001} - w_{00111} \\
u_{***00} &= w_{00000} + w_{01100} + w_{10100} + w_{11000} - w_{10000} - w_{01000} - w_{00100} - w_{11100} \\
u_{**0*0} &= w_{00000} + w_{01010} + w_{10010} + w_{11000} - w_{10000} - w_{01000} - w_{00010} - w_{11010} \\
u_{*0**0} &= w_{00000} + w_{00110} + w_{10010} + w_{10100} - w_{10000} - w_{00100} - w_{00010} - w_{10110} \\
u_{0***0} &= w_{00000} + w_{00110} + w_{01010} + w_{01100} - w_{01000} - w_{00100} - w_{00010} - w_{01110} \\
u_{**00*} &= w_{00000} + w_{01001} + w_{10001} + w_{11000} - w_{10000} - w_{01000} - w_{00001} - w_{11001} \\
u_{*00**} &= w_{00000} + w_{00011} + w_{10001} + w_{10010} - w_{10000} - w_{00010} - w_{00001} - w_{10011} \\
u_{*0*0*} &= w_{00000} + w_{00101} + w_{10001} + w_{10100} - w_{10000} - w_{00100} - w_{00001} - w_{10101} \\
u_{0**0*} &= w_{00000} + w_{00101} + w_{01001} + w_{01100} - w_{01000} - w_{00100} - w_{00001} - w_{01101} \\
u_{0*0**} &= w_{00000} + w_{00011} + w_{01001} + w_{01010} - w_{01000} - w_{00010} - w_{00001} - w_{01011}
\end{aligned}
$$

The following five standard interactions arise from the 4-way interaction described by u_{1111} above and involve four species out of the five, leaving out the remaining bacterial community:

$$
\begin{aligned}
u_{****0} &= w_{00000} + w_{11000} + w_{01100} + w_{00110} + w_{10100} + w_{01010} + w_{10010} + w_{11110} \\
&\quad - w_{10000} - w_{01000} - w_{00100} - w_{00010} - w_{10110} - w_{11010} - w_{01110} - w_{11100} \\
u_{***0*} &= w_{00000} + w_{11000} + w_{01100} + w_{00101} + w_{10100} + w_{01001} + w_{10001} + w_{11101} \\
&\quad - w_{10000} - w_{01000} - w_{00100} - w_{00001} - w_{10101} - w_{11001} - w_{01101} - w_{11100} \\
u_{**0**} &= w_{00000} + w_{11000} + w_{01010} + w_{00011} + w_{10010} + w_{01001} + w_{10001} + w_{11011} \\
&\quad - w_{10000} - w_{01000} - w_{00010} - w_{00001} - w_{10011} - w_{11001} - w_{01011} - w_{11010} \\
u_{*0***} &= w_{00000} + w_{10100} + w_{00110} + w_{00011} + w_{10010} + w_{00101} + w_{10001} + w_{10111} \\
&\quad - w_{10000} - w_{00100} - w_{00010} - w_{00001} - w_{10011} - w_{10101} - w_{00111} - w_{10110} \\
u_{0****} &= w_{00000} + w_{01100} + w_{00110} + w_{00011} + w_{01010} + w_{00101} + w_{01001} + w_{01111} \\
&\quad - w_{01000} - w_{00100} - w_{00010} - w_{00001} - w_{01011} - w_{01101} - w_{00111} - w_{01110}
\end{aligned}
$$

Geometrically, these interactions involve the 16 vertices of the specified 4-cubes inside G. The last standard

interaction is simply given by the following expression involving all five species and all 32 fitness values:

$$\begin{aligned}
u_{*****} &= u_{11111} \\
&= w_{00000} - w_{00001} - w_{00010} - w_{00100} - w_{01000} - w_{10000} + w_{11000} \\
&+ w_{10100} + w_{10010} + w_{10001} + w_{01100} + w_{01010} + w_{01001} + w_{00110} \\
&+ w_{00101} + w_{00011} - w_{11100} - w_{11010} - w_{11001} - w_{10110} - w_{10101} \\
&- w_{10011} - w_{01110} - w_{01101} - w_{01011} - w_{00111} + w_{11110} + w_{11101} \\
&+ w_{11011} + w_{10111} + w_{01111} - w_{11111}.
\end{aligned}$$

Geometrically, this interaction involves all vertices of G. In Figure 1 below we highlight the regions delimited by the vertices defining the 26 standard interactions described above inside a projection of a five cube G. In all cases, the sign ($+$ or $-$) of the summand (i.e. the phenotype being added) is defined by the the number of bacterial species involved.

For example, in a three bacterial system consisting of the following 8 bacterial combinations $G = \{000, 001, 010, 100, 110, 011, 101, 11\}$ there are three standard interactions involving two out of three species:

$$\begin{aligned}
u_{**0} &= w_{000} + w_{110} - w_{100} - w_{010} \\
u_{*0*} &= w_{000} + w_{101} - w_{100} - w_{001} \\
u_{0**} &= w_{000} + w_{011} - w_{010} - w_{001}
\end{aligned}$$

together with the 3-way interaction u_{111}, described above.

5.2 Non-standard interactions

In the following, we compare the results of these standard tests with the corresponding non-standard tests in a five species system. By this we mean all the interactions we described in Section 4 which are different from the standard interaction formulas described in Section 5.1. Thus, non-standard tests are higher-order interactions arising by allowing the species not present under the standard test, circuits and interaction coordinates to be occupied by bystanders, whose presence/absence is constant across the species considered in the standard test. It follows that non-standard tests measure how the context of additional species change the interaction effects.

The standard tests and non-standard tests we consider in this work are examples of the more generally defined marginal and conditional interactions in genetics context. To specify the species we condition on and to differentiate among types of conditional interactions, we refrain from using this terminology.

We computed 910 different non-standard interactions and 26 different standard interactions. From Section 4 and the recursive nature of the five dimensional cube, it follows that the non-standard interactions we inquire have a well understood biological and geometric interpretation.

Finally, we note that despite the formulas describing non-standard and standard interactions all being different, the two sets of interactions are not statistically independent as they use the same underlying data sets.

5.3 Significance testing of interactions

In order to decide whether an interaction is non-zero, we took the uncertainty in the phenotype measurements into account and devised a statistical test as follows. Since the phenotype measurements come with their standard errors, we computed the corresponding propagation of error for each interaction (standard and non). This error was determined by taking the square root of the sum of squares of the standard errors involved in each interaction formula. For example, for u_{11}, the propagated standard error $s(u_{11})$ is

$$s(u_{11}) = \sqrt{s_{00}^2 + s_{11}^2 + s_{01}^2 + s_{10}^2}$$

where s_{00} denotes the standard error of w_{00}, etc.

Moreover, if different formulas give rise to the same interaction term, we considered only the formula yielding the smallest propagation error, that is, the formula involving the least number of operations. For

instance, each interaction term $a, \ldots, t$ can be defined in two ways, one involving a difference between interaction coordinates and a direct way. The direct way involves less operations, and therefore a smaller propagation error.

We then determined significant interactions in the following way. We assumed that each interaction formula comes from a Gaussian distribution with mean 0 and standard deviation given by the computed propagation error. For each interaction u, we performed a two-sided null hypothesis test. The null hypothesis states that the true value of the interaction is $u = 0$, versus the alternative hypothesis that $u \neq 0$. We then considered an interaction statistically significant if the p-value was below 0.05. This means, that if the null hypothesis was true, the probability of obtaining the result of the interaction u we computed would be 5%.

In total, we tested 936 interactions (26 standard tests plus 910 non-standard tests). To account for the multiple comparisons, we corrected all p-values using the Benjamini-Hochberg procedure in order to control the false discovery rate at 5%; that is, we eventually considered an interaction statistically significant if the corrected p-value was below 0.05.

6 Supplemental Results of Interaction Testing

In this section we summarize the main findings we obtained through the computations described above. In the first part, we focus on the smaller family of higher-order interactions and on the second part we compare the outcomes of these interactions with non-standard interactions. In the last part, we focus on specific standard and non-standard interactions which compare the wild type and the triple mutants in different ways.

6.1 Results for standard interactions

To summarize the results of the standard interactions presented above we average the results of the standard 2-, 3-, 4- and 5-way tests. Thus, we let a_2 be the sum of all standard 2-way interactions u_{**} divided by the number of 2-way standard interactions (which is 10). Similarly, for a_3, a_4 and a_5.

Term	**CFU**	**dev**	**fd**	**td**
a_2	-117062	-0.235	-0.052	4.085
a_3	8920	-0.472	0.175	5.436
a_4	141250	-0.552	0.274	8.946
a_5	400470	-0.990	0.948	13.355

Table 2: The average interactions terms $a_2, \ldots, a_5$ for the standard interactions and the four phenotypes of fecundity per day (fd), CFU, time to death (td) and development rate (dev).

From Table 2 we deduce that in average the interaction tends to increase with the number of terms in the standard formulas. We also note that the sign of the interaction (positive or negative) is already determined by the sign of the average 2- and 3-way interaction. This means that when the average contribution of the two (or three) species combinations is higher than the average contribution of the single species combinations, also the average contribution of an even number of species is higher than the average contribution of an odd number of species (regardless of how many even or odd species are considered 2,4,8,16) as determined by the standard tests for interactions. Together with the results of Table 2, we also consider the average interactions of the standard 2-,3-, 4- and 5-way interactions normalized by the number of present bacterial species. That is $n_2 = a_2/2$,$n_3 = a_3/3$, $n_4 = a_4/4$ and $n_5 = a_5/5$, see Table 3 for the corresponding results.

Term	CFU	dev	fd	td
n_2	-58531	-0.117	-0.026	2.043
n_3	2972	-0.157	0.058	1.812
n_4	35312.5	-0.138	0.069	2.237
n_5	80094	-0.198	0.190	2.671

Table 3: The normalized terms $n_2, \ldots, n_5$ for then standard interactions, normalized by the number of bacteria present for the four phenotypes of fecundity per day (fd), CFU, time to death (time d) and development rate (dev).

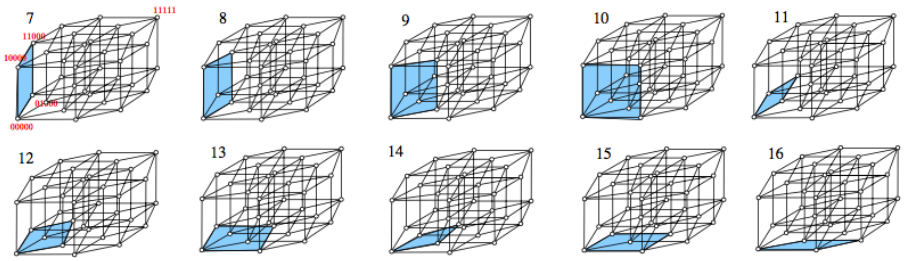
It is clear from examining the average interaction values (i.e. the terms $a_2, \ldots, a_5$ in Table 2) that the total contribution increases as the number of species increases. However, when we normalize these average values to the number of species, we see a more constant contribution for each individual species, see Table 3.

Thus, if interactions quantify the degree to which we can predict the phenotype of the microbiome when a new species is added, overall the microbiome is less predictable as we add additional species. However, on a per-species basis, the degree of unpredictability stays constant. And if we consider the number of combinations, the unpredictability decreases. Thus, while our analysis indicates that the microbiome problem increases in complexity as more species are added, there is reason for hope. For instance, if we discover a rule that determines *a priori* which non-standard test will be additive (versus showing interaction), the predictability of the microbiome will increase as we add species. However, our fundamental conclusion is that the relationships between species rather than the species themselves produce increasingly complex interactions. Therefore, our efforts at building a predictable framework should focus on the interactions between species as much as on the interactions between the individual species themselves.

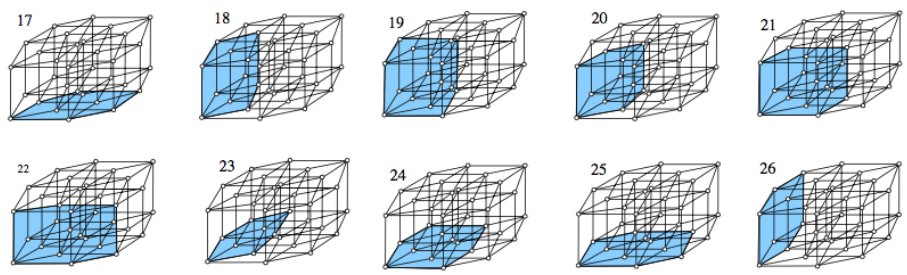
To this point, when we compare the distributions of the raw data with the results of the standard interactions, we see that linear trends in the raw data do not necessary translate to the same trends in the standard interaction data (Figure S6). For example, consider time to death: the raw data indicates that increasing the number of bacterial species results in a decrease in the time to death. However, the standard tests indicate that interaction tends to increase with the number of species involved (see Table 2), and hence with the numbers of terms in the standard interaction formulas. Finally, from the data in the Fig. S13, we deduce that the above observations remain valid by considering the (fewer) significant standard interactions rather than all standard interactions (Figure S6). For completeness, we also computed Spearman's correlation coefficients on all 26 standard interactions, and on the raw CFUs, development rate, daily fecundity, time to death data. See Table 4. The tests in bold in Table 4 which reached significance level ($p < 0.05$). We also see that the correlation might change sign when passing from the raw data to the computed standard interactions (as it is, for example, for time to death vs development rate). See Table 6 for the results of the normality test (Shapiro-Wilk).

	CFU	Development rate	Daily fecundity	Time to death
CFU	-	**-0.44** (p=0.0108)	-0.04 (p=0.8142)	**-0.35**(p=0.0465)
Development rate	**-0.40** (p=0.0412)	-	-0.28 (p=0.1246)	**0.39** (p=0.0257)
Daily fecundity	0.18 (p=0.3859)	**-0.49** (p=0.0117)	-	**-0.42** (p=0.0179)
Time to death	0.03 (p=0.8984)	**-0.48** (p=0.0127)	0.36 (p=0.0708)	-

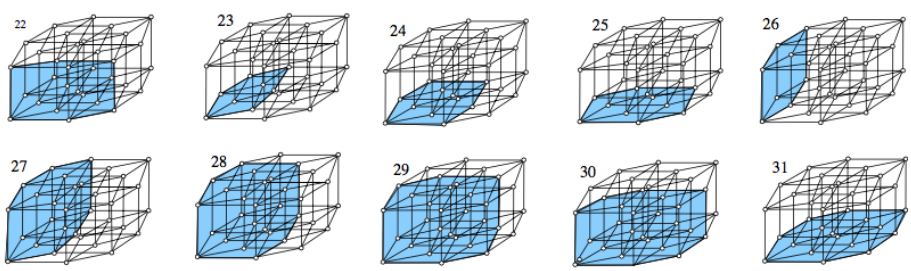
Table 4: **Spearman's correlation coefficients and significance level for all 26 standard interactions and raw data measurements**. Below the diagonal, we indicate Spearman's correlation coefficients and the corresponding p-values for all standard tested interactions (26 samples), similarly above the diagonal for the raw data (32 data points).



(A) The 10 standard 2-way interactions



(B) The 10 standard 3-way interactions



(C) The 5 standard 4-way interactionsl

Figure 1: **Geometric description of the 26 standard interactions**. The highlighted regions inside the projections of the 5-dimensional cube indicate the vertices (indicated by $*$ in the standard interaction formulas) involved in the corresponding interaction. The interaction u_{11111} is defined by all 32 vertices and therefore omitted.

6.2 Results for all computed interactions

In this section, we describe the results contained in Fig. 3 and Fig. S6 in the main text, obtained while computing the 910 non-standard interactions together with 26 standard interactions for the phenotypes of CFU, development rate, daily fecundity, time to death. As main findings we showed that there are many significant positive and negative interactions among bacterial species in fruit flies. Moreover, among all intricate ways bacterial species interact with each other we found that for some phenotypes (CFUs and daily fecundity) the standard interactions essentially fully capture interaction trends measured also by the non-standard interactions. However, for development rate and time to death we found that many new significant interactions arise under mixing procedures which cannot be quantified by the standard tests alone. Moreover, comparing the correlation Table 4 with Table 7 we found correlated interactions between phenotypes where neither the raw measurements nor the standard tests were correlated. More results, on all the computed correlations are in Figure S6 (main text).

The coefficients a_i and n_j for all non-standard interactions are then as indicated in the following Table 5. From the definition of the bipyramidal circuits (m -t) it is already clear that, the average contribution of these six circuits vanish, i. e. the sum of all six circuits m to t divided by 6 is zero, even if each interaction outcome is different. The same conclusion remains true, when one considers the average contribution of all 320 (different) higher order interactions obtained from the circuits m-t to the five species setting. For this reason, the average and normalized contributions of these 320 higher order interactions is omitted in Table 5.

Term	**CFU**	**dev**	**fd**	**td**
a_2	-49672.063	0.050	-0.023	0.211
a_3	-92665.250	0.158	-0.210	0.763
a_4	-176164.545	0.170	-0.376	1.492
a_5	-250353.846	0.147	-0.400	3.252
n_2	-24836.031	0.020	-0.011	0.106
n_3	-30888.417	0.053	-0.070	0.254
n_4	-44041.136	0.042	-0.094	0.373
n_5	-50070.769	0.029	- 0.080	0.650

Table 5: The columns *dev* indicates the coefficients $a_2, \ldots a_5$ and $n_2, n_3, \ldots, n_5$ for all (standard and non) interactions for tests development rate, *daily fecundity* for daily fecundity and $time_d$ for time to death.

For each phenotype, the coefficients $a_2, \ldots a_5$ from Table 5 might be visualized in Figure 3 (a)-(d) (in the main text), as the averages interactions in each of the 5 sectors. In fact, the first sector (from 0 to 320) indicates the interactions for all 2-way interactions and all interactions obtained from circuits a to l. The average of this interaction is given by $a2$ (this number is not represented in the figure). The second sector (from 321 to 640) indicates the interaction of all circuits m to t. As we just explained, their average interaction is zero. The third sector (from 641 to 800) indicates the interactions for all interactions arising from the interaction coordinates involving three species out of five. Their average interaction is given by a_3. Similarly, the sector from 801 to 911 summarizes all interactions coming from the interaction coordinates involving four species out of five. Their average interaction is a_4. Finally, the last sector (from test 912 to test 936) indicates the interactions obtained from the 26 interaction coordinates on five species. Their average contribution is given by a_5. In this same way, for standard interactions one can visualize in Figure S13 (a)-(d) (main text) the coefficients of Table 2. The sectors in Figure S13 indicate the species present in each standard interaction. More precisely, the first sector say in FS13 (a) shows the outcome of the 10 standard 2-way interactions (which involve two species out of five). The second sector show the outcome of the 10 standard 3-way interactions. The four the outcome of the 5 4-way interactions and the last sector

shows the outcome of u_{11111}, the only standard 5-way interaction. Geometrically, these interactions where described above in Figure 1.

Finally, we underpin the above findings by testing if the two probability distributions corresponding to standard and non-standard interactions come from the same continuous distributions. The set of non-standard test does not include the set of standard tests we computed. However, since standard tested and non-standard tests are computed from the same underlying data, the two sets of tests are statistically dependent. To test the difference between the distributions of the two sets we performed a two-sample and two-sided Kolmogorov-Smirnov (KS) test. The results of the KS test for the phenotypes of CFU (D = 0.2011, p-value = 0.2583) and daily fecundity (D = 0.13407, p-value = 0.7539) indicate that there is little evidence to reject the null-hypothesis of STD tests and non-STD tests coming from the same distributions. On the other hand, the results of the KS test for the phenotypes of development rate (D = 0.45934, p-value $= 4.661 \cdot 10^{-05}$) and time to death (D = 0.50769, p-value $= 4.384 \cdot 10^{-06}$) suggest that it is unlikely that the STD tests and non-STD tests come from the same distributions.

For completeness, we also ask if any of the STD tests, non-STD tests and significant standard test come from normal distributions. To test this hypothesis we performed a Shapiro-Wilk test. The results are summarized in Table 6. We see that for the non-standard interactions (significant and non) the Shapiro-Wilk tests are significant. Therefore we conclude that the corresponding distributions fail the normality test.

	W	***p*-value**
CFU	0.97077	$1.509 \cdot 10^{-12}$
CFU STD	0.95504	0.3034
CFU significant	0.91889	$1.008 \cdot 10^{-11}$
Development	0.96411	$3.619 \cdot 10^{-14}$
Development STD	0.95998	0.3912
Development significant	0.90617	$8.869 \cdot 10^{-09}$
Daily fecundity	0.97426	$1.337 \cdot 10^{-11}$
Daily fecundity STD	0.94793	0.2074
Daily fecundity significant	0.90894	$5.636 \cdot 10^{-09}$
Time to death	0.98732	$4.374 \cdot 10^{-07}$
Time to death STD	0.97510	0.7568
Time to death significant	0.90930	$4.233 \cdot 10^{-08}$

Table 6: **Shapiro-Wilk-Test**.For each phenotype we summarized the test statistic and the corresponding p-value. The p-values for all non-standard tests (significant and non) reached a significance level. We conclude that there is evidence that the various distributions are non-normal. 'STD' appended to a label indicates the interactions where standard, whereas if 'STD' is not appended, it is for the non-standard tests. If 'significant' is appended, it means that the non-standard tests we consider are the ones that reached statistical significance (p¡0.05) after multiple testing correction.

	CFU	Development	Daily fecundity	Time to death
CFU	-	**-0.14** ($p < 0.0005$)	**0.10** ($p < 0.0005$)	**-0.29** ($p < 0.0005$)
Development	**-0.65** ($p < 0.0005$)	-	**-0.35** ($p < 0.0005$)	**-0.17** ($p < 0.0005$)
Daily fecundity	**0.50** ($p < 0.0005$)	**-0.84** ($p < 0.0005$)	-	**-0.17** ($p < 0.0005$)
Time to death	-0.66 ($p = 0.0827$)	**-0.33** (p= 0.0104)	**-0.76** ($p < 0.0005$)	-

Table 7: **Spearman's correlation coefficients and significance level for all tested interactions**. Below the diagonal, we indicate Spearman's correlation coefficients and the corresponding p-values for statistically significant interactions ($p < 0.05$) after multiple testing correction with 27 to 78 pairwise complete comparisons). Similarly, above the diagonal we compute correlations for all non-standard tested interactions regardless of statistical significance (910 data points).

6.3 Comparisons for the triple mutants

In this section, we explain how interactions involving all 10 three-species combinations in a five species system behave under mixing, see Fig. S14-S17 in the main text. More specifically, to obtain these figures we consider the circuits g, i, k, m, n and the interaction coordinate u_{111}. We then lifted these interactions to the five species system, as described in Section 4.3. In particular, in this way, we obtained that each such interaction yields 40 interactions in the five species system (as $\binom{5}{3} = 10$ and there are 4 choices for the bystanders). In each figure (Fig. S14-S17 in the main text) we focused on one phenotype (CFU, development rate, daily fecundity or time to death). In each plot we represented with the same color the six interactions corresponding to g, i, k, m, n and u_{111}. The different colors then reflect the state of the bystanders: black if the bystanders are both 00, red if the bystanders are both mutated, and blue (resp. green) if only one bystander is mutated.

7 Discussion of Interaction Coordinates

Overall, when we utilize all of the interactions, including the non-standard tests, we see more significant interactions for the different phenotypes. Consistently, we see that the same interaction coordinates have correlated values across the different phenotypes that we measured. A simple explanation is that the rich microbial interactions underlying these phenotypes affect some central aspect of fly physiology that is reflected in multiple life history traits.

With the goal of quantifying higher order interactions we computed various interaction formulas. We also extracted a set of 26 interaction formulas (which we called standard interactions) and compared them with the results of 910 (different) interaction formulas. We found that for certain phenotypes the standard interactions approximate well the distribution of the other more involved computations. Since the number of all possible interactions is large, and unknown in general, it is important to find a more parsimonious approach based on fewer interactions that still capture the main interaction signals. This is particularly advisable since the number of all possible interactions increases with the number of species and the relative contributions coming from each test are mainly small. Moreover, analyzing smaller sets of particularly expressive and biologically interpretable interactions would not only be computationally more efficient, but also facilitate the comparisons of higher order genetic interactions arising in different biological context (for example infected fruit flies, or fruit flies treated with antibiotics). Moreover, the recursive approach we proposed here to define higher order interaction coordinates, as well as the computations we carried out, easily extends to settings with more species. Thus, our methodology provides a way to reduce the experimental burden of examining combinatorial interactions in the microbiome. We also observe that our computations are based on discrete data points. But since the nature of the phenotypes we analyzed is continuous, it would be interesting to extend our studies and consider a fitting continuous setting. Finally, our computations and conclusion consider the propagation of uncertainty in phenotype measurements, but

it would be interesting to develop a quantitative statistical framework accounting for possible sensible noise in the data set.

8 Supplements

8.1 Data transformation

All the interactions we compute are based on and generalize the additive genetic epistasis formula $u_{11} = w_{00} + w_{11} - w_{01} - w_{10}$. This additive formula relates to the multiplicative formula $m_{11} = w_{00}w_{11} - w_{01}w_{10}$ up to a logarithmic transformation. That is, composing the phenotype with a log transformation, we have:

$$\log(w_{00}) + \log(w_{11}) - \log(w_{01}) - \log(w_{10}) = \log(w_{00}w_{11}) - \log(w_{01}w_{10}).$$

To highlight that significant interactions we find do not depend on the additive approach we choose, we computed the same interactions as above also for the logarithmic (in base 2) transformation of the data. With no surprise, the interactions might dependent on the choice of the data transformation. More generally, we conclude by observing that transformations (possibly) depend on the true distribution of the observed (measured) data. Since the true distribution of our measurement remains unknown, we find it more reasonable to present our findings based on the actual measured data.

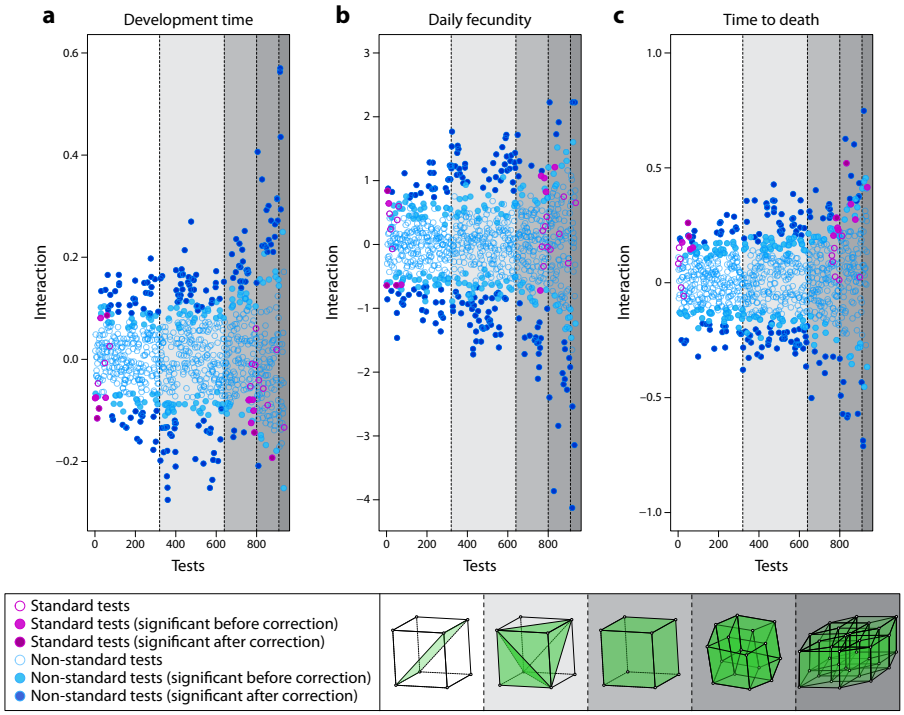


Figure 2: **Scatter plot of all tested interactions on** $\log_2$ **transformed data** (standard in purple and non standard interactions in blue). Interactions for (a) development time, (b) daily fecundity, (c) time to death. Filled circles indicate significant interactions, open circles represent non-significant interactions. Dark blue and dark purple filled circles indicate significance (p¡0.05) after multiple testing correction. Filled light blue and filled light purple dots indicate significance (p¡0.05) before corrections.

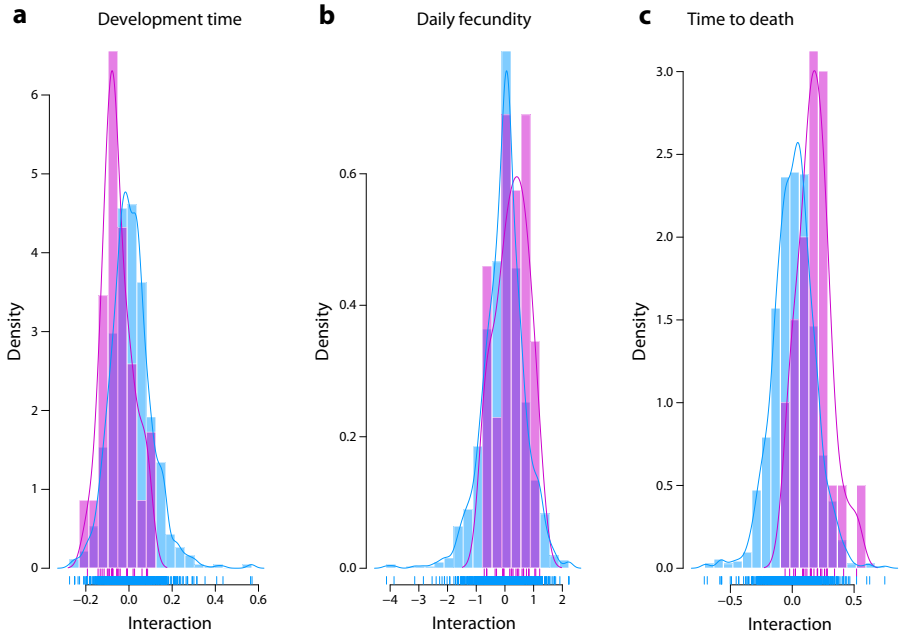


Figure 3: **Density plot of all tested interactions (standard in purple and non standard in blue) for the** $\log_2$ **transformed data.** Interactions for (a) development time, (b) daily fecundity, (c) time to death.

References

[1] N. Beerenwinkel, L. Pachter, B. Sturmfels, SF. Elena and RE. Lenski, *Analysis of epistatic interactions and fitness landscapes using a new geometric approach*, BMC Evolutionary Biology 20077:60 (2007)

[2] N. Beerenwinkel, L. Pachter, B. Sturmfels, *Epistasis and shapes of fitness landscapes* Statistica Sinica 17, 1317-1342 (2007)

[3] H. S. M. Coxeter.*Regular Polytopes*, London, Methuen,New York, Pitman 1949. 20-321 (1948)

[4] J. Grilli, G. Barabs, MJ. Michalska-Smith, S. Allesina, *Higher-order interactions stabilize dynamics in competitive network models*, Nature. 2017 Aug 10;548(7666):210-213. (2017)

[5] P. Huggins, B. Sturmfels, J. Yu, and DS. Yuster. *The Hyperdeterminant and Triangulations of the 4-Cube.* Mathematics of Computation 77, no. 263 (2008)

[6] P. D. Newell, A. E. Douglas, *Interspecies Interactions Determine the Impact of the Gut Microbiota on Nutrient Allocation in Drosophila melanogaster*, Appl. Environ. Microbiol, vol. 80 no. 2 788-796 (2014)

[7] R. T. Paine *Food-web analysis through field measurement of per capita interaction strength*, Nature 355, 73-75 (1992)