PNAS
www.pnas.org

Supplementary Information for

Microbiome interactions shape host fitness

Alison L. Gould,[1,2] Vivian Zhang,[1,2] Lisa Lamberti,[3,4] Eric W. Jones,[5] Benjamin Obadia,[1] Nikolaos Korasidis,[3] Alex Gavryushkin,[3,4,6] Jean M. Carlson,[5] Niko Beerenwinkel,[3,4] William B. Ludington[1,7*]

1. Dept Molecular & Cell Biology, Division of Genetics, Genomics & Development, Division of Immunology & Pathogenesis, UC Berkeley, Berkeley, CA 94720
2. equal contributors
3. Dept Biosystems Science & Engineering, ETH Zurich, 8092 Zurich, Switzerland
4. Swiss Institute of Bioinformatics, 4058 Basel, Switzerland
5. Complex Systems Group, Dept Physics, UC Santa Barbara, Santa Barbara, CA 93106
6. Dept Computer Science, University of Otago, Dunedin 9016, New Zealand
7. Dept Embryology, Carnegie Institution for Science, Baltimore, MD 21218

*Lead contact: ludington@carnegiescience.edu

This PDF file includes:

Supplemental Materials and Methods
Figs. S1 to S18
Table S1-S5
Math Supplement

No other supplementary materials

www.pnas.org/cgi/doi/10.1073/pnas.1809349115

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 (44). 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 400 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 using a sterile cotton swab to individual vials containing sterile fly medium (10% filter-sterilized glucose, 5% autoclaved active dry yeast, 1.2% autoclaved agar, 0.42% filter sterilized propionic acid). The resulting adults were maintained germ-free for at least three generations to mitigate any parental effects.

Bacteria strains: Five unique species were identified in our laboratory flies (17): *Lactobacillus plantarum*, *L. brevis*, *Acetobacter pasteurianus*, *A. tropicalis*, and *A. orientalis*, which were then isolated from *D. melanogaster* flies in our lab and checked by standard Sanger sequencing the complete 16S region (see PCR for fly bacteria section).
To prepare the inoculum for 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 (20) 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 mated flies 5-7 days post-eclosion were sorted into 10% glucose, 5% active dry yeast medium inoculated with a defined mixture of bacteria. Flies thus treated have been shown to have less variation in physiology and gut morphology (24). Inoculating in the adult phase allowed us to separate developmental differences from effects on adult flies. Each vial contained a total of $5x10^6$ CFUs (50 μL of 10^8 bacteria/mL in 1x PBS). Ten female and ten male flies were transferred into each vial. Gnotobiotic flies were transferred to freshly inoculated medium every 3 days for the duration of the concurrent lifespan-fecundity-development experiment (see following sections).

For the bacterial ecology calculations (Fig. 6), two different treatments were applied after an initial 10 days of inoculation where flies were inoculated as described in 'Adult gnotobiotic fly preparation.' Every three days the remaining live flies were transferred to fresh food vials inoculated with $5x10^6$ CFUs/vial (as with the initial inoculation) in order to reduce the effects of microbial interactions on the fly media. Vial swabs were performed to check that all inoculated species were still present on the third day. In the first treatment (n=24 flies per bacterial treatment), flies were immediately subjected to bacterial load

quantification on the 10th day of inoculation (see 'Bacterial load counts' section). This experiment lets us evaluate the load and relative abundance of bacteria during the fly fitness experiment.

A second treatment was undertaken to measure the steady state of bacterial population sizes in the absence of new colonization. After the initial 10 days of inoculation, these flies were transferred daily to fresh germ-free food for 5 days before subsequent bacterial load quantification (Fig S6C,D).

Check for contamination and correct colonization: All fly work including media preparation and transfers to fresh food was performed in a tissue culture hood using sterile technique. Contamination was assessed by two methods. First, groups of 10 flies were crushed and plated to determine whether they were colonized. Correct colonization was determined by colony morphology on MRS and MYPL media and by 16S PCR followed by Sanger sequencing to confirm species identities. Whole fly DNA extracts were also checked by PCR using both 16S and Wolbachia-specific primers. We perform these tests every two weeks to maintain our gnotobiotic flies. During the fitness experiment, the correct colonization was checked by swabbing the old vials after adults were transferred to fresh media.

PCR for fly bacteria

We used PCR to test for correct bacterial association and to ensure that our flies remained *Wolbachia*-free. 16S universal primers to the V4 region of the rRNA gene were used to check for proper bacterial association (16S V4 Forward: 5'- GTG TGC CAG CMG CCG CGG TAA; 16S V4 Reverse: 5'- CCG GAC TAC HVG GGT WTC TAA T). PCR reaction mix and cycling parameters were as follows:
KAPA2G Robust HotStart Kit, 15uL reaction: 3uL 5X KAPA2G Buffer B; 0.3uL dNTP mix; 0.12uL KAPA2G; Robust HotStart DNA Polymerase ; 0.5uL 16S V4 Forward primer; 0.5uL 16S V4 Reverse primer ; 1uL template DNA ; 9.58uL dH_2O.
Cycling Conditions: Initial denaturation: 98C – 45 seconds; 36 cycles: 98C – 15 s, 58C – 15 s, 72C – 15 s; Final Extension: 72C – 5 min; Hold at 4C.
Wolbachia-specific primers were used to check for infection every month (Wsp 81F: 5'-TGG TCC AAT AAG TGA TGA AGA AAC; Wsp 691R: 5'-AAA AAT TAA ACG CTA CTC CA). The same reaction mix and cycling parameters were used with the exception that denaturation, annealing, and polymerization steps were all extended to 1 minute each.
Standard Sanger sequencing was performed to validate contamination results.

Concurrent lifespan assay, fecundity (pupae counts), fly development: 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 as 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. In tests, we found that greater than 99% of pupae eclosed into adults and therefore used pupae counts as a proxy for adults. Total fecundity was calculated as the sum of all daily fecundity counts for a given vial of adult flies. 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: In the experiments presented in Fig. S10, 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 facilitate 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. S10 black points), development timing was assessed at ~3-h intervals.

Bacterial load counts from flies: To assess the number of bacterial CFUs per fly (Figs. 3B, S6), flies were shaken 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).

Bacterial load counts from food: To assess the bacterial load in fly food (Fig S9), a similar protocol was followed as for the whole flies. A 96-well plate containing 100 μL PBS and 80 μL 0.5 mm glass beads (Biospec) was prepared. This plate was placed on an analytical balance. A small metal spatula was then dipped into the fly food to gather ~10 mg of food. The food was scraped into a well of the plate and weighed. The spatula was sterilized in 70% ethanol and a flame between samples. Three samples were taken from separate regions of each fly vial and two separate biological replicates were made for each bacterial treatment. The plate was then heat sealed, bead beaten for 60 s, replica pinned onto selective media, and scored (as was done in making the bacterial counts from flies).

***In vitro* passaging assay:** Each bacterial combination was made in 1x PBS and 5 µL of 10^7 CFUs/mL was inoculated in triplicate into three rich media in 96 well plates: MRS, MYPL, and YG. YG is 50% diluted fly food without agar: 5% glucose, 2.5% boiled baker's yeast, and 0.21% proprionic acid. The yeast sediment has been removed by centrifugation. Culture volume was 150 µL per well. The cultures were allowed to grow for 48 hours at 25˚C under constant shaking. Cells were then passaged to fresh media in 96 well plates by diluting each well 10 fold and then replica pinning (Boekel 96 pin tool) to the fresh plate, which delivers ~2µL. A total of three passages were made, and the final passage was replica pinned onto selective agar, allowing discrimination of the 5 species' presence/absence. Selective agar were the following: MRS + Xgal grows *Lp* and *Lb* and *Lb* turns blue while *Lp* is yellowish white. MYLP + 10 µg/mL gentamycin grows *At* and *Ao*. MYPL + 5 µg/mL tetracycline grows *Ap* and *Ao*. *Ao*'s colony morphology is distinctive and can be distinguished by eye.

Fly activity assay: 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 9^{th} day (the third flip), gnotobiotic flies were flipped into a vial containing sterile gnotobiotic fly 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: We estimated fitness for each bacterial treatment using a Leslie matrix (1,000 replicates per treatment). For each replicate Leslie matrix, we randomly sampled from the experimental replicates of development time, daily 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 '1's 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) (28) 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(45) 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. S1). Model selection using the Akaike information criterion was applied to pick the best function. The same approach was applied to fit the daily fecundity data (Fig. S5), 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.

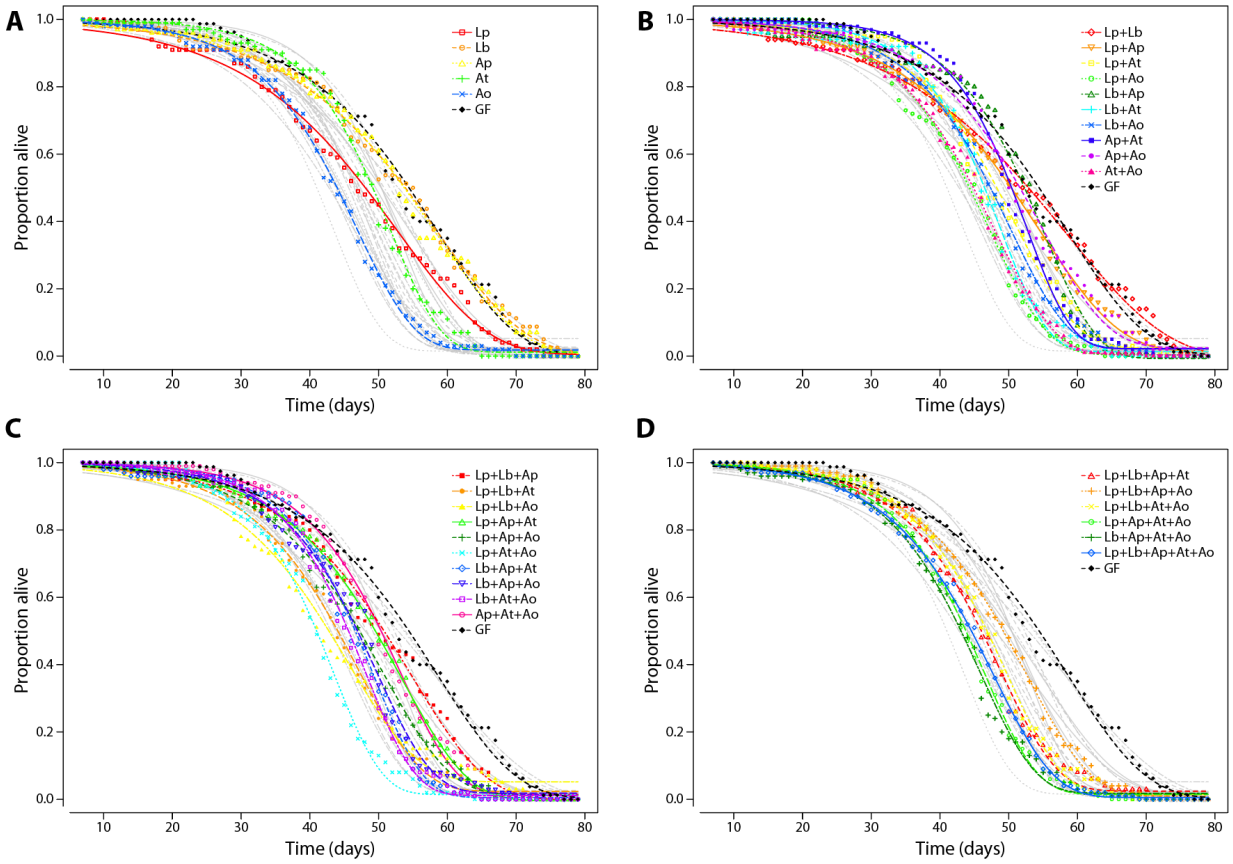


Figure S1. Curve fits to raw lifespan data aggregated from all 5 experimental replicates for each bacterial combination. Curve fits to a 3-parameter Gompertz distribution with an upper limit at one are depicted (see Methods). Bacterial combinations are grouped by the number of species. (**A**) Single 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 curves are displayed in grayscale as a reference.

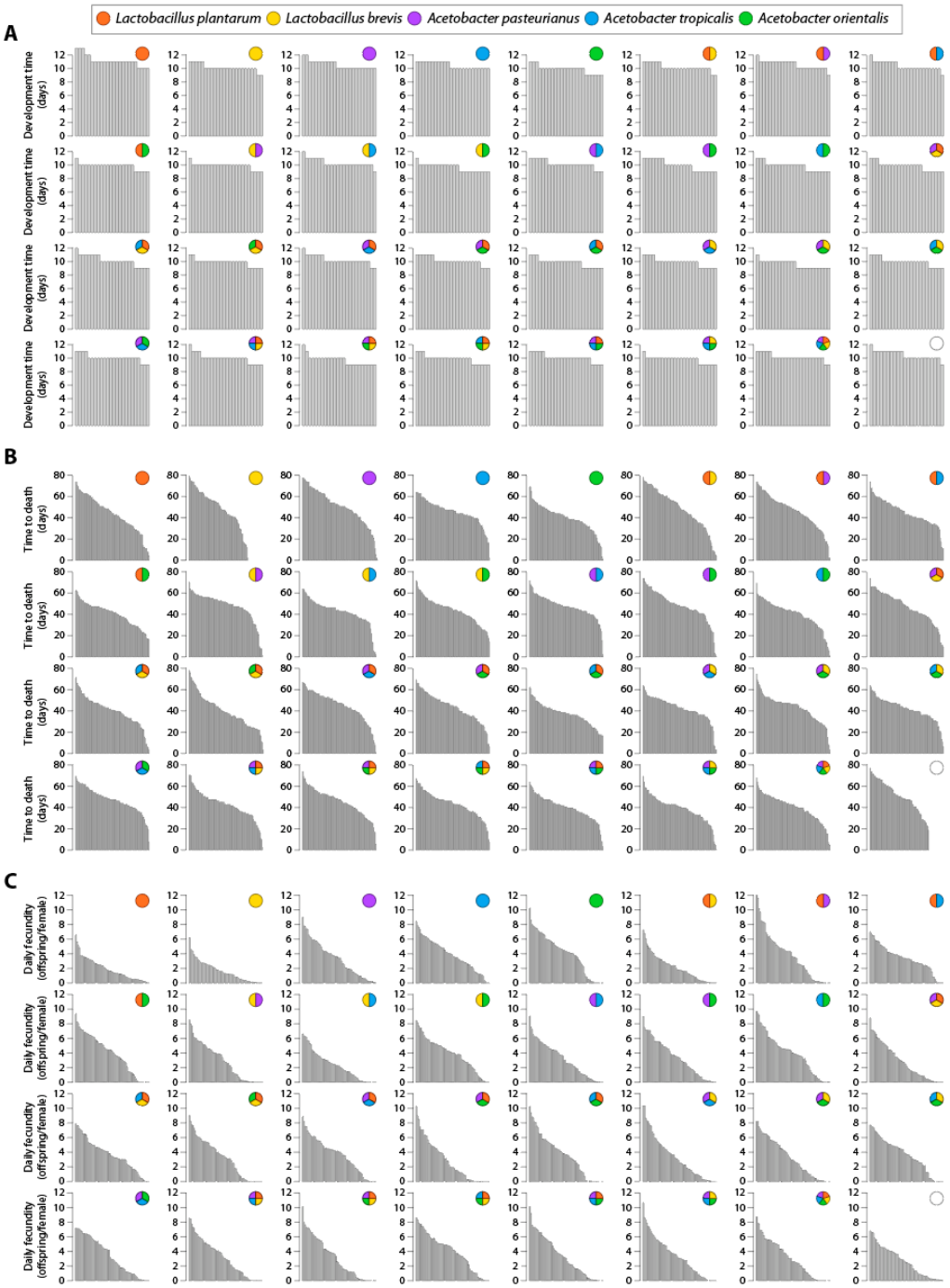


Figure S2. Raw data from development rate, fecundity, and time to death. (**a**) Raw data for development time by microbial treatment. Each bar within a treatment is the fastest-developing fly within a vial. (**b**) Raw data for time to death by microbial treatment. Each bar represents the lifespan of 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.

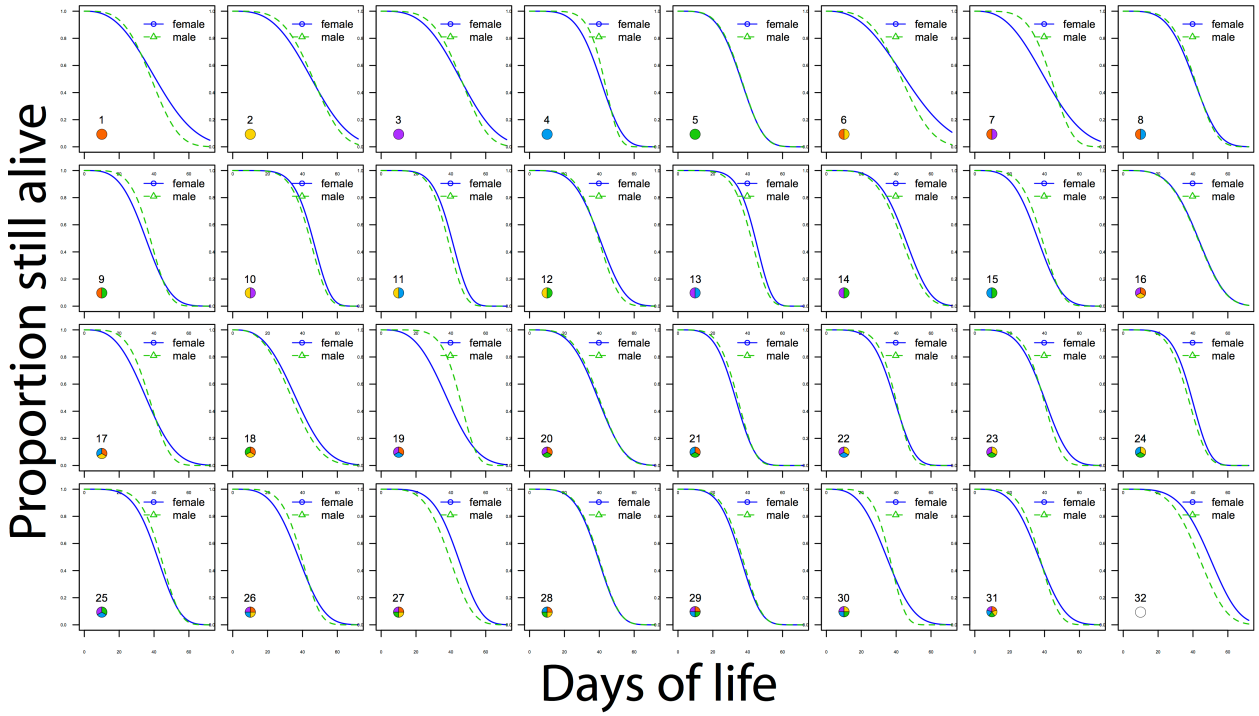


Figure S3. Lifespan differences between males and females are not significant overall. For each bacterial combination, a separate Gompertz curve (see Methods) was fit to male and female survival. No consistent difference is apparent overall, however, individual combinations may have differences if they were evaluated separately from the group.

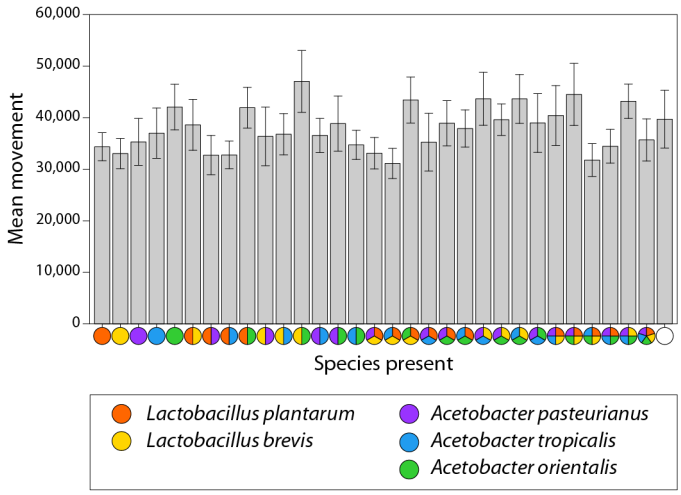


Figure S4. Average fly activity is unrelated to the fitness phenotypes. Fly movement is associated with overall metabolism, including food intake and energy expenditure. To search for behavior changes underlying the physiological differences in our bacterial treatments, we examined changes in fly motility 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.

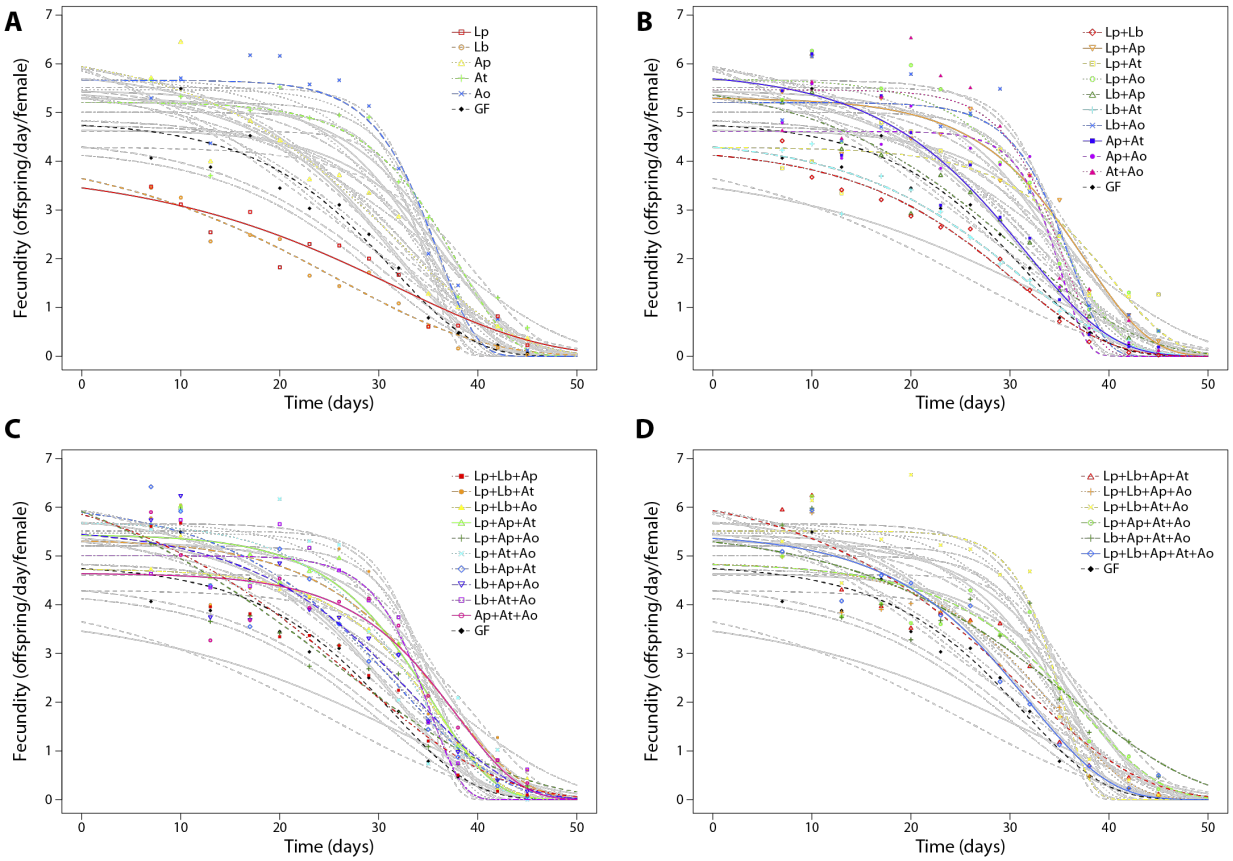


Figure S5. Curve fits to raw fecundity data aggregated from all 5 experimental replicates for each bacterial combination. Curve fits to a 3-parameter Gompertz distribution are depicted (see Methods). Bacterial combinations are grouped by the number of species. (**A**) Single species and germ-free flies. (**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.

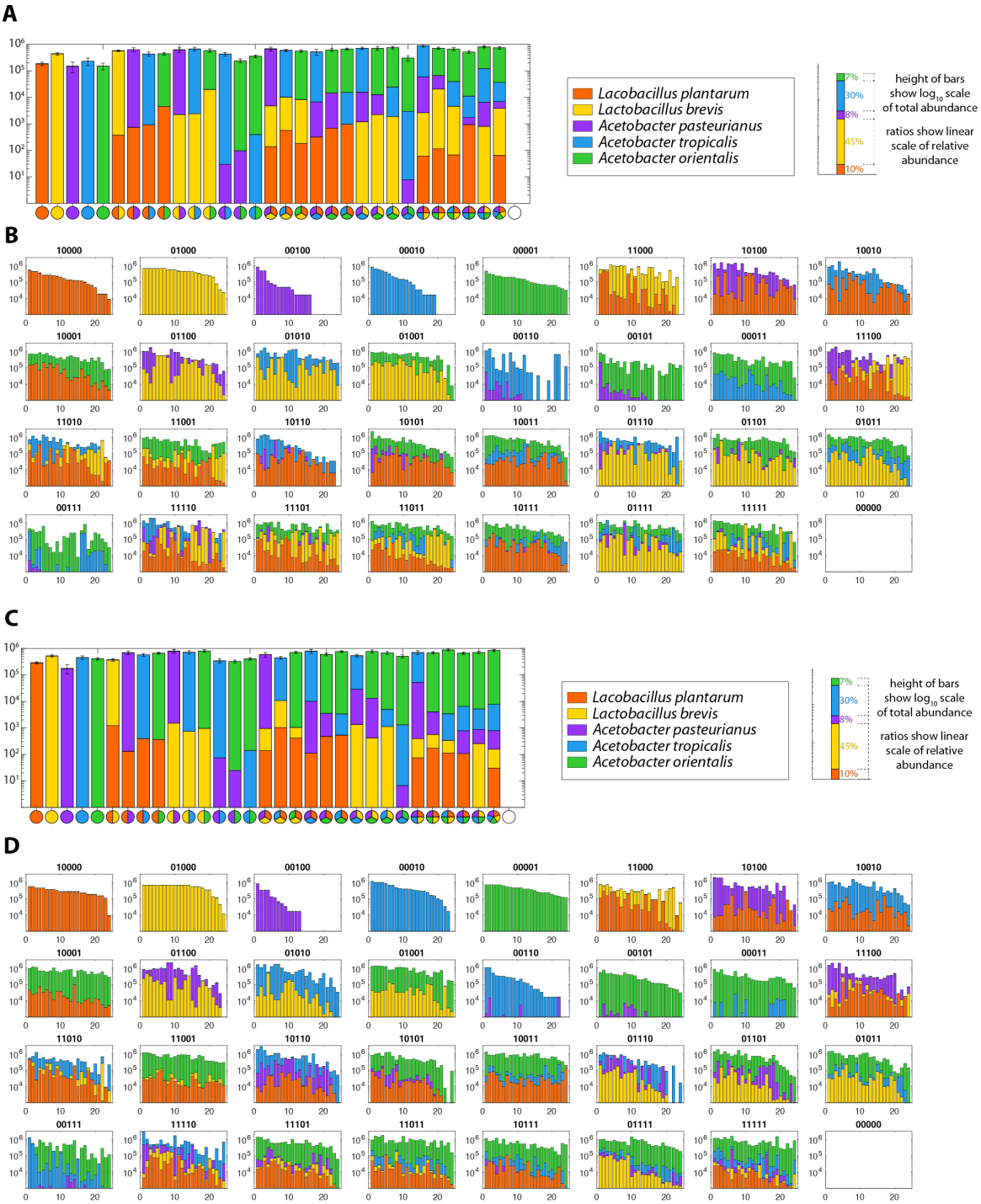
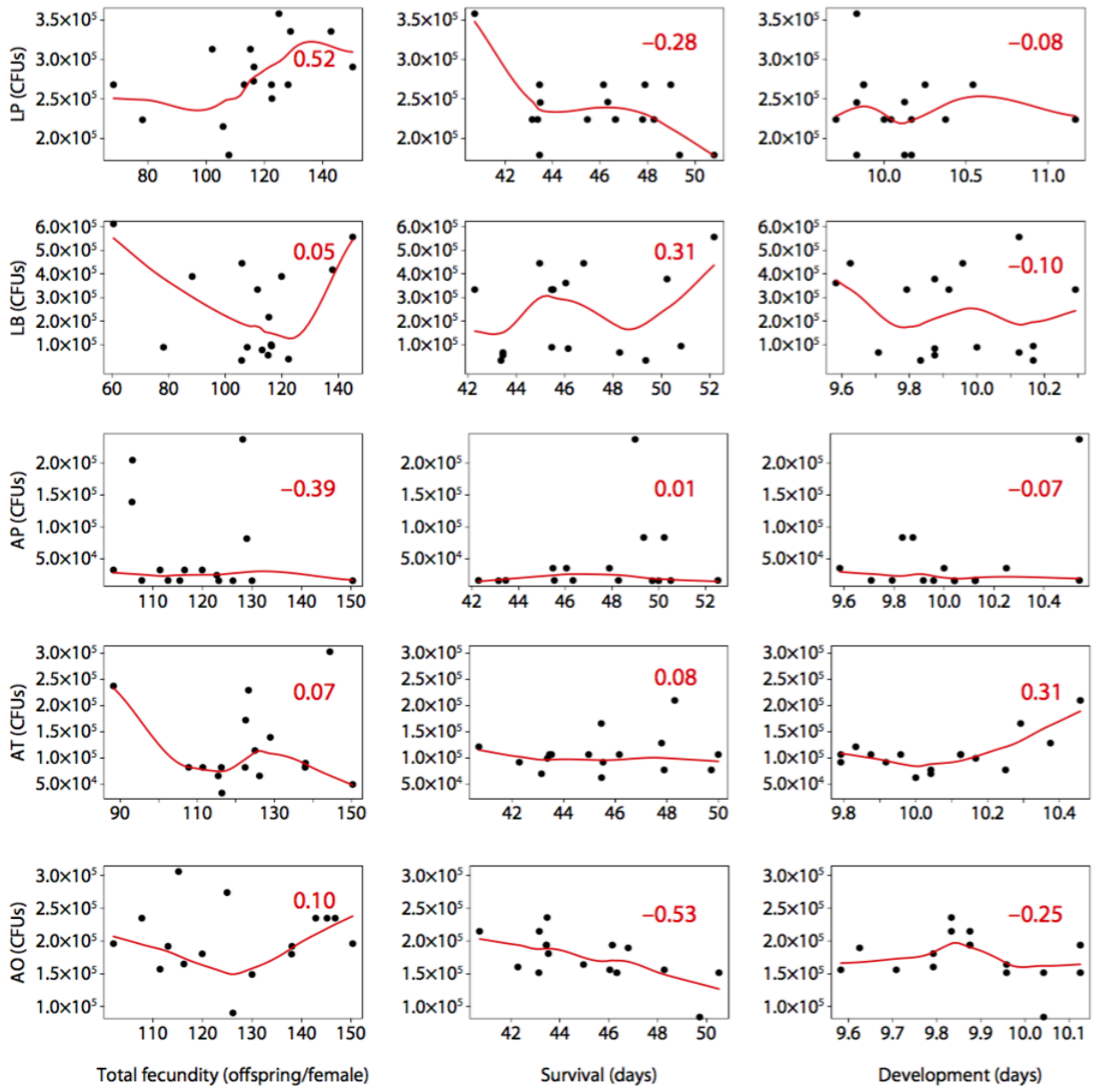


Figure S6. Raw bacterial abundance counts (CFUs) for each fly with each bacterial combination. Y-axes are CFUs on a $\log_{10}$ scale. The relative abundance of individual species is indicated by ratios on a linear scale. (**A**) Average CFUs for each bacterial combination where flies were fed defined bacteria continuously for 10 days and then crushed and CFUs enumerated. (**B**) Individual fly data for the same experiment represented in **A**. X-axes indicate the 24 individual flies. (**C**) Average CFUs for each bacterial combination. The experiment is identical to **A** with one difference: after the initial 10 days of inoculation, flies were transferred daily to fresh food for 5 days. Note that there are subtle differences between **A** and **C**. (**D**) Individual fly data for the same experiment represented in **C**. X-axes indicate the 24 individual flies.

We note that the limit of detection (~1,000-10,000 CFUs) can mask low-abundance colonization. We performed additional CFU enumeration of individual bacterial species in individual flies with a limit of detection of 10 CFUs. This experiment showed that flies which appeared uncolonized by one or more bacterial species, were likely colonized at levels below the limit of detection after five days of daily transfer to germ free food (12/12 flies colonized with 5 species had all 5 species; 12/12 flies colonized with *Ap* had *Ap*; 12/12 flies colonized with *Ap* + *Ao* had both species; 11/12 flies colonized with *Ap* + *At* were colonized by both species while one was missing *Ap*). These results are consistent with our previous results (17) and indicate that flies are stably colonized in our experimental conditions (Fig. 3B).



CFUs species fitness correlations plots

	Lp		Lb		Ap		At		Ao	
	r	p	r	p	r	p	r	p	r	p
TotalFecundity	**0.52**	**0.04**	0.05	0.86	-0.39	0.13	0.07	0.80	0.10	0.71
TimeDeath	-0.28	0.30	0.31	0.25	0.01	0.96	0.08	0.77	**-0.53**	**0.03**
Develop	-0.08	0.78	-0.10	0.71	-0.07	0.80	0.31	0.24	-0.25	0.35
C.V. TotalFecund	-0.21	0.42	-0.18	0.51	-0.22	0.41	0.31	0.24	0.47	0.07
C.V. TimeDeath	0.46	0.08	0.31	0.25	0.14	0.59	0.08	0.77	-0.43	0.10
C.V. Develop	0.14	0.60	0.35	0.19	0.14	0.61	0.34	0.20	0.08	0.77

Figure S7. Fly phenotype correlations with individual bacterial species abundances. Phenotype data from Fig. 2. CFU data from Fig. S6. P-values adjusted for multiple comparisons using Tukey's correction. Table shows Spearman correlations and p-values or each comparison. Coefficient of variation (C.V.) correlations are plotted in Fig. S8.

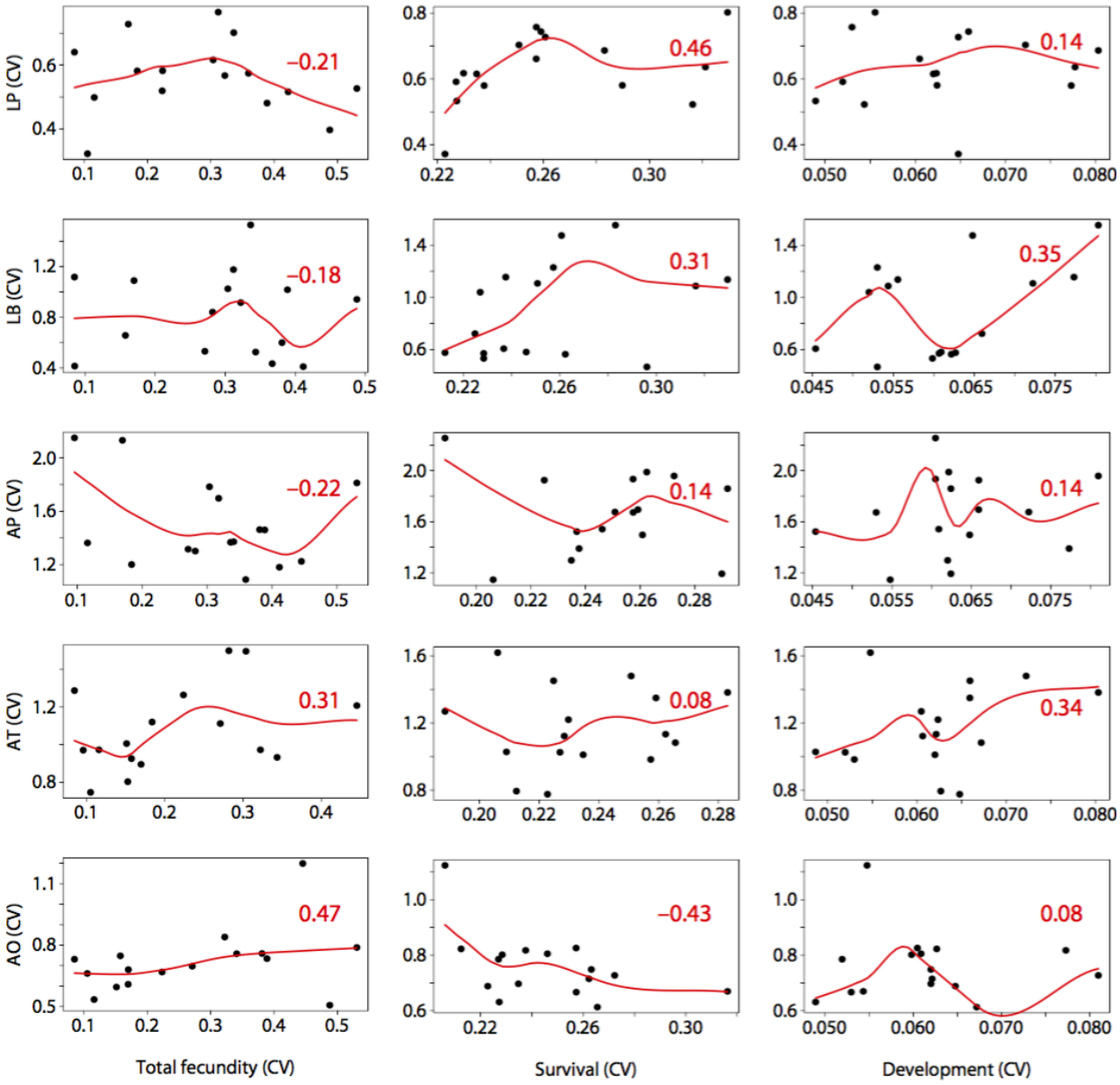


Figure S8. Correlation plots between the coefficient of variation for fly phenotypes and individual bacterial species abundances. None of the correlations are statistically significant (see Table in Fig. S7).

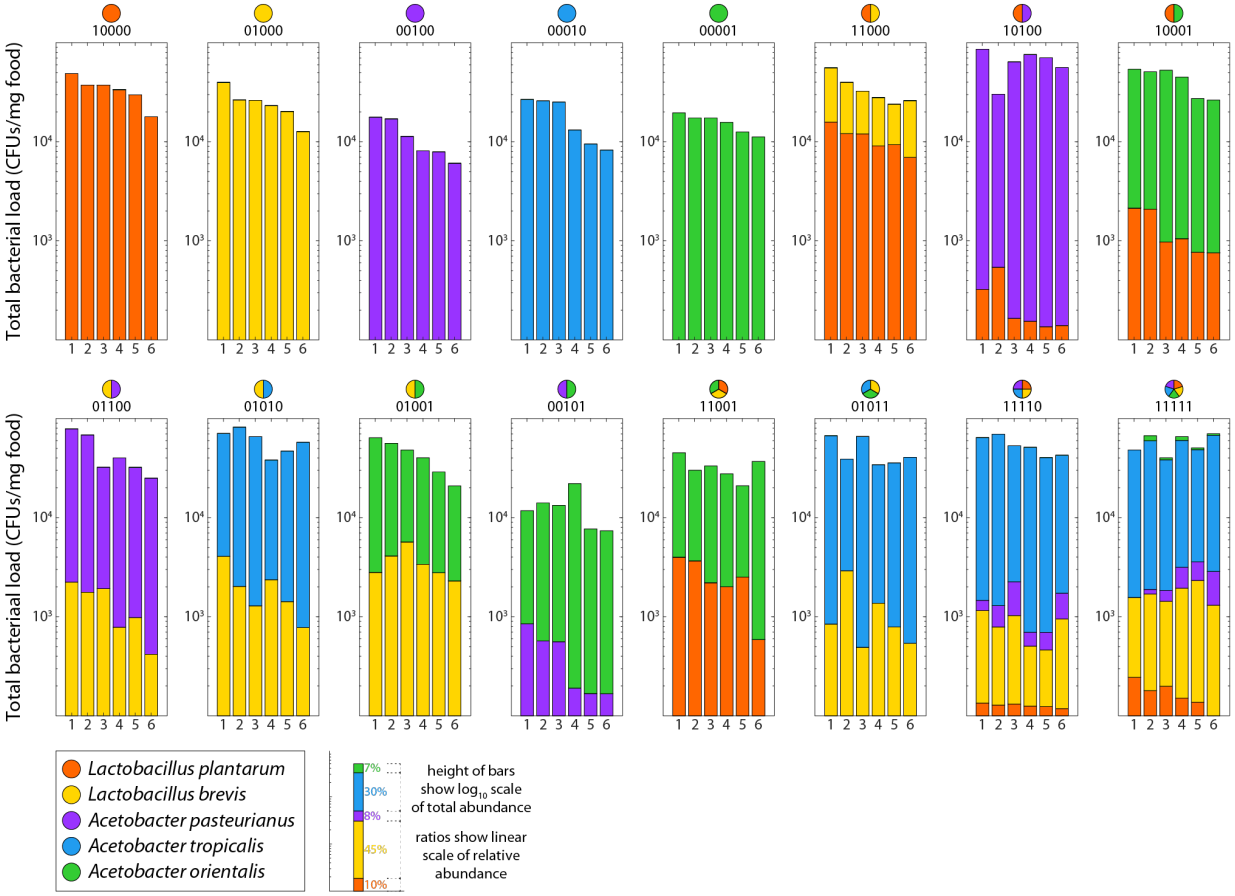


Figure S9. Raw bacterial abundance counts (CFUs) for fly food treatments with 16 selected bacterial combinations. X-axes indicate individual food samples 1 to 6. For each combination, the first three samples are from the first biological replicate, and the last three are from the second biological replicate. Y-axes are total CFUs on a $\log_{10}$ scale. The relative abundance of individual species is indicated by ratios on a linear scale.

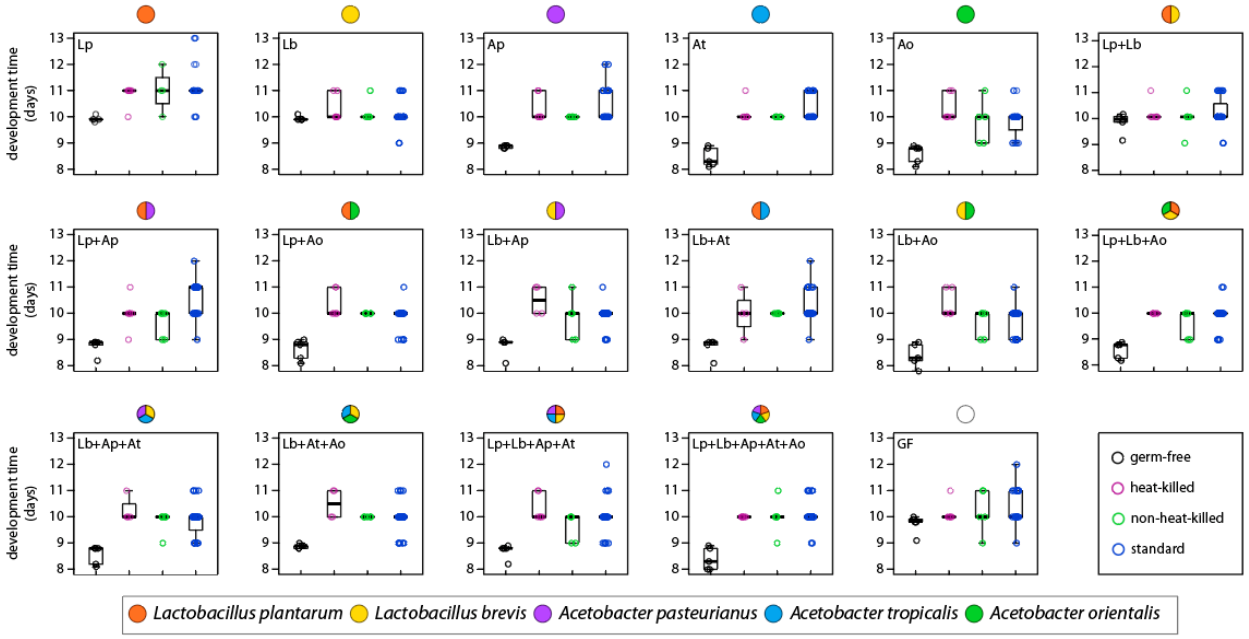


Figure S10. Parental effects and live bacteria influence offspring developmental pace. Sixteen microbial combinations and germ-free flies (indicated in upper left corners of boxes) were tested for their impacts on development time (number of days from embryo laid to adult emergence from pupal case). In the original experiments, the developmental time was measured from embryos that were directly laid by females continuously inoculated with their bacterial combination in the fitness experiment (blue data points in this figure; see also Fig. 1A, 2C). To test the role of parental effects, we experimentally varied the source of the embryos as well as the bacterial treatment. For all new treatments here (black, purple, and green data points), embryos were collected and dechorionated using 10% bleach, washed in 70% ethanol, and pipetted onto food in PBS with 0.1% Triton X detergent to prevent eggs sticking to one another (see Methods). Black points show data for n=20 embryos taken from germ-free mothers before placing in fresh inoculated vials. Green 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 in development between the original fitness experiment and this treatment (paired sample t-test, $p>0.18$, n=500). Purple points indicate development in vials that were heat-killed at 60˚C for 1 hour in a humid chamber to prevent drying (and tested for sterility) prior to the introduction of germ-free eggs. This treatment significantly increased the development time by 8 hours (paired sample t-test, $p<0.005$, n=178; see main text Fig. 3F). Finally, we tested whether bacteria deplete the food. Indeed, the fastest development times were for eggs introduced to fresh vials inoculated with bacteria but without previous fly occupation. In this treatment (black dots), there was very little variation between treatments except that flies lacking all *Acetobacter* species (Lp, Lb, Lp+Lb, germ-free) were delayed by 1 to 2 days with respect to their cohort, indicating that *Acetobacters* improve development most on fresh food. 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. The time resolution for these experiments is 1 day.

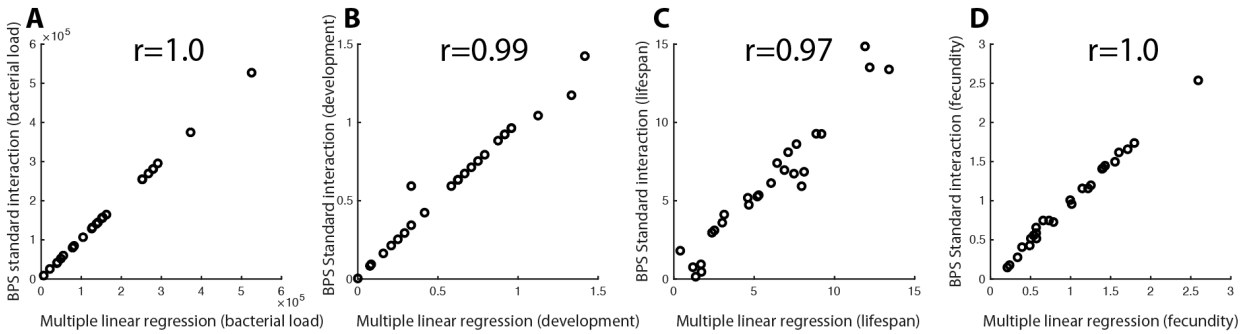


Figure S11. The BPS direct calculations of 'standard interactions' are highly correlated with multivariate linear regression least squares fitting of interaction coefficients. (A) CFUs in units of colony forming units (r^2=0.99, $p<10^{-30}$, n=26), (B) Development time in units of days (r^2=0.97, $p<10^{-19}$, n=26), (C) Lifespan in units of days (r^2=0.93, $p<10^{-14}$, n=26), and (D) Fecundity in units of viable offspring per female per day (r^2=0.99, $p<10^{-26}$, n=26).

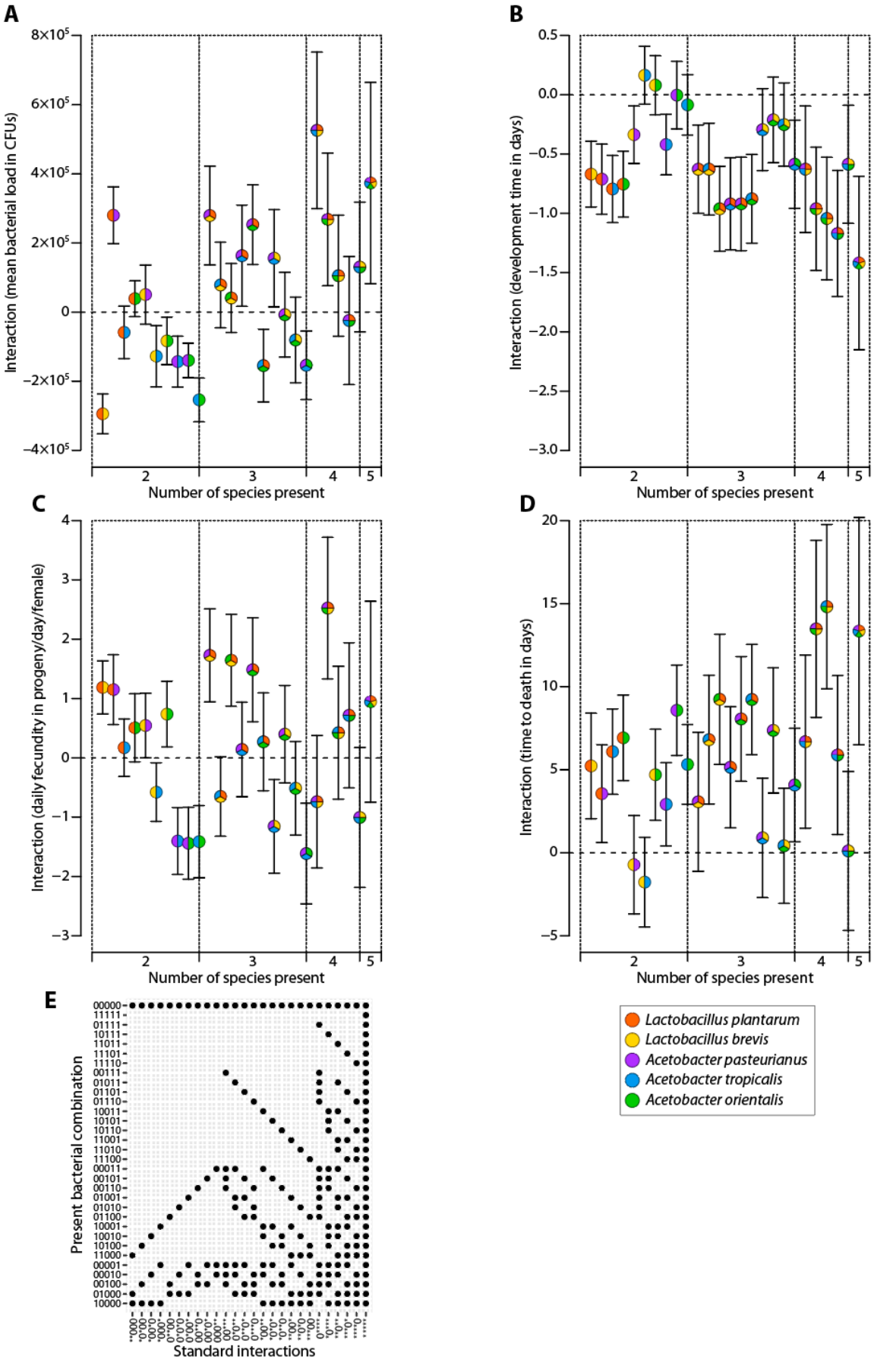
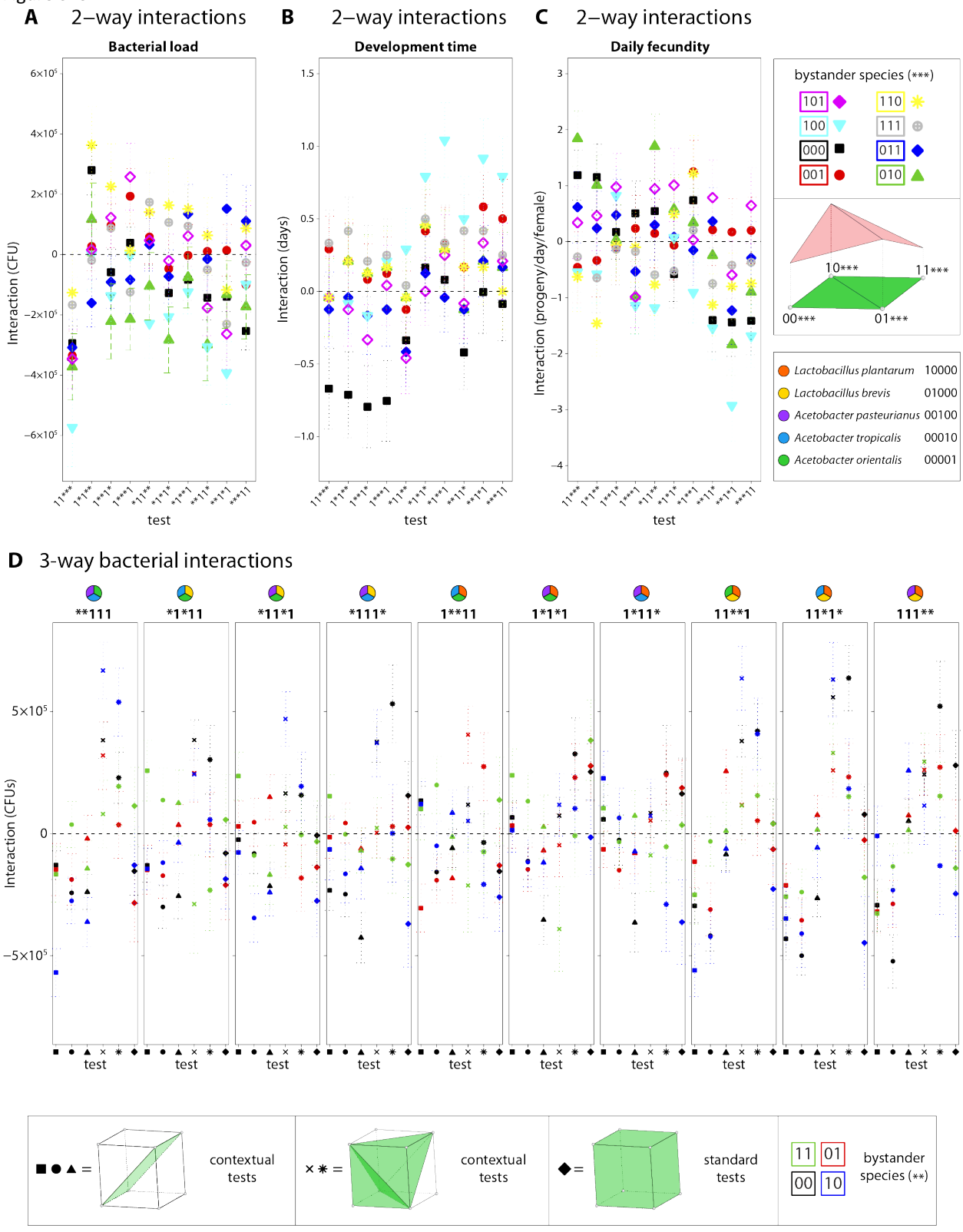

A
Interaction (mean bacterial load in CFUs)
8×10^5
6×10^5
4×10^5
2×10^5
0
−2×10^5
−4×10^5
2
3
4
5
Number of species present
B
Interaction (development time in days)
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5
−3.0
Number of species present
C
Interaction (daily fecundity in progeny/day/female)
4
3
2
1
0
−1
−2
−3
Number of species present
D
Interaction (time to death in days)
20
15
10
5
0
−5
Number of species present
Lactobacillus plantarum
Lactobacillus brevis
Acetobacter pasteurianus
Acetobacter tropicalis
Acetobacter orientalis
E
Present bacterial combination
Standard interactions

Figure S12. Standard interactions calculated for each phenotype in Fig. 2A,C,E,G. For all four phenotypes ((**A**) bacterial load, (**B**) development time, (**C**) fecundity, (**D**) lifespan) 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. (**E**) The combinations summed together to calculate the interactions are indicated by black dots. For instance, the interaction between LP and LB, indicated by **000 is calculated by combining the phenotype scores for 10000, 01000, 11000, and 00000. The 5-way interaction involves all combinations. 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 Fig. S13). See also T-polynomials (Math Supplement). Error bars are the propagated standard error from the raw phenotypes.

Figure S13


A 2–way interactions
B 2–way interactions
C 2–way interactions
Bacterial load
Development time
Daily fecundity
Interaction (CFU)
Interaction (days)
Interaction (progeny/day/female)
test
bystander species (***)
101
110
100
111
000
011
001
010
10***
11***
00***
01***
Lactobacillus plantarum 10000
Lactobacillus brevis 01000
Acetobacter pasteurianus 00100
Acetobacter tropicalis 00010
Acetobacter orientalis 00001
D 3-way bacterial interactions
**111
*1*11
*11*1
111
1**11
1*1*1
1*11*
11**1
11*1*
111**
Interaction (CFUs)
contextual tests
contextual tests
standard tests
bystander species (**)
11
01
00
10

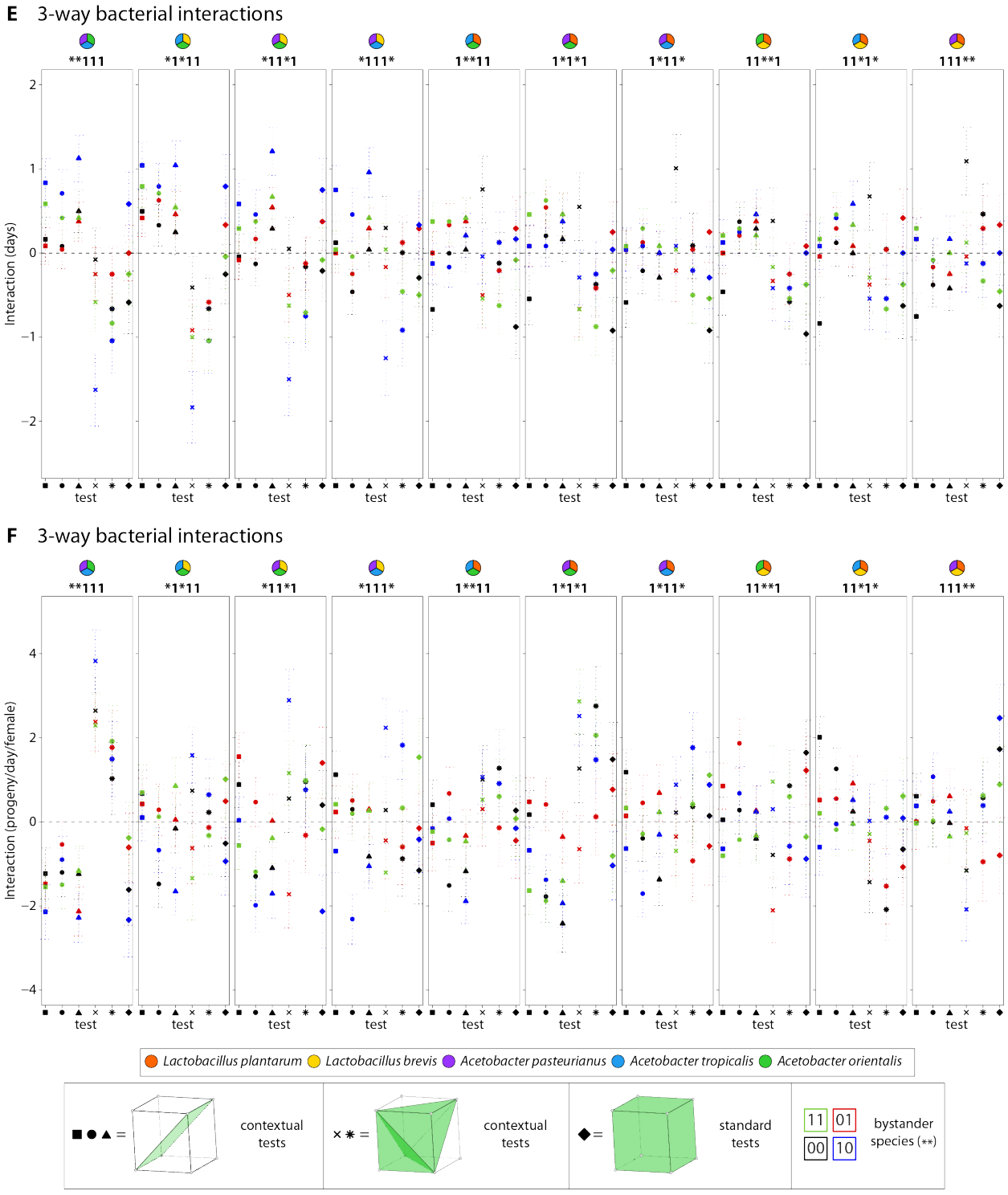


Figure S13. Detailed comparisons of the context-dependence of two-way and three-way interactions depending on bystander species. The pairwise interaction was calculated between each pair of species for each set of possible bystander species as in Fig. 5A for (**A**) bacterial load, (**B**) development time, and (**C**) fecundity. The standard three-way interaction as in Fig. 5B was compared with the related contextual tests as a function of bystander species present for (**D**) bacterial load, (**E**) development time, and (**F**) fecundity**.** 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 contextual tests can describe cases of non-additive standard tests and (ii) whether context of other species changes interactions (see Main Text).

Relevant equations (from Math Supplement):

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

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

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

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

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

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

Figure S14

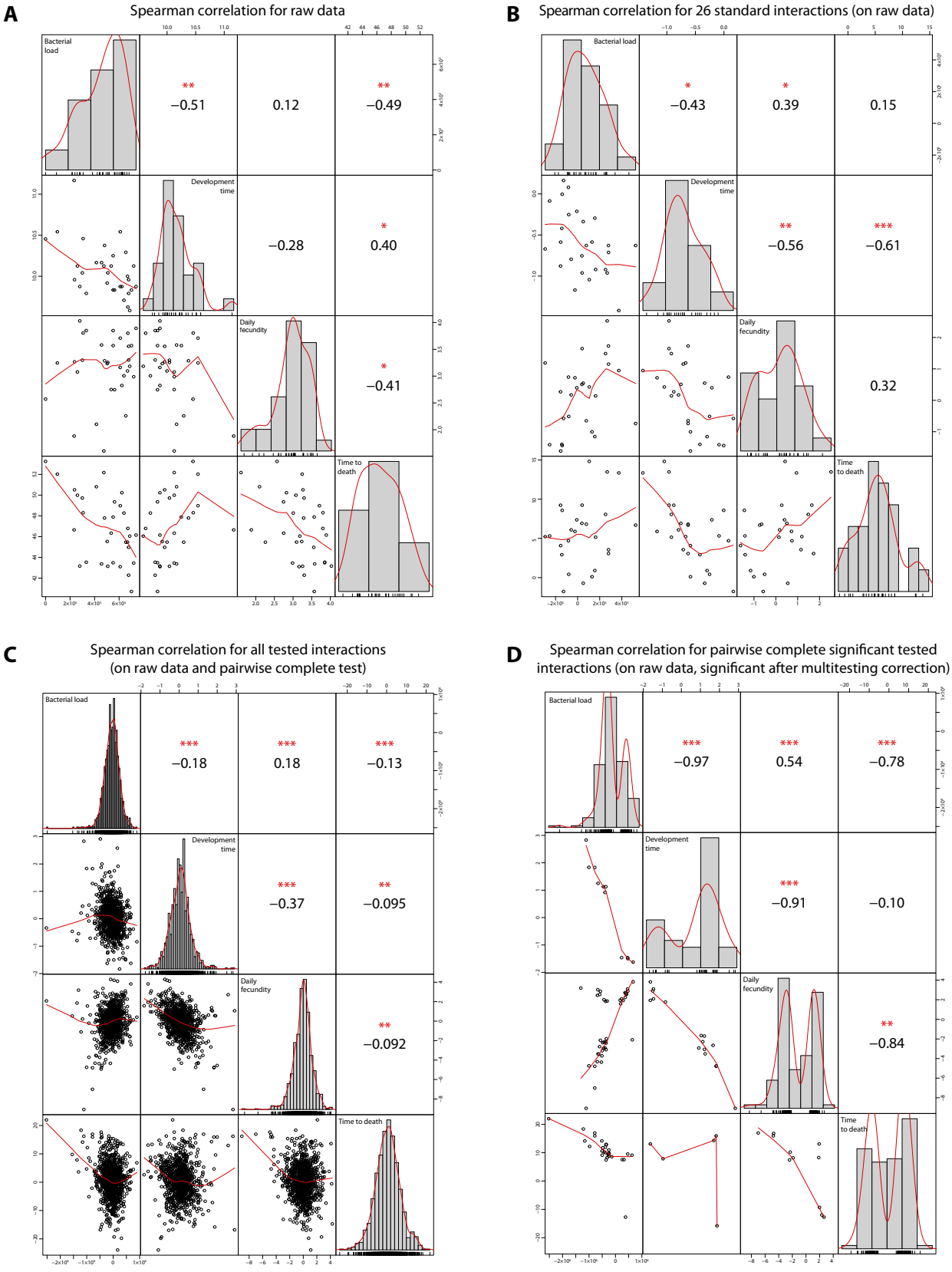

A
Spearman correlation for raw data
Bacterial load
Development time
Daily fecundity
Time to death
** −0.51
0.12
** −0.49
−0.28
* 0.40
* −0.41
B
Spearman correlation for 26 standard interactions (on raw data)
* −0.43
* 0.39
0.15
** −0.56
*** −0.61
0.32
C
Spearman correlation for all tested interactions
(on raw data and pairwise complete test)
*** −0.18
*** 0.18
*** −0.13
*** −0.37
** −0.095
** −0.092
D
Spearman correlation for pairwise complete significant tested
interactions (on raw data, significant after multitesting correction)
*** −0.97
*** 0.54
*** −0.78
*** −0.91
−0.10
** −0.84

Figure S14. 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 contextual 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 contextual 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 on histograms = kernel density function with Gaussian kernel). Correlations are all Spearman coefficients (data were not all normally distributed by Shapiro-Wilk tests, see results in Math Supplement). Significance values (*$p<0.005$; **$p<0.001$; ***$p<0.0001$) appear above the diagonal.

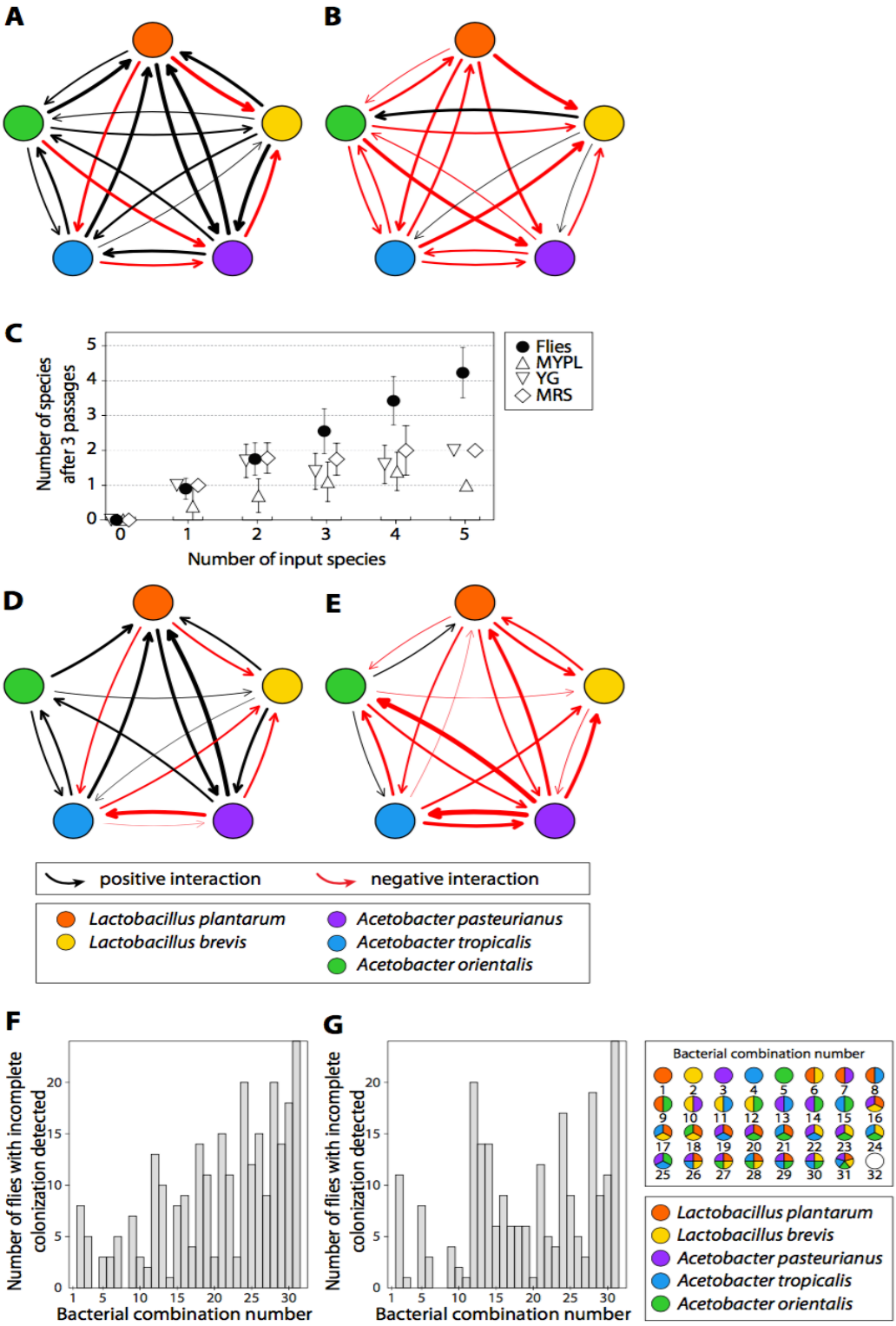


Figure S15. Pairwise bacterial interactions in the fly gut transition from positive to negative as diversity increases. (**A**) Interactions calculated by Paine's (33) method for mean CFU abundance data from bacterial combinations with one or two species. (**B**) Interactions calculated by Paine's (33) method for mean CFU abundance data from bacterial combinations with four or five species. (**C**) We were not able to maintain the 5 species

community in three rich growth liquid media formulated to (i) mimic the fly media (YG), (ii) optimize *Lactobacillus* growth (MRS), (iii) optimize *Acetobacter* growth (MYPL). In agreement with previous reports (32), we found that *in vitro* culture supported only low diversity, with a maximum of two species coexisting in the three rich media types, with the exception of one three member community. (**D**) Interactions calculated using the Lotka-Volterra fitting approach with all individual fly CFU abundance data from bacterial combinations with one or two species. (**E**) Interactions calculated using the Lotka-Volterra fitting approach with 3, 4 or 5 species combinations. (**F**) 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. (**G**) 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. For Fig. 6 and S15A,B, flies were eliminated from the analysis if not all inoculated bacterial species were detected.

Figure S16

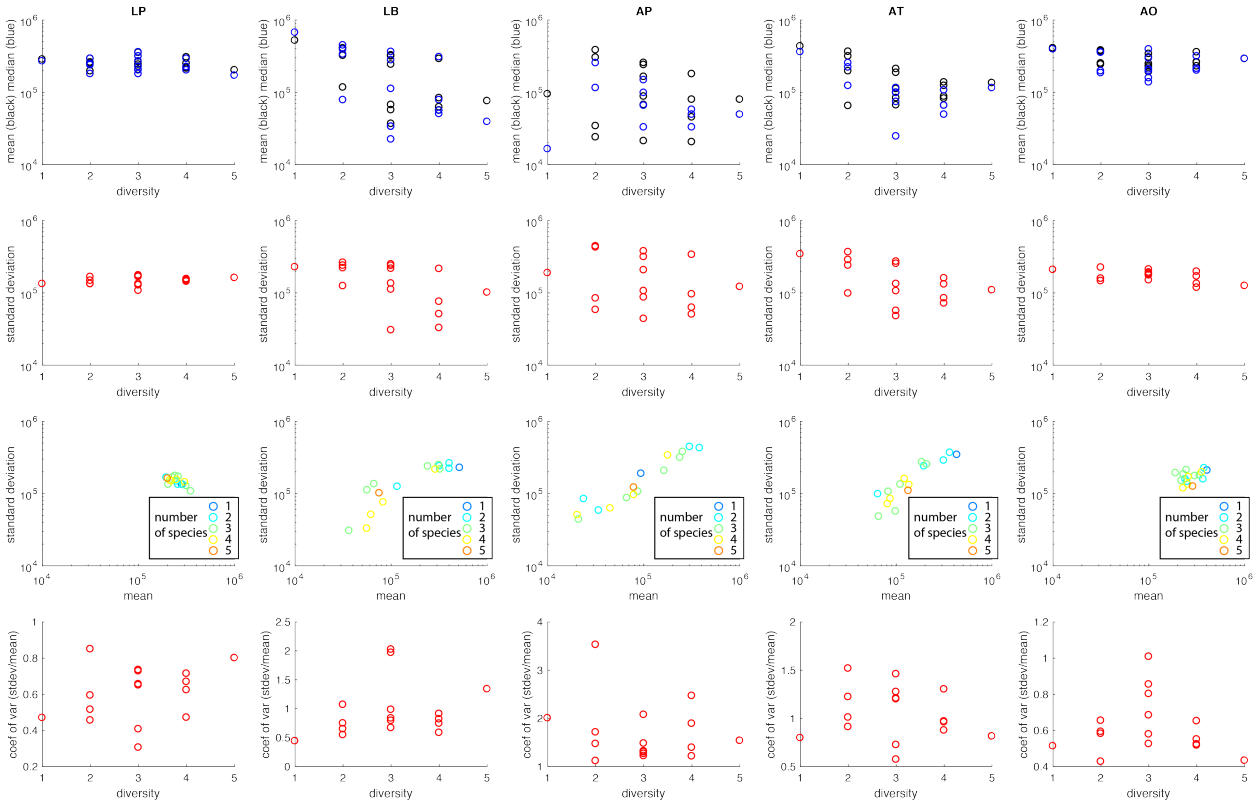


Figure S16: Variation in bacterial load decreases for increasing diversity.
Top row: We calculated the mean (black) and median (blue) bacterial load for each species as a function of increasing number of species, observing a decrease in abundance for increasing diversity. Second row: the Standard deviation also decreases for Lb, Ap, At, and Ao. Third row: the mean and standard deviation were correlated for Lb, Ap, and At. They were anti-correlated for Lp, and not significantly correlated for Ao. Fourth row: the coefficient of variation was roughly constant for each species. However, for the complete microbiome, the coefficient of variation decreased

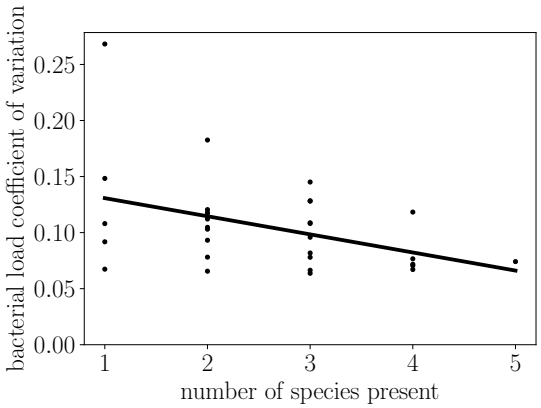


Figure S17. Bacterial load coefficient of variation decreases with increasing diversity. We calculated the coefficient of variation in the total bacterial load as a function of increasing gut diversity, finding a decrease (p=0.02, Wald test; Math Supplement Section 10.5).

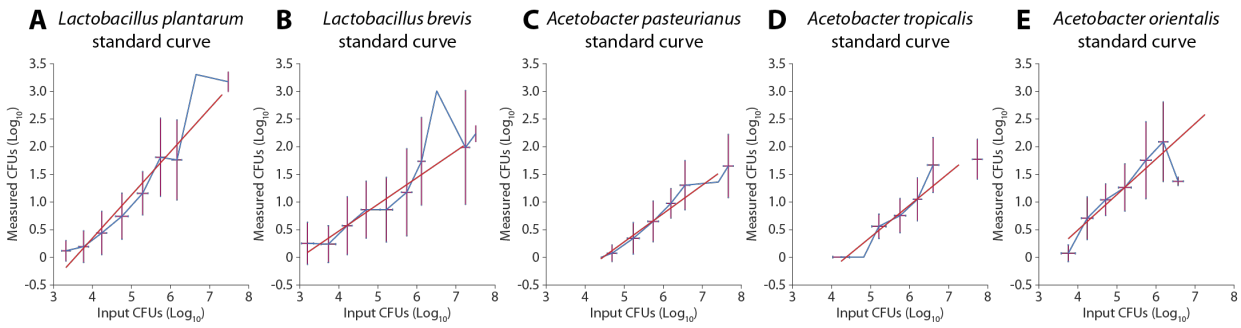


Figure S18. Standard curves used to calculate total CFUs from colony counts (Fig. S4, S6, S9). A standard curve was constructed by plating known concentrations of bacteria on selective medium with a 96-pin replicator as in *Bacterial load counts* (Methods). Counts were fit to a power law curve using the NLINFIT function in MATLAB v2017a.

Table S1. Data means and standard errors (SE) from Fig. 2 that were used to calculate interactions in Figures 2,4,5. Binary ID corresponds to Fig 1A and Figure 4A.

Binary ID	Treatment	Daily Fecundity (eggs/female/day)	Daily Fecundity (SE)	Daily Fecundity (N)	Total Fecundity	Total Fecundity (SE)	Time to Death (days)	Time to Death (SE)	Time to Death (N)	Development Time (days)	Development Time (SE)	Development Time (N)	Mean Bacterial Load (CFUs)	Bacterial Load (SE)	Bacterial Load (N)
00000	32	2.570	0.282	52	130.43	5.2141	53.25	1.45	80	10.455	0.143	22	0	0	48
10000	1	1.881	0.183	65	96.463	9.5197	46.65	1.50	100	11.167	0.177	24	233245.42	21412.15	48
01000	2	1.603	0.194	52	85.656	15.41	52.18	1.73	80	10.125	0.110	24	475712.50	32052.66	48
00100	3	3.246	0.293	65	165.46	6.8996	52.02	1.51	100	10.542	0.134	24	96439.58	28008.52	48
00010	4	3.803	0.266	65	181.43	2.9718	48.30	1.01	100	10.458	0.104	24	306531.25	45831.30	48
00001	5	4.030	0.334	65	191.63	2.4205	43.16	1.14	100	9.875	0.125	24	279266.67	30165.37	48
11000	6	2.102	0.222	65	108.59	5.4624	50.81	1.67	100	10.167	0.115	24	414921.67	42826.24	48
10100	7	3.708	0.385	65	180.15	3.2493	48.98	1.42	100	10.542	0.134	24	609631.25	74105.91	48
10010	8	3.286	0.224	65	157.07	2.8465	47.79	1.10	100	10.375	0.132	24	481337.50	56720.93	48
10001	9	3.848	0.326	65	184.53	3.5578	43.48	0.99	100	9.833	0.098	24	551514.58	36281.12	48
01100	10	2.825	0.305	65	145.34	7.3124	50.23	1.19	100	9.875	0.092	24	622841.67	74050.39	48
01010	11	2.258	0.236	65	115.47	5.3909	45.46	1.08	100	10.292	0.127	24	654837.50	68718.72	48
01001	12	3.801	0.278	65	178.32	2.4106	46.79	1.07	100	9.625	0.118	24	671834.17	52555.70	48
00110	13	3.077	0.285	65	156.53	4.6458	49.99	0.94	100	10.125	0.125	24	259983.33	50392.53	48
00101	14	3.266	0.301	65	155.63	2.713	50.51	1.31	100	9.958	0.165	24	236139.58	28015.66	48
00011	15	3.850	0.329	65	182.46	2.2346	43.53	1.16	100	9.792	0.134	24	332337.50	31499.85	48
11100	16	2.746	0.296	65	144.93	10.75	49.35	1.29	100	9.833	0.130	24	562772.92	72784.20	48
11010	17	3.580	0.266	65	174.49	3.8842	43.37	1.23	100	10.167	0.167	24	457034.17	44228.94	48
11001	18	3.161	0.322	65	153.97	3.3452	43.10	1.34	100	9.875	0.110	24	609145.83	40513.51	48
10110	19	3.569	0.294	65	173.59	3.5142	47.89	1.24	100	10.250	0.138	24	551583.33	81965.04	48
10101	20	2.749	0.338	65	145.53	15.304	46.33	1.19	100	10.125	0.125	24	535355.00	59542.90	48
10011	21	3.569	0.350	65	174.18	3.4082	40.71	0.91	100	9.833	0.130	24	700770.83	44902.50	48
01110	22	3.233	0.347	65	164.88	5.8153	45.53	1.02	100	9.917	0.133	24	503093.75	55993.51	48
01101	23	3.184	0.313	65	160.32	5.373	46.05	1.13	100	9.583	0.119	24	686791.67	56946.65	48
01011	24	3.556	0.287	65	166.53	2.2059	44.97	0.96	100	9.958	0.127	24	677964.58	53131.35	48
00111	25	3.297	0.305	65	159.62	3.3159	49.72	1.02	100	10.042	0.112	24	299869.58	40955.49	48
11110	26	3.096	0.311	65	159.82	8.3454	45.47	1.14	100	10.000	0.147	24	668435.42	80306.35	48
11101	27	3.006	0.334	65	149.54	4.1132	48.27	1.15	100	9.708	0.153	24	639993.75	45986.21	48
11011	28	3.892	0.310	65	185.49	2.2084	46.15	1.05	100	9.875	0.125	24	738166.67	53267.80	48
10111	29	3.256	0.334	65	159.95	4.3024	43.14	1.01	100	10.042	0.127	24	507893.75	40493.53	48
01111	30	3.301	0.326	65	166.47	8.5605	42.28	1.11	100	9.792	0.104	24	678083.33	46462.03	48
11111	31	2.988	0.306	65	149.56	4.4131	43.45	1.12	100	10.125	0.110	24	719135.42	54260.31	48

Table S2: Multivariate linear regression analysis of fecundity data to calculate interactions

Model	SE (residual)	DF	R^2 (adj.)	F	P	Parameter	Estimate	SE	t
Lp	2.46	2052	-3.5E-04	0.27	6.0E-01	Intercept	3.209	0.077	**41.50**
						Lp	-0.057	0.109	-0.52
Lb	2.46	2052	0.003	6.45	1.1E-02	Intercept	3.318	0.077	**43.23**
						Lb	-0.276	0.109	**-2.54**
Ap	2.46	2052	-4.1E-04	0.17	6.8E-01	Intercept	3.203	0.077	**41.42**
						Ap	-0.044	0.109	-0.41
At	2.46	2052	0.004	9.73	1.8E-03	Intercept	3.009	0.077	**39.02**
						At	0.338	0.108	**3.12**
Ao	2.45	2052	0.009	20.48	6.4E-06	Intercept	2.933	0.077	**38.11**
						Ao	0.489	0.108	**4.53**
Lp*Lb	2.46	2050	0.003	2.73	4.3E-02	Intercept	3.413	0.109	**31.26**
						Lp	-0.188	0.154	-1.22
						Lb	-0.408	0.154	**-2.64**
						Lp:Lb	0.262	0.217	1.21
Lp*Ap	2.46	2050	-0.001	0.16	9.3E-01	Intercept	3.241	0.111	**29.24**
						Lp	-0.075	0.155	-0.49
						Ap	-0.063	0.155	-0.41
						Lp:Ap	0.035	0.217	0.16
Lp*At	2.46	2050	0.004	4.01	7.4E-03	Intercept	3.117	0.111	**28.20**
						Lp	-0.210	0.154	-1.36
						At	0.180	0.154	1.17
						Lp:At	0.311	0.217	1.44
Lp*Ao	2.45	2050	0.010	7.81	3.5E-05	Intercept	2.866	0.110	**26.00**
						Lp	0.130	0.154	0.85
						Ao	0.670	0.154	**4.35**
						Lp:Ao	-0.357	0.216	-1.65
Lb*Ap	2.46	2050	0.002	2.29	7.7E-02	Intercept	3.368	0.109	**30.80**
						Lb	-0.330	0.154	**-2.14**
						Ap	-0.099	0.154	-0.64
						Lb:Ap	0.108	0.217	0.50
Lb*At	2.45	2050	0.007	5.57	8.3E-04	Intercept	3.184	0.109	**29.25**
						Lb	-0.350	0.154	**-2.27**
						At	0.266	0.153	1.74
						Lb:At	0.146	0.217	0.67
Lb*Ao	2.45	2050	0.013	9.87	1.9E-06	Intercept	3.157	0.109	**29.06**
						Lb	-0.449	0.154	**-2.92**
						Ao	0.319	0.153	**2.09**
						Lb:Ao	0.342	0.216	1.58

Ap*At	2.46	2050	0.005	4.65	3.0E-03	Intercept	2.916	0.111	**26.39**
						Ap	0.182	0.154	1.18
						At	0.560	0.154	**3.63**
						Ap:At	-0.438	0.217	**-2.02**
Ap*Ao	2.44	2050	0.021	15.81	3.7E-10	Intercept	2.664	0.110	**24.31**
						Ap	0.523	0.153	**3.42**
						Ao	1.050	0.153	**6.86**
						Ap:Ao	-1.108	0.215	**-5.15**
At*Ao	2.44	2050	0.017	12.51	4.2E-08	Intercept	2.611	0.110	**23.77**
						At	0.627	0.153	**4.08**
						Ao	0.775	0.153	**5.06**
						At:Ao	-0.556	0.216	**-2.58**
Lp*Lb*Ap	2.46	2046	0.004	2.13	3.7E-02	Intercept	3.615	0.156	**23.12**
						Lp	-0.482	0.219	**-2.21**
						Lb	-0.748	0.221	**-3.38**
						Ap	-0.394	0.218	-1.80
						Lp:Lb	0.814	0.309	**2.64**
						Lp:Ap	0.579	0.307	1.89
						Lb:Ap	0.662	0.309	**2.14**
						Lp:Lb:Ap	-1.087	0.434	**-2.51**
Lp*Lb*At	2.45	2046	0.007	3.02	3.7E-03	Intercept	3.315	0.156	**21.23**
						Lp	-0.255	0.218	-1.17
						Lb	-0.396	0.221	-1.80
						At	0.191	0.218	0.88
						Lp:Lb	0.090	0.308	0.29
						Lp:At	0.141	0.306	0.46
						Lb:At	-0.023	0.308	-0.08
						Lp:Lb:At	0.340	0.433	0.79
Lp*Lb*Ao	2.45	2046	0.013	4.97	1.4E-05	Intercept	3.206	0.156	**20.60**
						Lp	-0.095	0.217	-0.44
						Lb	-0.680	0.220	**-3.09**
						Ao	0.405	0.217	1.86
						Lp:Lb	0.450	0.307	1.46
						Lp:Ao	-0.175	0.306	-0.57
						Lb:Ao	0.529	0.307	1.72
						Lp:Lb:Ao	-0.363	0.432	-0.84
Lp*Ap*At	2.46	2046	0.005	2.52	1.4E-02	Intercept	3.102	0.161	**19.33**
						Lp	-0.354	0.221	-1.60
						Ap	0.028	0.221	0.13
						At	0.264	0.221	1.19

						Lp:Ap	0.290	0.309	0.94
						Lp:At	0.572	0.309	1.85
						Ap:At	-0.167	0.309	-0.54
						Lp:Ap:At	-0.523	0.434	-1.21
Lp*Ap*Ao	2.44	2046	0.021	7.24	1.4E-08	Intercept	2.611	0.159	**16.39**
						Lp	0.101	0.220	0.46
						Ap	0.484	0.220	**2.21**
						Ao	1.198	0.220	**5.46**
						Lp:Ap	0.083	0.306	0.27
						Lp:Ao	-0.290	0.306	-0.95
						Ap:Ao	-1.031	0.306	**-3.37**
						Lp:Ap:Ao	-0.160	0.430	-0.37
Lp*At*Ao	2.44	2046	0.017	6.07	5.0E-07	Intercept	2.614	0.160	**16.38**
						Lp	-0.004	0.220	-0.02
						At	0.479	0.220	**2.18**
						Ao	0.957	0.220	**4.35**
						Lp:At	0.294	0.307	0.96
						Lp:Ao	-0.362	0.307	-1.18
						At:Ao	-0.549	0.307	-1.79
						Lp:At:Ao	-0.014	0.431	-0.03
Lb*Ap*At	2.45	2046	0.007	3.04	3.5E-03	Intercept	3.109	0.156	**19.92**
						Lb	-0.387	0.221	-1.75
						Ap	0.146	0.218	0.67
						At	0.506	0.218	**2.32**
						Lb:Ap	0.072	0.308	0.23
						Lb:At	0.108	0.308	0.35
						Ap:At	-0.477	0.306	-1.56
						Lb:Ap:At	0.077	0.433	0.18
Lb*Ap*Ao	2.43	2046	0.024	8.15	8.0E-10	Intercept	2.902	0.155	**18.75**
						Lb	-0.475	0.219	**-2.17**
						Ap	0.498	0.216	**2.31**
						Ao	0.912	0.216	**4.22**
						Lb:Ap	0.050	0.306	0.16
						Lb:Ao	0.277	0.306	0.91
						Ap:Ao	-1.173	0.304	**-3.86**
						Lb:Ap:Ao	0.128	0.430	0.30
Lb*At*Ao	2.44	2046	0.019	6.75	6.2E-08	Intercept	2.866	0.155	**18.47**
						Lb	-0.510	0.219	**-2.32**
						At	0.568	0.217	**2.62**
						Ao	0.619	0.216	**2.86**

						Lb:At	0.118	0.306	0.39
						Lb:Ao	0.312	0.306	1.02
						At:Ao	-0.586	0.305	-1.92
						Lb:At:Ao	0.062	0.431	0.14
Ap*At*Ao	2.42	2046	0.034	11.27	4.5E-14	Intercept	2.034	0.158	**12.86**
						Ap	1.098	0.218	**5.03**
						At	1.198	0.218	**5.49**
						Ao	1.676	0.218	**7.69**
						Ap:At	-1.085	0.304	**-3.57**
						Ap:Ao	-1.743	0.304	**-5.73**
						At:Ao	-1.189	0.304	**-3.91**
						Ap:At:Ao	1.206	0.428	**2.82**
Lp*Lb*Ap*At	2.45	2038	0.010	2.36	2.3E-03	Intercept	3.381	0.227	**14.93**
						Lp	-0.516	0.312	-1.65
						Lb	-0.557	0.320	-1.74
						Ap	-0.125	0.312	-0.40
						At	0.445	0.312	1.43
						Lp:Lb	0.324	0.442	0.73
						Lp:Ap	0.514	0.435	1.18
						Lb:Ap	0.305	0.442	0.69
						Lp:At	0.093	0.436	0.21
						Lb:At	-0.362	0.442	-0.82
						Ap:At	-0.514	0.436	-1.18
						Lp:Lb:Ap	-0.450	0.616	-0.73
						Lp:Lb:At	0.955	0.616	1.55
						Lp:Ap:At	0.104	0.612	0.17
						Lb:Ap:At	0.694	0.616	1.13
						Lp:Lb:Ap:At	-1.250	0.865	-1.44
Lp*Lb*Ap*Ao	2.43	2038	0.028	5.01	8.6E-10	Intercept	3.255	0.224	**14.51**
						Lp	-0.671	0.309	**-2.17**
						Lb	-1.288	0.317	**-4.06**
						Ap	-0.094	0.309	-0.30
						Ao	0.685	0.309	**2.21**
						Lp:Lb	1.545	0.437	**3.53**
						Lp:Ap	1.149	0.432	**2.66**
						Lb:Ap	1.156	0.437	**2.64**
						Lp:Ao	0.419	0.432	0.97
						Lb:Ao	1.027	0.437	**2.35**
						Ap:Ao	-0.564	0.432	-1.31
						Lp:Lb:Ap	-2.131	0.610	**-3.49**

						Lp:Lb:Ao	-1.416	0.610	**-2.32**
						Lp:Ap:Ao	-1.180	0.606	-1.95
						Lb:Ap:Ao	-0.934	0.610	-1.53
						Lp:Lb:Ap:Ao	2.040	0.857	**2.38**
Lp*Lb*At*Ao	2.44	2038	0.019	3.67	2.2E-06	Intercept	2.945	0.225	**13.07**
						Lp	-0.151	0.311	-0.49
						Lb	-0.664	0.319	**-2.08**
						At	0.495	0.311	1.59
						Ao	0.703	0.311	**2.26**
						Lp:Lb	0.293	0.439	0.67
						Lp:At	0.138	0.434	0.32
						Lb:At	-0.030	0.439	-0.07
						Lp:Ao	-0.174	0.433	-0.40
						Lb:Ao	0.508	0.439	1.16
						At:Ao	-0.569	0.434	-1.31
						Lp:Lb:At	0.312	0.613	0.51
						Lp:Lb:Ao	-0.377	0.613	-0.62
						Lp:At:Ao	-0.029	0.609	-0.05
						Lb:At:Ao	0.041	0.613	0.07
						Lp:Lb:At:Ao	0.028	0.861	0.03
Lp*Ap*At*Ao	2.42	2038	0.033	5.66	1.6E-11	Intercept	2.087	0.237	**8.79**
						Lp	-0.095	0.319	-0.30
						Ap	0.949	0.319	**2.98**
						At	0.944	0.319	**2.96**
						Ao	1.829	0.319	**5.74**
						Lp:Ap	0.287	0.438	0.66
						Lp:At	0.498	0.438	1.14
						Ap:At	-0.824	0.438	-1.88
						Lp:Ao	-0.315	0.438	-0.72
						Ap:Ao	-1.639	0.438	**-3.74**
						At:Ao	-1.156	0.438	**-2.64**
						Lp:Ap:At	-0.513	0.610	-0.84
						Lp:Ap:Ao	-0.196	0.610	-0.32
						Lp:At:Ao	-0.054	0.610	-0.09
						Ap:At:Ao	1.110	0.610	1.82
						Lp:Ap:At:Ao	0.180	0.856	0.21
Lb*Ap*At*Ao	2.42	2038	0.036	6.17	7.1E-13	Intercept	2.188	0.223	**9.79**
						Lb	-0.308	0.316	-0.97
						Ap	1.289	0.308	**4.19**
						At	1.357	0.308	**4.41**

						Ao	1.752	0.308	**5.69**
						Lb:Ap	-0.384	0.436	-0.88
						Lb:At	-0.318	0.436	-0.73
						Ap:At	-1.511	0.430	**-3.52**
						Lb:Ao	-0.151	0.436	-0.35
						Ap:Ao	-2.194	0.429	**-5.11**
						At:Ao	-1.608	0.430	**-3.74**
						Lb:Ap:At	0.851	0.608	1.40
						Lb:Ap:Ao	0.902	0.608	1.48
						Lb:At:Ao	0.839	0.608	1.38
						Ap:At:Ao	1.973	0.604	**3.27**
						Lb:Ap:At:Ao	-1.533	0.854	-1.80
Lp*Lb*Ap*At*Ao	2.41	2022	0.040	3.75	2.1E-11	Intercept	2.570	0.335	**7.68**
						Lp	-0.689	0.449	-1.54
						Lb	-0.968	0.473	**-2.05**
						Ap	0.675	0.449	1.50
						At	1.233	0.449	**2.75**
						Ao	1.459	0.449	**3.25**
						Lp:Lb	1.188	0.635	1.87
						Lp:Ap	1.152	0.617	1.87
						Lb:Ap	0.547	0.635	0.86
						Lp:At	0.172	0.617	0.28
						Lb:At	-0.577	0.635	-0.91
						Ap:At	-1.401	0.617	**-2.27**
						Lp:Ao	0.508	0.617	0.82
						Lb:Ao	0.739	0.635	1.16
						Ap:Ao	-1.438	0.617	**-2.33**
						At:Ao	-1.412	0.617	**-2.29**
						Lp:Lb:Ap	-1.729	0.872	**-1.98**
						Lp:Lb:At	0.651	0.872	0.75
						Lp:Ap:At	-0.143	0.860	-0.17
						Lb:Ap:At	1.154	0.872	1.32
						Lp:Lb:Ao	-1.646	0.872	-1.89
						Lp:Ap:Ao	-1.430	0.859	-1.67
						Lb:Ap:Ao	-0.401	0.872	-0.46
						Lp:At:Ao	-0.318	0.860	-0.37
						Lb:At:Ao	0.512	0.872	0.59
						Ap:At:Ao	1.611	0.860	1.88
						Lp:Lb:Ap:At	-0.739	1.216	-0.61
						Lp:Lb:Ap:Ao	2.469	1.215	**2.03**

	Lp:Lb:At:Ao	0.521	1.216	0.43
	Lp:Ap:At:Ao	0.646	1.207	0.54
	Lb:Ap:At:Ao	-1.002	1.216	-0.82
	Lp:Lb:Ap:At:Ao	-0.925	1.706	-0.54

Table S3: Multivariate linear regression analysis of development data to calculate interactions

Model	SE (residual)	DF	R^2 (adj.)	F	P	Parameter	Estimate	SE	t
Lp	0.70	764	0.003	3.67	5.6E-02	Intercept	10.024	0.036	**281.79**
						Lp	0.096	0.050	1.92
Lb	0.68	764	0.039	32.16	2.0E-08	Intercept	10.212	0.035	**292.37**
						Lb	-0.280	0.049	**-5.67**
Ap	0.70	764	0.003	2.96	8.6E-02	Intercept	10.115	0.036	**284.23**
						Ap	-0.087	0.050	-1.72
At	0.70	764	-0.001	0.07	7.9E-01	Intercept	10.079	0.036	**282.67**
						At	-0.013	0.050	-0.27
Ao	0.67	764	0.077	64.87	3.1E-15	Intercept	10.267	0.034	**299.92**
						Ao	-0.389	0.048	**-8.05**
Lp*Lb	0.68	762	0.042	12.07	1.0E-07	Intercept	10.153	0.049	**205.26**
						Lp	0.118	0.070	1.69
						Lb	-0.257	0.070	**-3.68**
						Lp:Lb	-0.045	0.099	-0.46
Lp*Ap	0.69	762	0.005	2.21	8.6E-02	Intercept	10.068	0.050	**199.75**
						Lp	0.093	0.071	1.31
						Ap	-0.089	0.071	-1.26
						Lp:Ap	0.006	0.100	0.06
Lp*At	0.70	762	0.003	1.72	1.6E-01	Intercept	10.000	0.050	**198.20**
						Lp	0.156	0.071	**2.20**
						At	0.047	0.071	0.66
						Lp:At	-0.120	0.101	-1.19
Lp*Ao	0.67	762	0.079	22.97	3.0E-14	Intercept	10.221	0.048	**210.83**
						Lp	0.091	0.068	1.34
						Ao	-0.393	0.068	**-5.75**
						Lp:Ao	0.008	0.097	0.08
Lb*Ap	0.68	762	0.043	12.47	5.8E-08	Intercept	10.221	0.049	**206.79**
						Lb	-0.211	0.070	**-3.02**
						Ap	-0.018	0.070	-0.26
						Lb:Ap	-0.138	0.098	-1.41
Lb*At	0.68	762	0.054	15.48	8.7E-10	Intercept	10.311	0.049	**209.78**
						Lb	-0.462	0.069	**-6.66**
						At	-0.196	0.069	**-2.83**
						Lb:At	0.363	0.098	**3.70**
Lb*Ao	0.65	762	0.129	38.76	< 2.2e-16	Intercept	10.489	0.047	**222.45**
						Lb	-0.443	0.067	**-6.65**
						Ao	-0.552	0.067	**-8.30**

						Lb:Ao	0.323	0.094	**3.44**
Ap*At	0.70	762	0.000	1.12	3.4E-01	Intercept	10.137	0.051	**200.68**
						Ap	-0.116	0.071	-1.63
						At	-0.043	0.071	-0.61
						Ap:At	0.059	0.101	0.58
Ap*Ao	0.66	762	0.095	27.69	< 2.2e-16	Intercept	10.400	0.048	**216.34**
						Ap	-0.265	0.068	**-3.90**
						Ao	-0.567	0.068	**-8.36**
						Ap:Ao	0.353	0.096	**3.69**
At*Ao	0.67	762	0.083	24.00	7.6E-15	Intercept	10.337	0.048	**213.61**
						At	-0.139	0.068	**-2.04**
						Ao	-0.514	0.068	**-7.53**
						At:Ao	0.248	0.096	**2.58**
Lp*Lb*Ap	0.68	758	0.044	6.06	6.8E-07	Intercept	10.138	0.070	**144.37**
						Lp	0.164	0.099	1.66
						Lb	-0.138	0.099	-1.40
						Ap	0.028	0.099	0.29
						Lp:Lb	-0.143	0.139	-1.03
						Lp:Ap	-0.091	0.139	-0.65
						Lb:Ap	-0.237	0.139	-1.70
						Lp:Lb:Ap	0.195	0.197	0.99
Lp*Lb*At	0.68	758	0.056	7.52	9.1E-09	Intercept	10.202	0.070	**146.20**
						Lp	0.215	0.098	**2.19**
						Lb	-0.400	0.098	**-4.08**
						At	-0.098	0.098	-1.00
						Lp:Lb	-0.121	0.138	-0.87
						Lp:At	-0.194	0.138	-1.40
						Lb:At	0.285	0.138	**2.06**
						Lp:Lb:At	0.152	0.196	0.78
Lp*Lb*Ao	0.65	758	0.133	17.69	< 2.2e-16	Intercept	10.394	0.067	**155.35**
						Lp	0.190	0.094	**2.02**
						Lb	-0.342	0.094	**-3.63**
						Ao	-0.477	0.094	**-5.07**
						Lp:Lb	-0.200	0.133	-1.51
						Lp:Ao	-0.148	0.133	-1.12
						Lb:Ao	0.164	0.133	1.24
						Lp:Lb:Ao	0.315	0.188	1.68
Lp*Ap*At	0.69	758	0.007	1.74	9.8E-02	Intercept	10.011	0.072	**139.83**
						Lp	0.250	0.101	**2.48**
						Ap	-0.021	0.101	-0.21

						At	0.114	0.101	1.14
						Lp:Ap	-0.187	0.142	-1.32
						Lp:At	-0.312	0.142	**-2.20**
						Ap:At	-0.135	0.142	-0.95
						Lp:Ap:At	0.385	0.201	1.92
Lp*Ap*Ao	0.66	758	0.096	12.62	2.3E-15	Intercept	10.330	0.068	**151.26**
						Lp	0.139	0.096	1.45
						Ap	-0.215	0.096	**-2.24**
						Ao	-0.517	0.096	**-5.38**
						Lp:Ap	-0.097	0.136	-0.72
						Lp:Ao	-0.097	0.136	-0.72
						Ap:Ao	0.246	0.136	1.82
						Lp:Ap:Ao	0.212	0.191	1.11
Lp*At*Ao	0.67	758	0.085	11.14	1.8E-13	Intercept	10.245	0.069	**149.09**
						Lp	0.182	0.097	1.89
						At	-0.047	0.097	-0.48
						Ao	-0.484	0.097	**-5.01**
						Lp:At	-0.182	0.136	-1.34
						Lp:Ao	-0.057	0.136	-0.42
						At:Ao	0.182	0.136	1.34
						Lp:At:Ao	0.130	0.193	0.68
Lb*Ap*At	0.68	758	0.056	7.46	1.1E-08	Intercept	10.330	0.070	**147.99**
						Lb	-0.382	0.098	**-3.89**
						Ap	-0.038	0.098	-0.39
						At	-0.215	0.098	**-2.19**
						Lb:Ap	-0.160	0.139	-1.15
						Lb:At	0.340	0.139	**2.46**
						Ap:At	0.038	0.139	0.28
						Lb:Ap:At	0.045	0.196	0.23
Lb*Ap*Ao	0.64	758	0.149	20.09	< 2.2e-16	Intercept	10.617	0.066	**160.20**
						Lb	-0.430	0.093	**-4.61**
						Ap	-0.252	0.093	**-2.71**
						Ao	-0.784	0.093	**-8.41**
						Lb:Ap	-0.029	0.132	-0.22
						Lb:Ao	0.430	0.132	**3.27**
						Ap:Ao	0.461	0.132	**3.50**
						Lb:Ap:Ao	-0.211	0.186	-1.14
Lb*At*Ao	0.64	758	0.152	20.53	< 2.2e-16	Intercept	10.681	0.066	**161.43**
						Lb	-0.681	0.093	**-7.32**
						At	-0.379	0.093	**-4.07**

						Ao	-0.733	0.093	**-7.87**
						Lb:At	0.473	0.131	**3.60**
						Lb:Ao	0.431	0.131	**3.28**
						At:Ao	0.358	0.131	**2.73**
						Lb:At:Ao	-0.212	0.185	-1.14
Ap*At*Ao	0.66	758	0.099	12.99	7.6E-16	Intercept	10.479	0.068	**153.67**
						Ap	-0.281	0.096	**-2.93**
						At	-0.156	0.096	-1.62
						Ao	-0.677	0.096	**-7.05**
						Ap:At	0.031	0.135	0.23
						Ap:Ao	0.322	0.135	**2.38**
						At:Ao	0.218	0.135	1.61
						Ap:At:Ao	0.063	0.191	0.33
Lp*Lb*Ap*At	0.68	750	0.059	4.22	1.5E-07	Intercept	10.152	0.100	**101.94**
						Lp	0.348	0.139	**2.50**
						Lb	-0.277	0.139	**-1.99**
						Ap	0.098	0.139	0.70
						At	-0.027	0.139	-0.20
						Lp:Lb	-0.202	0.196	-1.03
						Lp:Ap	-0.264	0.196	-1.35
						Lb:Ap	-0.244	0.196	-1.24
						Lp:At	-0.369	0.196	-1.88
						Lb:At	0.277	0.196	1.41
						Ap:At	-0.139	0.196	-0.71
						Lp:Lb:Ap	0.160	0.277	0.58
						Lp:Lb:At	0.119	0.277	0.43
						Lp:Ap:At	0.348	0.277	1.26
						Lb:Ap:At	0.014	0.277	0.05
						Lp:Lb:Ap:At	0.069	0.391	0.18
Lp*Lb*Ap*Ao	0.64	750	0.151	10.06	< 2.2e-16	Intercept	10.457	0.095	**110.50**
						Lp	0.314	0.132	**2.37**
						Lb	-0.248	0.132	-1.87
						Ap	-0.123	0.132	-0.93
						Ao	-0.623	0.132	**-4.71**
						Lp:Lb	-0.356	0.186	-1.91
						Lp:Ap	-0.252	0.186	-1.35
						Lb:Ap	-0.189	0.186	-1.02
						Lp:Ao	-0.314	0.186	-1.69
						Lb:Ao	0.207	0.186	1.11
						Ap:Ao	0.290	0.186	1.56

						Lp:Lb:Ap	0.314	0.263	1.20
						Lp:Lb:Ao	0.439	0.263	1.67
						Lp:Ap:Ao	0.335	0.263	1.28
						Lb:Ap:Ao	-0.082	0.263	-0.31
						Lp:Lb:Ap:Ao	-0.252	0.371	-0.68
Lp*Lb*At*Ao	0.64	750	0.154	10.31	< 2.2e-16	Intercept	10.500	0.094	**111.20**
						Lp	0.354	0.132	**2.68**
						Lb	-0.500	0.132	**-3.78**
						At	-0.208	0.132	-1.58
						Ao	-0.583	0.132	**-4.41**
						Lp:Lb	-0.354	0.186	-1.91
						Lp:At	-0.333	0.186	-1.79
						Lb:At	0.313	0.186	1.68
						Lp:Ao	-0.292	0.186	-1.57
						Lb:Ao	0.188	0.186	1.01
						At:Ao	0.208	0.186	1.12
						Lp:Lb:At	0.313	0.262	1.19
						Lp:Lb:Ao	0.479	0.262	1.83
						Lp:At:Ao	0.292	0.262	1.11
						Lb:At:Ao	-0.042	0.262	-0.16
						Lp:Lb:At:Ao	-0.333	0.370	-0.90
Lp*Ap*At*Ao	0.66	750	0.104	6.93	2.8E-14	Intercept	10.283	0.097	**105.80**
						Lp	0.384	0.136	**2.82**
						Ap	-0.074	0.136	-0.55
						At	0.092	0.136	0.68
						Ao	-0.533	0.136	**-3.92**
						Lp:Ap	-0.405	0.191	**-2.12**
						Lp:At	-0.488	0.191	**-2.55**
						Ap:At	-0.280	0.191	-1.46
						Lp:Ao	-0.280	0.191	-1.46
						Ap:Ao	0.095	0.191	0.50
						At:Ao	0.033	0.191	0.17
						Lp:Ap:At	0.613	0.270	**2.27**
						Lp:Ap:Ao	0.447	0.270	1.66
						Lp:At:Ao	0.363	0.270	1.35
						Ap:At:Ao	0.301	0.270	1.11
						Lp:Ap:At:Ao	-0.467	0.381	-1.23
Lb*Ap*At*Ao	0.64	750	0.168	11.32	< 2.2e-16	Intercept	10.826	0.094	**115.61**
						Lb	-0.680	0.131	**-5.19**
						Ap	-0.284	0.131	**-2.17**

						At	-0.409	0.131	**-3.12**
						Ao	-0.972	0.131	**-7.42**
						Lb:Ap	-0.007	0.184	-0.04
						Lb:At	0.493	0.184	**2.67**
						Ap:At	0.055	0.184	0.30
						Lb:Ao	0.576	0.184	**3.13**
						Ap:Ao	0.472	0.184	**2.56**
						At:Ao	0.368	0.184	**2.00**
						Lb:Ap:At	-0.034	0.260	-0.13
						Lb:Ap:Ao	-0.284	0.260	-1.09
						Lb:At:Ao	-0.284	0.260	-1.09
						Ap:At:Ao	-0.014	0.260	-0.05
						Lb:Ap:At:Ao	0.139	0.367	0.38
Lp*Lb*Ap*At*Ao	0.63	734	0.176	6.27	< 2.2e-16	Intercept	10.455	0.135	**77.56**
						Lp	0.712	0.187	**3.82**
						Lb	-0.330	0.187	-1.77
						Ap	0.087	0.187	0.47
						At	0.004	0.187	0.02
						Ao	-0.580	0.187	**-3.11**
						Lp:Lb	-0.670	0.261	**-2.57**
						Lp:Ap	-0.712	0.261	**-2.73**
						Lb:Ap	-0.337	0.261	-1.29
						Lp:At	-0.795	0.261	**-3.05**
						Lb:At	0.163	0.261	0.62
						Ap:At	-0.420	0.261	-1.61
						Lp:Ao	-0.754	0.261	**-2.89**
						Lb:Ao	0.080	0.261	0.31
						Ap:Ao	-0.004	0.261	-0.02
						At:Ao	-0.087	0.261	-0.33
						Lp:Lb:Ap	0.629	0.367	1.71
						Lp:Lb:At	0.629	0.367	1.71
						Lp:Ap:At	0.920	0.367	**2.51**
						Lb:Ap:At	0.295	0.367	0.81
						Lp:Lb:Ao	0.962	0.367	**2.62**
						Lp:Ap:Ao	0.920	0.367	**2.51**
						Lb:Ap:Ao	0.212	0.367	0.58
						Lp:At:Ao	0.879	0.367	**2.39**
						Lb:At:Ao	0.254	0.367	0.69
						Ap:At:Ao	0.587	0.367	1.60
						Lp:Lb:Ap:At	-0.629	0.518	-1.22

	Lp:Lb:Ap:Ao	-0.962	0.518	-1.86
	Lp:Lb:At:Ao	-1.045	0.518	**-2.02**
	Lp:Ap:At:Ao	-1.170	0.518	**-2.26**
	Lb:Ap:At:Ao	-0.587	0.518	-1.13
	Lp:Lb:Ap:At:Ao	1.420	0.731	1.94

Table S4: Multivariate linear regression analysis of lifespan data to calculate interactions

Model	SE (residual)	DF	R^2 (adj.)	F	P	Parameter	Estimate	SE	t
Lp	12.26	3158	0.004	14.96	1.1E-04	Intercept	47.62	0.31	**153.45**
						Lp	-1.69	0.44	**-3.87**
Lb	12.28	3158	0.001	2.91	8.8E-02	Intercept	47.14	0.31	**152.58**
						Lb	-0.75	0.44	-1.71
Ap	12.27	3158	0.003	10.08	1.5E-03	Intercept	46.06	0.31	**148.32**
						Ap	1.39	0.44	**3.18**
At	12.22	3158	0.011	35.69	2.6E-09	Intercept	48.08	0.31	**155.44**
						At	-2.60	0.43	**-5.97**
Ao	12.17	3158	0.019	60.59	9.5E-15	Intercept	48.47	0.31	**157.32**
						Ao	-3.37	0.43	**-7.78**
Lp*Lb	12.24	3156	0.008	9.37	3.6E-06	Intercept	48.70	0.44	**111.15**
						Lp	-3.07	0.62	**-4.99**
						Lb	-2.15	0.62	**-3.47**
						Lp:Lb	2.78	0.87	**3.19**
Lp*Ap	12.24	3156	0.007	8.27	1.8E-05	Intercept	46.91	0.44	**105.65**
						Lp	-1.66	0.62	**-2.67**
						Ap	1.38	0.62	**2.22**
						Lp:Ap	-0.02	0.87	-0.03
Lp*At	12.19	3156	0.015	17.30	3.8E-11	Intercept	49.09	0.44	**111.02**
						Lp	-1.97	0.62	**-3.19**
						At	-2.87	0.62	**-4.65**
						Lp:At	0.49	0.87	0.57
Lp*Ao	12.14	3156	0.023	25.69	< 2.2e-16	Intercept	49.46	0.44	**112.28**
						Lp	-1.92	0.62	**-3.12**
						Ao	-3.58	0.62	**-5.82**
						Lp:Ao	0.37	0.86	0.43
Lb*Ap	12.24	3156	0.007	8.40	1.5E-05	Intercept	45.67	0.44	**104.20**
						Lb	0.79	0.62	1.28
						Ap	2.90	0.62	**4.71**
						Lb:Ap	-3.03	0.87	**-3.48**
Lb*At	12.20	3156	0.013	14.92	1.2E-09	Intercept	47.91	0.44	**109.65**
						Lb	0.33	0.62	0.54
						At	-1.53	0.61	**-2.49**
						Lb:At	-2.13	0.87	**-2.46**
Lb*Ao	12.16	3156	0.020	22.40	2.4E-14	Intercept	49.26	0.44	**113.12**
						Lb	-1.57	0.62	**-2.55**
						Ao	-4.19	0.61	**-6.84**

						Lb:Ao	1.63	0.87	1.89
Ap*At	12.20	3156	0.014	15.54	4.9E-10	Intercept	47.15	0.44	**106.54**
						Ap	1.82	0.62	**2.94**
						At	-2.11	0.62	**-3.42**
						Ap:At	-0.92	0.87	-1.06
Ap*Ao	12.15	3156	0.022	24.94	6.1E-16	Intercept	48.25	0.44	**109.51**
						Ap	0.43	0.62	0.70
						Ao	-4.27	0.62	**-6.93**
						Ap:Ao	1.80	0.86	**2.09**
At*Ao	12.09	3156	0.031	34.58	< 2.2e-16	Intercept	50.31	0.44	**114.70**
						At	-3.59	0.61	**-5.86**
						Ao	-4.35	0.61	**-7.11**
						At:Ao	1.87	0.86	**2.17**
Lp*Lb*Ap	12.19	3152	0.015	7.83	2.0E-09	Intercept	46.73	0.63	**74.72**
						Lp	-2.08	0.87	**-2.38**
						Lb	0.36	0.88	0.41
						Ap	3.83	0.87	**4.38**
						Lp:Lb	0.84	1.24	0.68
						Lp:Ap	-1.90	1.23	-1.55
						Lb:Ap	-4.90	1.24	**-3.97**
						Lp:Lb:Ap	3.75	1.74	**2.16**
Lp*Lb*At	12.16	3152	0.020	10.20	1.1E-12	Intercept	49.55	0.62	**79.43**
						Lp	-3.19	0.87	**-3.66**
						Lb	-0.92	0.88	-1.04
						At	-1.67	0.87	-1.91
						Lp:Lb	2.44	1.23	**1.98**
						Lp:At	0.19	1.22	0.15
						Lb:At	-2.41	1.23	-1.96
						Lp:Lb:At	0.61	1.73	0.36
Lp*Lb*Ao	12.12	3152	0.027	13.61	< 2.2e-16	Intercept	50.77	0.62	**81.68**
						Lp	-2.94	0.87	**-3.39**
						Lb	-2.62	0.88	**-2.98**
						Ao	-4.04	0.87	**-4.65**
						Lp:Lb	2.04	1.23	1.66
						Lp:Ao	-0.38	1.22	-0.31
						Lb:Ao	0.91	1.23	0.74
						Lp:Lb:Ao	1.49	1.72	0.87
Lp*Ap*At	12.18	3152	0.018	9.07	4.0E-11	Intercept	48.41	0.64	**75.44**
						Lp	-2.40	0.88	**-2.72**
						Ap	1.29	0.88	1.46

						At	-2.85	0.88	**-3.22**
						Lp:Ap	0.93	1.23	0.76
						Lp:At	1.34	1.23	1.09
						Ap:At	0.03	1.23	0.02
						Lp:Ap:At	-1.77	1.73	-1.02
Lp*Ap*Ao	12.12	3152	0.026	13.14	< 2.2e-16	Intercept	49.47	0.64	**77.43**
						Lp	-2.32	0.88	**-2.63**
						Ap	-0.03	0.88	-0.03
						Ao	-4.86	0.88	**-5.52**
						Lp:Ap	0.80	1.23	0.65
						Lp:Ao	1.06	1.23	0.87
						Ap:Ao	2.56	1.23	**2.08**
						Lp:Ap:Ao	-1.39	1.73	-0.80
Lp*At*Ao	12.06	3152	0.036	17.68	< 2.2e-16	Intercept	51.83	0.64	**81.52**
						Lp	-2.88	0.88	**-3.29**
						At	-4.51	0.88	**-5.15**
						Ao	-5.20	0.88	**-5.94**
						Lp:At	1.69	1.22	1.38
						Lp:Ao	1.55	1.22	1.27
						At:Ao	3.01	1.22	**2.46**
						Lp:At:Ao	-2.12	1.72	-1.24
Lb*Ap*At	12.17	3152	0.019	9.80	3.9E-12	Intercept	46.29	0.62	**74.17**
						Lb	1.72	0.88	1.95
						Ap	3.17	0.87	**3.64**
						At	-1.20	0.87	-1.38
						Lb:Ap	-2.71	1.23	**-2.20**
						Lb:At	-1.82	1.23	-1.48
						Ap:At	-0.57	1.22	-0.47
						Lb:Ap:At	-0.70	1.73	-0.40
Lb*Ap*Ao	12.11	3152	0.028	14.18	< 2.2e-16	Intercept	48.77	0.62	**78.52**
						Lb	-1.04	0.88	-1.19
						Ap	0.95	0.87	1.09
						Ao	-6.05	0.87	**-6.98**
						Lb:Ap	-1.03	1.23	-0.84
						Lb:Ao	3.57	1.23	**2.91**
						Ap:Ao	3.76	1.22	**3.08**
						Lb:Ap:Ao	-3.91	1.72	**-2.27**
Lb*At*Ao	12.07	3152	0.035	17.39	< 2.2e-16	Intercept	50.07	0.62	**80.88**
						Lb	0.49	0.88	0.57
						At	-1.57	0.86	-1.82

						Ao	-4.20	0.86	**-4.85**
						Lb:At	-4.03	1.22	**-3.30**
						Lb:Ao	-0.31	1.22	-0.26
						At:Ao	-0.02	1.21	-0.02
						Lb:At:Ao	3.78	1.72	**2.20**
Ap*At*Ao	12.06	3152	0.036	17.87	< 2.2e-16	Intercept	50.50	0.64	**79.45**
						Ap	-0.36	0.88	-0.41
						At	-4.27	0.88	**-4.87**
						Ao	-6.37	0.88	**-7.27**
						Ap:At	1.35	1.22	1.10
						Ap:Ao	4.01	1.22	**3.28**
						At:Ao	3.98	1.22	**3.25**
						Ap:At:Ao	-4.20	1.72	**-2.44**
Lp*Lb*Ap*At	12.12	3144	0.026	6.72	1.6E-14	Intercept	47.64	0.90	**52.74**
						Lp	-2.58	1.25	**-2.07**
						Lb	1.54	1.28	1.20
						Ap	3.62	1.25	**2.91**
						At	-1.73	1.25	-1.39
						Lp:Lb	0.35	1.76	0.20
						Lp:Ap	-1.03	1.74	-0.59
						Lb:Ap	-4.66	1.76	**-2.65**
						Lp:At	0.91	1.74	0.53
						Lb:At	-2.24	1.76	-1.27
						Ap:At	0.32	1.74	0.18
						Lp:Lb:Ap	3.93	2.46	1.60
						Lp:Lb:At	0.86	2.46	0.35
						Lp:Ap:At	-1.64	2.44	-0.67
						Lb:Ap:At	-0.59	2.46	-0.24
						Lp:Lb:Ap:At	-0.24	3.45	-0.07
Lp*Lb*Ap*Ao	12.05	3144	0.037	9.14	< 2.2e-16	Intercept	50.50	0.90	**56.21**
						Lp	-3.28	1.24	**-2.65**
						Lb	-2.06	1.27	-1.62
						Ap	0.51	1.24	0.41
						Ao	-7.16	1.24	**-5.78**
						Lp:Lb	1.93	1.75	1.10
						Lp:Ap	0.71	1.73	0.41
						Lb:Ap	-1.07	1.75	-0.61
						Lp:Ao	2.03	1.73	1.18
						Lb:Ao	4.59	1.75	**2.62**
						Ap:Ao	6.27	1.73	**3.63**

						Lp:Lb:Ap	0.17	2.44	0.07
						Lp:Lb:Ao	-1.93	2.44	-0.79
						Lp:Ap:Ao	-4.84	2.43	**-1.99**
						Lb:Ap:Ao	-7.42	2.44	**-3.03**
						Lp:Lb:Ap:Ao	6.91	3.43	**2.01**
Lp*Lb*At*Ao	12.01	3144	0.044	10.64	< 2.2e-16	Intercept	52.57	0.90	**58.71**
						Lp	-4.75	1.23	**-3.85**
						Lb	-1.47	1.27	-1.16
						At	-3.42	1.23	**-2.77**
						Ao	-5.73	1.23	**-4.64**
						Lp:Lb	3.74	1.75	**2.14**
						Lp:At	3.45	1.72	**2.00**
						Lb:At	-2.18	1.75	-1.25
						Lp:Ao	2.82	1.72	1.64
						Lb:Ao	1.06	1.75	0.61
						At:Ao	3.21	1.72	1.87
						Lp:Lb:At	-3.51	2.44	-1.44
						Lp:Lb:Ao	-2.54	2.44	-1.04
						Lp:At:Ao	-6.22	2.42	**-2.57**
						Lb:At:Ao	-0.41	2.44	-0.17
						Lp:Lb:At:Ao	8.19	3.42	**2.39**
Lp*Ap*At*Ao	12.03	3144	0.040	9.88	< 2.2e-16	Intercept	52.71	0.95	**55.41**
						Lp	-3.98	1.28	**-3.12**
						Ap	-1.59	1.28	-1.24
						At	-5.83	1.28	**-4.57**
						Ao	-7.74	1.28	**-6.06**
						Lp:Ap	2.02	1.75	1.15
						Lp:At	2.68	1.75	1.53
						Ap:At	2.47	1.75	1.41
						Lp:Ao	2.30	1.75	1.31
						Ap:Ao	4.89	1.75	**2.79**
						At:Ao	5.11	1.75	**2.91**
						Lp:Ap:At	-1.80	2.44	-0.74
						Lp:Ap:Ao	-1.32	2.44	-0.54
						Lp:At:Ao	-1.82	2.44	-0.74
						Ap:At:Ao	-4.02	2.44	-1.65
						Lp:Ap:At:Ao	-0.79	3.43	-0.23
Lb*Ap*At*Ao	12.00	3144	0.046	11.08	< 2.2e-16	Intercept	49.58	0.89	**55.43**
						Lb	1.83	1.26	1.45
						Ap	0.92	1.23	0.74

						At	-1.54	1.23	-1.25
						Ao	-6.26	1.23	**-5.08**
						Lb:Ap	-2.54	1.74	-1.46
						Lb:At	-5.46	1.74	**-3.13**
						Ap:At	-0.02	1.72	-0.01
						Lb:Ao	-0.21	1.74	-0.12
						Ap:Ao	4.18	1.72	**2.43**
						At:Ao	0.34	1.72	**0.20**
						Lb:Ap:At	2.73	2.43	1.12
						Lb:Ap:Ao	-0.34	2.43	-0.14
						Lb:At:Ao	7.28	2.43	**2.99**
						Ap:At:Ao	-0.77	2.42	-0.32
						Lb:Ap:At:Ao	-6.85	3.42	**-2.01**
Lp*Lb*Ap*At*Ao	11.93	3128	0.057	7.12	< 2.2e-16	Intercept	53.25	1.33	**39.92**
						Lp	-6.60	1.79	**-3.69**
						Lb	-1.08	1.89	-0.57
						Ap	-1.23	1.79	-0.69
						At	-4.95	1.79	**-2.77**
						Ao	-10.09	1.79	**-5.64**
						Lp:Lb	5.24	2.53	**2.07**
						Lp:Ap	3.56	2.46	1.45
						Lb:Ap	-0.72	2.53	-0.28
						Lp:At	6.09	2.46	**2.48**
						Lb:At	-1.77	2.53	-0.70
						Ap:At	2.92	2.46	1.19
						Lp:Ao	6.92	2.46	**2.81**
						Lb:Ao	4.71	2.53	1.86
						Ap:Ao	8.58	2.46	**3.49**
						At:Ao	5.32	2.46	**2.16**
						Lp:Lb:Ap	-3.08	3.48	-0.88
						Lp:Lb:At	-6.82	3.48	-1.96
						Lp:Ap:At	-5.15	3.43	-1.50
						Lb:Ap:At	-0.91	3.48	-0.26
						Lp:Lb:Ao	-9.25	3.48	**-2.66**
						Lp:Ap:Ao	-8.06	3.43	**-2.35**
						Lb:Ap:Ao	-7.38	3.48	**-2.12**
						Lp:At:Ao	-9.23	3.43	**-2.69**
						Lb:At:Ao	-0.43	3.48	-0.12
						Ap:At:Ao	-4.08	3.43	-1.19
						Lp:Lb:Ap:At	6.70	4.85	1.38

Lp:Lb:Ap:Ao	13.49	4.85	**2.78**
Lp:Lb:At:Ao	14.83	4.85	**3.06**
Lp:Ap:At:Ao	5.89	4.81	1.23
Lb:Ap:At:Ao	0.12	4.85	0.02
Lp:Lb:Ap:At:Ao	-13.36	6.80	**-1.96**

Table S5: Multivariate linear regression analysis of bacterial abundance data to calculate interactions

Model	SE (residual)	DF	R^2 (adj.)	F	P	Parameter	Estimate	SE	t
Lp	395300	1534	0.029	46.42	1.4E-11	Intercept	423858	14265	**29.71**
						Lp	137451	20173	**6.81**
Lb	383300	1534	0.087	147.40	< 2.2E-16	Intercept	373869	13830	**27.03**
						Lb	237429	19558	**12.14**
Ap	400800	1534	0.001	3.29	7.0E-02	Intercept	474039	14464	**32.78**
						Ap	37089	20454	1.81
At	399200	1534	0.010	16.19	6.0E-05	Intercept	451600	14403	**31.35**
						At	81966	20369	**4.02**
Ao	396500	1534	0.023	36.87	1.6E-09	Intercept	431150	14308	**30.13**
						Ao	122866	20235	**6.07**
Lp*Lb	368900	1532	0.154	94.14	< 2.2E-16	Intercept	226321	18828	**12.02**
						Lp	295096	26626	**11.08**
						Lb	395074	26626	**14.84**
						Lp:Lb	-315290	37655	**-8.37**
Lp*Ap	394700	1532	0.032	17.91	2.0E-11	Intercept	424810	20140	**21.09**
						Lp	98457	28482	**3.46**
						Ap	-1905	28482	-0.07
						Lp:Ap	77988	40280	1.94
Lp*At	393400	1532	0.038	21.18	1.9E-13	Intercept	383628	20078	**19.11**
						Lp	135944	28394	**4.79**
						At	80459	28394	**2.83**
						Lp:At	3013	40155	0.08
Lp*Ao	390800	1532	0.051	28.51	< 2.2E-16	Intercept	364930	19941	**18.30**
						Lp	132440	28200	**4.70**
						Ao	117856	28200	**4.18**
						Lp:Ao	10021	39881	0.25
Lb*Ap	383000	1532	0.088	50.48	< 2.2E-16	Intercept	360626	19546	**18.45**
						Lb	226827	27643	**8.21**
						Ap	26486	27643	0.96
						Lb:Ap	21205	39093	0.54
Lb*At	381000	1532	0.098	56.44	< 2.2E-16	Intercept	317699	19443	**16.34**
						Lb	267803	27497	**9.74**
						At	112339	27497	**4.09**
						Lb:At	-60747	38887	-1.56
Lb*Ao	378500	1532	0.110	63.93	< 2.2E-16	Intercept	317344	19316	**16.43**
						Lb	227612	27317	**8.33**

						Ao	113050	27317	**4.14**
						Lb:Ao	19634	38632	0.51
Ap*At	398000	1532	0.016	9.19	5.0E-06	Intercept	404455	20308	**19.92**
						Ap	94291	28720	**3.28**
						At	139167	28720	**4.85**
						Ap:At	-114403	40616	**-2.82**
Ap*Ao	394800	1532	0.031	17.47	3.7E-11	Intercept	377952	20148	**18.76**
						Ap	106395	28494	**3.73**
						Ao	192173	28494	**6.74**
						Ap:Ao	-138612	40296	**-3.44**
At*Ao	394400	1532	0.033	18.52	8.3E-12	Intercept	376946	20128	**18.73**
						At	108409	28465	**3.81**
						Ao	149310	28465	**5.25**
						At:Ao	-52886	40256	-1.31
Lp*Lb*Ap	368400	1528	0.157	41.70	< 2.2E-16	Intercept	229534	26586	**8.63**
						Lp	262183	37599	**6.97**
						Lb	390553	37599	**10.39**
						Ap	-6426	37599	-0.17
						Lp:Lb	-327453	53173	**-6.16**
						Lp:Ap	65825	53173	1.24
						Lb:Ap	9041	53173	0.17
						Lp:Lb:Ap	24327	75197	0.32
Lp*Lb*At	366400	1528	0.166	44.57	< 2.2E-16	Intercept	152962	26441	**5.79**
						Lp	329475	37393	**8.81**
						Lb	461334	37393	**12.34**
						At	146719	37393	**3.92**
						Lp:Lb	-387062	52882	**-7.32**
						Lp:At	-68759	52882	-1.30
						Lb:At	-132519	52882	**-2.51**
						Lp:Lb:At	143544	74786	1.92
Lp*Lb*Ao	364200	1528	0.176	47.78	< 2.2E-16	Intercept	165738	26281	**6.31**
						Lp	303211	37166	**8.16**
						Lb	398383	37166	**10.72**
						Ao	121165	37166	**3.26**
						Lp:Lb	-341541	52561	**-6.50**
						Lp:Ao	-16231	52561	-0.31
						Lb:Ao	-6618	52561	-0.13
						Lp:Lb:Ao	52503	74333	0.71
Lp*Ap*At	392000	1528	0.045	11.35	4.3E-14	Intercept	356703	28289	**12.61**
						Lp	95504	40006	**2.39**

						Ap	53850	40006	1.35
						At	136214	40006	**3.41**
						Lp:Ap	80882	56578	1.43
						Lp:At	5906	56578	0.10
						Ap:At	-111510	56578	-1.97
						Lp:Ap:At	-5787	80013	-0.07
Lp*Ap*Ao	387800	1528	0.065	16.37	< 2.2E-16	Intercept	359270	27985	**12.84**
						Lp	37364	39576	0.94
						Ap	11319	39576	0.29
						Ao	131080	39576	**3.31**
						Lp:Ap	190152	55970	**3.40**
						Lp:Ao	122184	55970	**2.18**
						Ap:Ao	-26449	55970	-0.47
						Lp:Ap:Ao	-224327	79153	**-2.83**
Lp*At*Ao	388700	1528	0.061	15.28	< 2.2E-16	Intercept	298748	28050	**10.65**
						Lp	156394	39668	**3.94**
						At	132363	39668	**3.34**
						Ao	169760	39668	**4.28**
						Lp:At	-47908	56099	-0.85
						Lp:Ao	-40900	56099	-0.73
						At:Ao	-103807	56099	-1.85
						Lp:At:Ao	101842	79336	1.28
Lb*Ap*At	379800	1528	0.103	26.30	< 2.2E-16	Intercept	266007	27410	**9.71**
						Lb	276897	38764	**7.14**
						Ap	103385	38764	**2.67**
						At	189238	38764	**4.88**
						Lb:Ap	-18188	54821	-0.33
						Lb:At	-100140	54821	-1.83
						Ap:At	-153796	54821	**-2.81**
						Lb:Ap:At	78786	77529	1.02
Lb*Ap*Ao	376700	1528	0.118	30.42	< 2.2E-16	Intercept	255278	27183	**9.39**
						Lb	245348	38442	**6.38**
						Ap	124131	38442	**3.23**
						Ao	210694	38442	**5.48**
						Lb:Ap	-35471	54366	-0.65
						Lb:Ao	-37042	54366	-0.68
						Ap:Ao	-195289	54366	**-3.59**
						Lb:Ap:Ao	113352	76884	1.47
Lb*At*Ao	376000	1528	0.121	31.27	< 2.2E-16	Intercept	234829	27136	**8.65**
						Lb	284233	38376	**7.41**

						At	165030	38376	**4.30**
						Ao	165740	38376	**4.32**
						Lb:At	-113242	54272	**-2.09**
						Lb:Ao	-32861	54272	-0.61
						At:Ao	-105381	54272	-1.94
						Lb:At:Ao	104989	76753	1.37
Ap*At*Ao	391700	1528	0.047	11.72	1.4E-14	Intercept	280970	28266	**9.94**
						Ap	191952	39974	**4.80**
						At	193965	39974	**4.85**
						Ao	246970	39974	**6.18**
						Ap:At	-171113	56532	**-3.03**
						Ap:Ao	-195322	56532	**-3.46**
						At:Ao	-109596	56532	-1.94
						Ap:At:Ao	113418	79948	1.42
Lp*Lb*Ap*At	364400	1520	0.175	22.66	< 2.2E-16	Intercept	139633	37192	**3.75**
						Lp	252747	52597	**4.81**
						Lb	434140	52597	**8.25**
						Ap	26656	52597	0.51
						At	179801	52597	**3.42**
						Lp:Lb	-314486	74384	**-4.23**
						Lp:Ap	153457	74384	**2.06**
						Lb:Ap	54387	74384	0.73
						Lp:At	18873	74384	0.25
						Lb:At	-87173	74384	-1.17
						Ap:At	-66164	74384	-0.89
						Lp:Lb:Ap	-145151	105195	-1.38
						Lp:Lb:At	-25934	105195	-0.25
						Lp:Ap:At	-175265	105195	-1.67
						Lb:Ap:At	-90692	105195	-0.86
						Lp:Lb:Ap:At	338956	148768	**2.28**
Lp*Lb*Ap*Ao	361000	1520	0.190	25.03	< 2.2E-16	Intercept	153266	36842	**4.16**
						Lp	204026	52102	**3.92**
						Lb	412009	52102	**7.91**
						Ap	24946	52102	0.48
						Ao	152536	52102	**2.93**
						Lp:Lb	-333323	73683	**-4.52**
						Lp:Ap	198370	73683	**2.69**
						Lb:Ap	-27253	73683	-0.37
						Lp:Ao	116315	73683	1.58
						Lb:Ao	-42912	73683	-0.58

						Ap:Ao	-62743	73683	-0.85
						Lp:Lb:Ap	-16436	104204	-0.16
						Lp:Lb:Ao	11739	104204	0.11
						Lp:Ap:Ao	-265091	104204	**-2.54**
						Lb:Ap:Ao	72589	104204	0.70
						Lp:Lb:Ap:Ao	81528	147366	0.55
Lp*Lb*At*Ao	361300	1520	0.189	24.83	< 2.2E-16	Intercept	48220	36871	1.31
						Lp	373218	52144	**7.16**
						Lb	501057	52144	**9.61**
						At	235038	52144	**4.51**
						Ao	209483	52144	**4.02**
						Lp:Lb	-433648	73742	**-5.88**
						Lp:At	-140015	73742	-1.90
						Lb:At	-205349	73742	**-2.79**
						Lp:Ao	-87487	73742	-1.19
						Lb:Ao	-79448	73742	-1.08
						At:Ao	-176637	73742	**-2.40**
						Lp:Lb:At	184214	104287	1.77
						Lp:Lb:Ao	93174	104287	0.89
						Lp:At:Ao	142512	104287	1.37
						Lb:At:Ao	145660	104287	1.40
						Lp:Lb:At:Ao	-81341	147485	-0.55
Lp*Ap*At*Ao	384600	1520	0.081	9.99	< 2.2E-16	Intercept	237856	39252	**6.06**
						Lp	86227	55511	1.55
						Ap	121784	55511	**2.19**
						At	242828	55511	**4.37**
						Ao	237694	55511	**4.28**
						Lp:Ap	140334	78505	1.79
						Lp:At	-97726	78505	-1.25
						Ap:At	-220930	78505	**-2.81**
						Lp:Ao	18552	78505	0.24
						Ap:Ao	-135869	78505	-1.73
						At:Ao	-213228	78505	**-2.72**
						Lp:Ap:At	99635	111022	0.90
						Lp:Ap:Ao	-118905	111022	-1.07
						Lp:At:Ao	207264	111022	1.87
						Ap:At:Ao	218840	111022	**1.97**
						Lp:Ap:At:Ao	-210844	157009	-1.34
Lb*Ap*At*Ao	373100	1520	0.135	16.96	< 2.2E-16	Intercept	116623	38078	**3.06**
						Lb	328694	53851	**6.10**

						Ap	236413	53851	**4.39**
						At	277312	53851	**5.15**
						Ao	298768	53851	**5.55**
						Lb:Ap	-88922	76156	-1.17
						Lb:At	-166693	76156	**-2.19**
						Ap:At	-224564	76156	**-2.95**
						Lb:Ao	-103595	76156	-1.36
						Ap:Ao	-266056	76156	**-3.49**
						At:Ao	-176148	76156	**-2.31**
						Lb:Ap:At	106902	107701	0.99
						Lb:Ap:Ao	141468	107701	1.31
						Lb:At:Ao	133105	107701	1.24
						Ap:At:Ao	141535	107701	1.31
						Lb:Ap:At:Ao	-56232	152313	-0.37
Lp*Lb*Ap*At*Ao	356300	1504	0.211	14.25	< 2.2E-16	Intercept	-1.0E-09	5.1E+04	0.00
						Lp	2.3E+05	7.3E+04	**3.21**
						Lb	4.8E+05	7.3E+04	**6.54**
						Ap	9.6E+04	7.3E+04	1.33
						At	3.1E+05	7.3E+04	**4.22**
						Ao	2.8E+05	7.3E+04	**3.84**
						Lp:Lb	-2.9E+05	1.0E+05	**-2.86**
						Lp:Ap	2.8E+05	1.0E+05	**2.72**
						Lb:Ap	5.1E+04	1.0E+05	0.49
						Lp:At	-5.8E+04	1.0E+05	-0.57
						Lb:At	-1.3E+05	1.0E+05	-1.24
						Ap:At	-1.4E+05	1.0E+05	-1.39
						Lp:Ao	3.9E+04	1.0E+05	0.38
						Lb:Ao	-8.3E+04	1.0E+05	-0.81
						Ap:Ao	-1.4E+05	1.0E+05	-1.36
						At:Ao	-2.5E+05	1.0E+05	**-2.46**
						Lp:Lb:Ap	-2.8E+05	1.5E+05	-1.92
						Lp:Lb:At	-7.9E+04	1.5E+05	-0.54
						Lp:Ap:At	-1.6E+05	1.5E+05	-1.12
						Lb:Ap:At	-1.6E+05	1.5E+05	-1.07
						Lp:Lb:Ao	-4.1E+04	1.5E+05	-0.28
						Lp:Ap:Ao	-2.5E+05	1.5E+05	-1.74
						Lb:Ap:Ao	7.4E+03	1.5E+05	0.05
						Lp:At:Ao	1.5E+05	1.5E+05	1.06
						Lb:At:Ao	8.0E+04	1.5E+05	0.55
						Ap:At:Ao	1.5E+05	1.5E+05	1.06

Lp:Lb:Ap:At	5.3E+05	2.1E+05	**2.56**
Lp:Lb:Ap:Ao	2.7E+05	2.1E+05	1.30
Lp:Lb:At:Ao	1.1E+05	2.1E+05	0.51
Lp:Ap:At:Ao	-2.4E+04	2.1E+05	-0.12
Lb:Ap:At:Ao	1.3E+05	2.1E+05	0.63
Lp:Lb:Ap:At:Ao	-3.7E+05	2.9E+05	-1.28

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

November 14, 2018

Contents

1 Availability of code

The Python code for computing interactions coordinates and circuits in a system of n-bacterial species is available on Github at `https://github.com/cbg-ethz/epistasis-formulas`, see [8]. The code can also be used to compute the magnitude of an interaction as well as the interval in which an interaction is contained if the starting measurements are given as intervals.

2 Introduction to Geometric Interactions (Figure 4)

It is in open question in the microbiome field of how to quantify the many possible interactions between different bacterial species within a microbial community and to measure how these interactions impact host physiology traits. Here, we apply the mathematics of genetic epistasis to the microbiome. We explicitly make an analogy between genetic loci in a genome and bacterial species in a microbiome in order to calculate the interactions between species in a microbiome and their effects on host physiology. The basic assumption is that if two bacterial species have independent effects on the host, their phenotypes will be additive. The interaction is the degree to which this assumption is incorrect.

For an n-genotype system, genetic interactions, *i.e.* epistasis, are commonly described in terms of 'interaction coordinates' and 'circuits' (see work of Beerenwinkel, Pachter, and Sturmfels[2]). Interaction coordinates are equations that calculate the effect of interactions between genetic loci on organismal traits. 'Circuits' use the interaction coordinates as basis vectors and can give a richer description of the complete interaction space (see [2] for details). Circuits can, for instance, ask how much of a three-way interaction can be explained by a two-way interaction between individual species. Beerenwinkel, Pachter, and Sturmfels (BPS) described a formal mathematical framework to quantify interactions between genetic loci in an n-locus system [2]. The combinatorial nature and flexibility of the approach make the BPS framework generalizable to different types of high dimensional interacting systems beyond gene networks. Here we apply the formal framework with its interaction coordinates and circuits to the microbiome.

Besides a graspable biological interpretation, interaction coordinates and 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. These higher order interactions generalize several well-known and widely used notions of gene interaction, including Fourier-Walsh coefficients (see Box 1 in Weinreich, Lan, et al. 2013 [12]). 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.

Here, we study interactions of up to five bacterial species, which we have found to be the minimal set of consistently occurring, stably associated species in wild and laboratory *Drosophila melanogaster* fruit flies. In this work, we exploit the combinatorial geometry of the 5-cube. We examine 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. By this we mean, we compare the formulas of the following type:

$$u_{10100} = w_{00000} + w_{10100} - w_{10000} - w_{00100}$$
$$u_{10110} = w_{00010} + w_{10110} - w_{10010} - w_{00110},$$

where u_{10100} specifies an interaction between the species indicated in binary notation (10100 indicates the first and third species are present; see Fig. 1A and Box 1), and w_{10100} indicates the physiology trait score with the species indicated in binary. Generalizing, we extend the two-way interaction case up to the complete 5-way interactions along with combinatorial associations at intermediate diversity.

Clearly, extracting relevant and biologically meaningful interactions from among the many possible interaction coordinates and circuits must avoid redundant analyses. Therefore, in this paper we focus on the

interactions we find biologically interpretable and comparable with other studies (e.g. Newell and Douglas 2014 [9]).

In this Supplement, we give examples of how the mathematical approach applies to other biological systems specifically to bacterial interactions in the gut microbiome and explain how this approach generalizes to n-dimensional systems and to all lower rank interactions inside this system.

3 Glossary for interactions - Mathematical Terminology

The terminology we use is an adaptation of genetic epistasis 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). We refer to a bacterial combination within the n-species system using binary code. For instance, 00000 indicates no species present in a 5-species system, 111 indicates all species present in a 3-species system, and 1010 indicates the first and third species present in a 4-species system. Thus, each unique species is assigned an index with the binary string.

- **n-cube (n dimensional unit hypercube)**: is an n-dimensional generalization of a square (2-cube) or cube (3-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 [4].

- **phenotype**: is a measured, quantitative trait that is associated with a particular bacterial combination. In this work, we consider the number of bacterial CFUs, the development time of a fly from embryo to adult, the fecundity of a female fly, and the lifespan of a fly. We consider each phenotype separately, and we use w to refer generally to any phenotype. The phenotype associated with a specific bacterial combination is given by w_{XXX} where XXX is a binary string of length n referring to a bacterial combination.

- **Interaction coordinates**: are equations that describe the non-additivity of phenotypes associated with sets of species. For instance, if we consider two species, the interaction coordinate is just the degree to which w_{11} cannot be determined from knowing w_{00}, w_{01}, and w_{10}. We use u to refer generally to an interaction coordinate and we associate a specific interaction coordinate with its binary string. For instance, u_{11} indicates the interaction we are considering dependent upon w_{00}, w_{01}, w_{10}, and w_{11}. If $u_{11} = 0$, the phenotypes are completely additive, and we say there is no interaction. More generally, in the n-species system, interaction coordinates are given by linear combinations of the measured traits associated with each bacterial combination.

- **Circuits**: are certain linear combinations of interaction coordinates. In this sense, interaction coordinates form the basis elements for the interaction space, which can be more completely explored using circuits. Many different types of circuits exist, which we classify based on their symmetry groups. We assign a letter to each of these symmetry groups. For instance, the circuit b (described in a later section), is defined as the difference between u_{110} and u_{111} and it asks if the interaction between the first two species changes when the third species is added.

- **Triangulation**: is the local shape of the phenotypic landscape imposed by the interactions between bacterial species (see Box 1 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. An example is u_{11}, described in the definition for 'interaction coordinates'.

- **Contextual interactions (or contextual tests)**: are higher-order interactions (higher order here means interactions with more than 4 summands) arising as a generalization of the standard test. For instance, interaction coordinates for the 2, 3 and 4 and 5 species systems can be generalized by allowing the species not present to be occupied by bystanders, whose presence/absence is constant across the species considered in the standard test. Circuit b is an example of a contextual interaction of order three.

Summary of glossary section In this work, we focus on a 5 species system consisting of 32 bacterial combinations. We encode the different bacterial combinations by a fixed 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 defines a unique vertex of the 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. The phenotypes associated with each bacterial combination in binary notation are denoted simply by w_S. The order of the bacterial species are fixed and as described in the main text.

4 Multivariate linear model to detect interactions between bacteria

We use standard methods in R to calculate interactions (see Text for references). The mathematical formulas are provided here for completeness. Consider the Taylor expansion of the fitness landscape

$$f(x_1, \ldots, x_5) = \beta_0 + \sum_i \beta_i x_i + \sum_{i<j} \beta_{i,j} x_i x_j + \sum_{i<j<k} \beta_{i,j,k} x_i x_j x_k + \ldots$$

Here, x_i is the binary variable corresponding to the presence (1) or absence (0) of species i, for i from 1 to 5. The coefficients β's are hence the contributions of the corresponding interactions between species to the total fitness of the population. For example, $f(1,1,1,0,0)$ is the fitness of the population when only the first three species are present. This fitness is then decomposed into the sum of the contributions of the three species alone: $\beta_1+\beta_2+\beta_3$, plus the contribution of all pairwise interactions: $\beta_{1,2}+\beta_{1,3}+\beta_{2,3}$, plus the contribution of the three-way interaction: $\beta_{1,2,3}$. The rest of summands in the Taylor expansion of $f(1,1,1,0,0)$ are 0.

Given a set of estimates of the values of f by the data, we can fit the multiple linear regression model to estimate the coefficients β. This allows us to disentangle the contributions of the species and those of the interactions between the species.

We have implemented this method at `https://github.com/gavruskin/microinteractions/blob/master/taylor_lin_fit.ipynb`, where all necessary python code to reproduce the analysis can be found.

Note that in the code we denote the parameters by $a, b, c, d, \ldots$ instead of β's.

5 Two disjoint families of interactions: standard and contextual tests

In the following, we formally define two disjoint families of interaction formulas (tests) in the five bacterial species setting. The first family, standard tests, is smaller in size than the second, contextual tests. Standard tests allow us to determine whether the phenotype of a group of bacterial species can be computed from the sum of its parts. These tests can be linked the coefficients of the multivariate linear regression method presented in Section 4 (main text Figure S10).

The family of contextual tests describe how standard test change according to the presence of additional bacterial species and includes circuits [2]. These two families of test we define in this work are new and inspired by [1] and [6] (the second reference for standard tests).

5.1 Standard tests

Consider the two species 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 bacterial combination 00, which is germ-free. Similarly, w_{11} (both species), w_{01} (one species) and w_{10} (the other species). Biologically, the meaning of this formula is well understood: it compares the

phenotype contributions of the two-species association with the single species associations. Geometrically, the summands of u_{11}, which are w_{00}, w_{11}, w_{01} and w_{10}, can be indexed by the four vertices of a unit 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 this work, 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 of the terms change according to equation (4). As previous authors have described, this sign change results from a Fourier transform, see [2]. Biologically, one can say that these tests compare the phenotypes of bacterial combinations when an even number of bacterial species are present versus when an odd number of bacterial species are present (e.g. w_{00} and w_{11} vs w_{01} and w_{10}).

Quantifying the number of tests within symmetry groups of the 5-cube. Examining the symmetries of the 5-cube, we can see that there are 10 possible 2-way interactions, u_{11}. Similarly there are 10 different 3-way interactions and five 4-way interactions. Together, this approach gives the

$$\binom{5}{2} + \binom{5}{3} + \binom{5}{4} + \binom{5}{5} = 26 \tag{1}$$

different standard tests in a five bacterial species system. The number of standard tests we find in Equation (1) matches the results of [6].

To be explicit, the standard tests 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 tests arise from the 2-way interaction and always involve two bacterial 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 test involving three species out of the five are:

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

These tests involve different combinations of vertices of G and define three dimensional cubes. The following five standard tests arise from the 4-way interaction described by u_{1111} above and involve four species out of the five, leaving out the remaining bacterial species:

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

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:

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

Geometrically, this test involves all vertices of G. In Figure 1 below we highlight the regions delimited by the vertices defining the above 26 standard tests inside a projection of the five cube G.

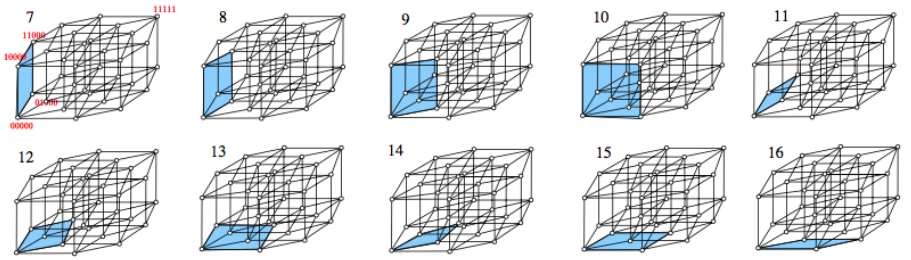
For example, in a system with three bacterial species consisting of the following 8 bacterial combinations $G = \{000, 001, 010, 100, 110, 011, 101, 11\}$ there are three standard tests involving two out of three species:

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

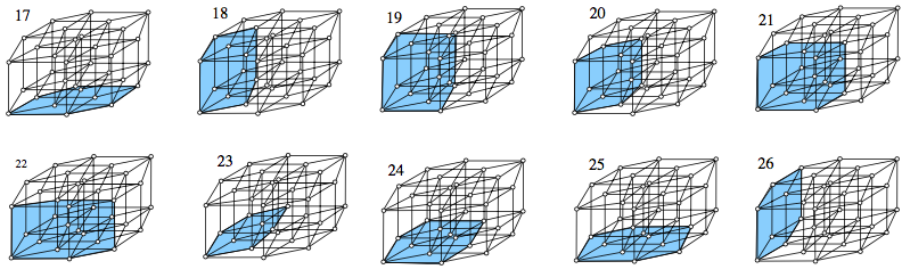
together with the 3-way interaction u_{111}, described above.

On the other hand, in an n-species system there are

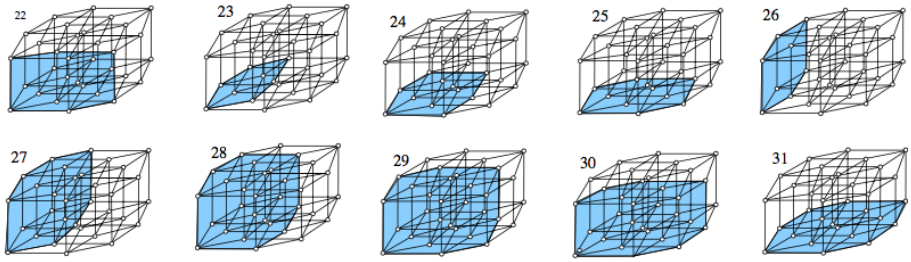
$$\sum_{2 \leq l \leq n} \binom{n}{l} \tag{2}$$



(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 involved in the corresponding test. The interaction u_{11111} is defined by all 32 vertices and therefore omitted.

standard tests.

5.2 Contextual tests

In the following, we compare the results of the standard tests with the results of the contextual tests in a five species system and for various phenotypes. The contextual tests include **interaction coordinates**, which are generalizations of the standard test, and **circuits**, which are linear combinations of the interaction coordinates.

In an n-species system there are

$$\sum_{2\leq l\leq n}\binom{n}{l}2^{n-l}(2^l-l-1)+\binom{n}{3}2^{n-3}\cdot 14-\sum_{2\leq l\leq n}\binom{n}{l} \tag{3}$$

different contextual test. Compare with Equation (2) and (6).

In this section, we make precise the notion of interaction coordinates and circuits, see also Glossary [2] for binary systems of n-bacterial species, $\{0,1\}^n$. We first describe these formulas abstractly for arbitrary values of n and later focus on the case where there are only 5 bacterial species.

5.3 Interaction coordinates

For a fixed $n\in\mathbb{N}$, let $j_1,j_2,\ldots,j_n\in\{0,1\}^n$ be a binary strings of lengths n. We view each such string as a vertex of an n-dimensional cube. Let $i=i_1,i_2,\ldots,i_n\in\{0,1\}^n$ be such a binary string 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:

$$u_{i_1,i_2,\ldots,i_n}:=\frac{1}{2^n-1}\sum_{j_1=0}^{1}\sum_{j_2=0}^{1}\cdots\sum_{j_n=0}^{1}(-1)^{i_1j_1+i_2j_2+\cdots+i_nj_n}w_{j_1j_2\ldots j_n} \tag{4}$$

where $w_{\_}$ are values of a corresponding phenotype and indexed by the vertices of the n-dimensional cube. There are 2^n-n-1 interaction coordinates. Moreover, these coordinates are linearly independent and form a vector space basis of the interaction space. Interaction coordinates include the so called higher-order interactions arising by introducing the species not present under the lower dimensional standard tests. For instance, we previously noted that there are 10 possible pairs of species in the 5-species system. For any of these pairs, we can consider how the interaction between the pair changes in the presence of a third, fourth, or fifth species.

5.4 Circuits

Certain linear combinations of interaction coordinates give rise to circuits, that is minimal dependency sets of configurations of points in a space of dimension n, see [2]. Among all possible circuits, we will focus on circuits that have a simple biological meaning and that originate from the following circuits defined for the

3-cube in [2]:

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

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

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

Linear combinations of these circuits, which we do not consider here, yield more interactions contained in the interaction space. Biological and geometric interpretations of these circuits were first given in [2, §.3] for the genetic setting. For instance, circuits m to t relate the three-way interactions to the total two-way interactions. Examining the equation for m, we see that it is equal to the sum of the phenotype terms for each of the single species combinations minus the sum of the three-species association and the germ free flies. Similarly in n, the sum of the two-species combinations is compared to the sum of the three-species combination and the germ free flies. The biological interpretation is that m tells us whether the single species associations predict the three-species combination, and n tells us whether the two-species associations predict the three-species combination. Of note, the signs of the circuits m to t do not have a two-locus interpretation, making them truly of higher-order.

Later we will see that the above circuits become circuits of the 5-cube,in a similar way as we described above for standard tests. Varying the presence, resp. absence, of the bystander species gives rise to a new class of circuits inside the 5-cube which have an established biological meaning.

5.5 Recursively constructing higher order interactions from lower order interactions

We now describe interaction coordinates in lower dimensional cubes and how the circuits a-t extend to interaction formulas in higher dimensional cubes. In light of the data analyzed in this paper, we focus on the case of five bacterial species. However, the approach we present here easily extends to systems with n-bacterial species. See discussion at the end of this section.

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.

$$\begin{aligned}
\alpha_{**klm} &= w_{00klm} + w_{11klm} - w_{01klm} - w_{10klm} \\
\alpha_{*k*lm} &= w_{0k0lm} + w_{1k1lm} - w_{0k1lm} - w_{1k0lm} \\
\alpha_{*kl*m} &= w_{0kl0m} + w_{1kl1m} - w_{0kl1m} - w_{1kl0m} \\
\alpha_{*klm*} &= w_{0klm0} + w_{1klm1} - w_{0klm1} - w_{1klm0} \\
\alpha_{k**lm} &= w_{k00lm} + w_{k11lm} - w_{k01lm} - w_{k10lm} \\
\alpha_{k*l*m} &= w_{k0l0m} + w_{k1l1m} - w_{k0l1m} - w_{k1l0m} \\
\alpha_{k*lm*} &= w_{k0lm0} + w_{k1lm1} - w_{l0lm1} - w_{k1lm0} \\
\alpha_{kl**m} &= w_{kl00m} + w_{kl11m} - w_{kl01m} - w_{kl10m} \\
\alpha_{kl*m*} &= w_{kl0m0} + w_{kl1m1} - w_{kl0m1} - w_{kl1m0} \\
\alpha_{klm**} &= w_{klm00} + w_{klm11} - w_{klm01} - w_{klm10},
\end{aligned} \tag{5}$$

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. Similarly, extending to the five species system, the four interaction coordinates u_{111}, $u_{110}, u_{011}, u_{101}$ are defined by (4) above in the following fashion:

$$\begin{aligned}
\beta_{***kl} &= \sum_{j_1=0}^{1}\sum_{j_2=0}^{1}\sum_{j_3=0}^{1}(-1)^{*j_1+*j_2+*j_3} w_{j_1 j_2 j_3 kl} \\
\beta_{**k*l} &= \sum_{j_1=0}^{1}\sum_{j_2=0}^{1}\sum_{j_4=0}^{1}(-1)^{*j_1+*j_2+*j_4} w_{j_1 j_2 k j_4 l} \\
\beta_{**kl*} &= \sum_{j_1=0}^{1}\sum_{j_2=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_1+*j_2+*j_5} w_{j_1 j_2 k l j_5} \\
\beta_{*k*l*} &= \sum_{j_1=0}^{1}\sum_{j_3=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_1+*j_3+*j_5} w_{j_1 k j_3 l j_5} \\
\beta_{*kl**} &= \sum_{j_1=0}^{1}\sum_{j_4=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_1+*j_4+*j_5} w_{j_1 k l j_4 j_5} \\
\beta_{k*l**} &= \sum_{j_2=0}^{1}\sum_{j_4=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_2+*j_4+*j_5} w_{k j_2 l j_4 j_5} \\
\beta_{kl***} &= \sum_{j_3=0}^{1}\sum_{j_4=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_3+*j_4+*j_5} w_{k l j_3 j_4 j_5} \\
\beta_{*k**l} &= \sum_{j_1=0}^{1}\sum_{j_3=0}^{1}\sum_{j_4=0}^{1}(-1)^{*j_1+*j_3+*j_4} w_{j_1 k j_3 j_4 l} \\
\beta_{k**l*} &= \sum_{j_2=0}^{1}\sum_{j_3=0}^{1}\sum_{j_5=0}^{1}(-1)^{*j_2+*j_3+*j_5} w_{k j_2 j_3 l j_5} \\
\beta_{k***l} &= \sum_{j_2=0}^{1}\sum_{j_3=0}^{1}\sum_{j_4=0}^{1}(-1)^{*j_2+*j_3+*j_4} w_{k j_2 j_3 j_4 l}.
\end{aligned}$$

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 $kl = 10$. The biological significance of these interactions is examined in the main text Figs. 5 and S12.

We also 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.

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

Thus in a 5-species system, there are

$$\sum_{2\leq l\leq 5} \binom{5}{l} 2^{5-l}(2^l - l - 1) = 376$$

interaction coordinates together with

$$\binom{5}{3} 2^{5-3} \cdot 20 = 800$$

possible extensions of the circuits a-t. Notice however that the circuits a-f are two way interactions extended to the three species setting, hence only the circuits g-t provide new possible extended circuits. It follows, that only 560 of these extended circuits are different among each other and disjoint from the previous interaction coordinates. Thus, together this approach yields 936 different extended interaction coordinates and circuits.

In an n-species system, the approach described above gives

$$\sum_{2\leq l\leq n} \binom{n}{l} 2^{n-l}(2^l - l - 1) + \binom{n}{3} 2^{n-3} \cdot 14 \tag{6}$$

different extended interaction coordinates together with the disjoint set of all possible and different extensions of the circuits g-t. As for the five species setting, in equation (6) we omit the circuits a-f.

The biological and geometric interpretation of the interactions enumerated in equation (6) can be deduced from the lower dimensional interpretations given in [2], for the three species case $\{0, 1\}^3$. Clearly, more interactions might be obtained by considering linear combinations of the above formulas.

6 Significance testing of interactions

Despite the standard and contextual tests being disjoint families, the two sets of tests are not statistically independent as they derive from the same underlying data sets. In order to determine whether a test is non-zero, we took the uncertainty in the phenotype measurements into account and devised a statistical test as follows. Since the phenotype measurements were computed together with their standard errors, we can compute the corresponding propagation of error for each test (standard and contextual). This error was determined by taking the square root of the sum of squares of the standard errors involved in each test. 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 of error, that is, the formula involving the least number of operations. For instance, each circuit test $a, \ldots, t$ can be defined in two ways, one involving a difference between interaction coordinates and a direct way see [2]. The direct way involves less operations, and therefore a smaller propagated error.

We then determined significance in the following way. We assumed that each interaction formula (test) comes from a Gaussian distribution with mean 0 and standard deviation given by the propagated 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%.

To account for the multiple comparisons, we corrected all the above p-values using the Benjamini-Hochberg multiple testing correction procedure. In this way, we are able to control the false discovery rate at 5%; that is, we eventually considered an interaction statistically significant if the corrected p-value falls below 0.05.

7 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 standard tests and on the second part we compare the outcomes of these tests with contextual tests. In the last part, we focus on specific standard and contextual tests which examine how these test change as the number of species increases.

7.1 Computed standard tests

Nr.	Combination	Daily Fecundity	Development	Bacterial load	Time to death
7	u_{**000}	**1.19**	**-0.67**	**-294036**	5.23
8	u_{*0*00}	1.15	**-0.71**	**279946**	3.05
9	u_{*00*0}	0.17	**-0.80**	-58439	**6.09**
10	u_{*000*}	0.51	**-0.75**	39002	**6.92**
11	u_{0**00}	0.55	-0.34	50689	-0,72
12	u_{0*0*0}	-0.58	0.16	-127406	-1.77
13	u_{0*00*}	0.74	0.08	-83144	4.70
14	u_{00**0}	**-1.40**	-0.42	-142987	2.92
15	u_{00*0*}	**-1.44**	0.00	**-139566**	**8.58**
16	u_{000**}	**-1.41**	-0.09	**-253460**	**5.32**
17	u_{***00}	**1.73**	-0.63	279224	3.07
18	u_{**0*0}	-0.65	-0.63	78573	6.81
19	u_{**00*}	**1.65**	**-0.96**	40900	**9.24**
20	u_{*0**0}	0.14	**-0.92**	163152	5.15
21	u_{*0*0*}	1.49	**-0.92**	**252978**	**8.06**
22	u_{*00**}	0.27	**-0.88**	-154624	**9.23**
23	u_{0***0}	-1.16	-0.30	155885	0.9
24	u_{0**0*}	0.40	-0.21	-7395	7.37
25	u_{0*0**}	-0.51	-0.25	-80465	0.42
26	u_{00***}	-1.61	-0.59	-153646	4.08
27	u_{****0}	-0.74	-0.63	**525575**	6.69
28	u_{***0*}	**2.53**	-0.96	268147	**13.48**
29	u_{**0**}	0.42	**-1.04**	105278	**14.82**
30	u_{*0***}	0.72	**-1.17**	-24224	5.89
31	u_{0****}	-1.00	-0.59	130387	0.11
32	u_{*****}	0.95	**-1.42**	373239	**13.35**

Table 1: **Result for the 26 standard tests** computed on the raw fecundity per day (daily fec), bacterial load (CFU), time to death (time d) and development rate (dev) data. The first two columns indicate the performed standard tests separated according to the number of bacterial species present in the fly gut (the type of species is indexed by the position of the symbol $*$). Results in bold reached statistical significance ($p < 0.05$) after BH-multiple comparison correction. The geometric description of these tests is illustrated in Figure 1.

To summarize the results of the standard test presented above in Table 1 we averaged the results of all the standard 2-, 3-, 4- and 5-way interaction at each level of species diversity (Table 2). 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, we calculated a_3, a_4 and a_5.

Term	CFU	dev	fd	td	unif	nor
a_2	-72940.29	-0.353	-0.053	4.083	-0.197	1.477
a_3	57458	-0.628	0.174	5.433	-0.112	1.585
a_4	201032	-0.878	0.385	8.198	0.170	2.216
a_5	373239	-1.419	0.947	13.35	0.650	4.080

Table 2: The average interactions terms $a_2, \ldots, a_5$ for the standard tests and the four phenotypes of number of bacterial cells per fly (CFU), development time in days (dev), fecundity per day (fd), time to death in days (td), as well as for synthetic data from a uniform distribution on $[0, 1]$ and from a standard normal distribution.

From Table 2 we see that the average 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 4-way interactions (or at even lower dimensions). Together with the results of Table 2, we also consider the average interaction 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	unif	nor
n_2	-36470	-0.177	-0.027	2.04	-0.099	0.739
n_3	19153	-0.209	0.058	1.81	-0.037	0.528
n_4	50258	-0.220	0.096	2.05	0.043	0.554
n_5	74648	-0.284	0.189	2.67	0.130	0.816

Table 3: The normalized terms $n_2, \ldots, n_5$ for the standard tests, normalized by the number of bacteria present for the four phenotypes of number of bacterial cells per fly (CFU), development time in days (dev), fecundity per day (fd), time to death in days (td), as well as for synthetic data from a uniform distribution on $[0, 1]$ and from a standard normal distribution.

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 cannot predict the phenotype of the microbiome when a new species is added, the microbiome becomes 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 results of the interaction tends to increase together with its statistical significance, in this sense the result of the interaction tends to become more predictable. 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 contextual test will be additive (versus showing an 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 individual species themselves.

7.2 Comparing the relative importance of individual bacterial species versus their interactions in determining physiological traits

When we compare the distributions of the raw data with the results of the standard tests, we see that linear trends in the raw data do not necessary translate to the same trends in the outcomes of standard test (see Fig. S13). 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 (see Fig. 2). However, the standard tests indicate that interaction magnitude tends to increase with the number of species involved (see Table 2), and hence with the numbers of terms occurring in the standard tests. Finally, from the data in the Fig. S13, we deduce that the above observations remain valid by considering the (fewer) standard tests which reached statistical significance rather than considering all of them. For completeness, we also computed Spearman's correlation coefficients on all 26 standard tests, and on the raw CFUs, development rate, daily fecundity, time to death data. See Table 4. The tests in bold in Table 4 reached significance at $p < 0.05$. See Table 5 for the results of the normality test (Shapiro-Wilk).

	CFU	**Development**	**Daily fecundity**	**Time to death**
CFU	-	**-0.51** (p=0.0028)	0.12(p=0.5031)	**-0.49**(p=0.0242)
Development rate	**-0.43** (p=0.0272)	-	-0.28 (p=0.1246)	**0.40** (p=0.0242)
Daily fecundity	0.39 (p=0.0463)	**-0.56** (p=0.0031)	-	**-0.41** (p=0.0202)
Time to death	0.15 (p=0.4622)	**-0.61** (p=0.0010)	0.32 (p=0.1082)	-

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

7.3 Results for all computed interactions from Figs. 4, 5, and S12

In this section, we describe the results contained in Figs. 4, 5, and S12 in the main text, obtained from computing the 910 contextual tests together with 26 standard tests for the phenotypes of CFUs, development rate, daily fecundity, time to death. We found many significant positive and negative interactions among bacterial species in fruit flies. Moreover, we found that for some phenotypes (CFUs and daily fecundity) the standard tests fully capture the interaction trends measured by the contextual tests However, for development rate and time to death we found that many new significant tests arise when considering contextual tests, indicating that the impacts of the bacterial community on the fly depend more on the context of which other bacterial species are present. Finally, comparing the Table 4 with Table 6 we found correlated interactions between phenotypes where neither the raw measurements nor the standard tests were correlated. The complete correlations between the measured phenotypes for the same interactions are shown in Figure S13. This finding suggests that interactions more than the individual species themselves shape host physiology.

One of the major findings of this work is that interactions between bacteria and their effects on the host are highly dependent on context. By comparing the standard tests to the contextual tests (*e.g.* Figs. 4, 5 and S12), we quantitatively demonstrated this point for specific combinations of bacteria. Here, we extend that analysis to compare the two probability distributions corresponding to standard and contextual tests to ask whether they come from the same continuous distributions. The set of contextual tests does not include the set of standard tests we computed. However, since standard tested and contextual tests are computed from the same underlying data, the two sets of tests are statistically dependent. To test the difference between the distributions, we performed a two-sample and two-sided Kolmogorov-Smirnov (KS) test. The results of the KS test for the phenotypes of bacterial CFUs (D = 0.26264, p-value = 0.06117) and daily fecundity (D = 0.17912, p-value = 0.392) indicate that there is little evidence to reject the null-hypothesis of standard tests and contextual tests coming from the same distributions. On the other hand, the results of the KS test for the phenotypes of development time (D = 0.58022, p-value = $8.1 \cdot 10^{-08}$) and time to death (D = 0.46681,

p-value $= 1.252 \cdot 10^{-05}$) suggest that it is unlikely that the standard tests and contextual tests come from the same distributions. Thus, for these phenotypes, the difference between the standard and contextual tests strongly supports the notion that interactions are highly dependent on context not just for specific cases but also on a global scale.

For completeness, we also ask if any of the standard tests, contextual tests and significant standard tests come from normal distributions. To test this hypothesis we performed a Shapiro-Wilk test. The results are summarized in Table 5. We see that for the contextual tests (significant after multiple testing) the Shapiro-Wilk tests reach significance. We then conclude that the corresponding distributions fail the normality test.

	W	p**-value**
CFU	0.94732	$2.2 \cdot 10^{-16}$
CFU STD	0.97287	0.6985
CFU significant	0.91657	$8.131 \cdot 10^{-9}$
Development	0.97131	$2.092 \cdot 10^{-12}$
Development STD	0.97793	0.8272
Development significant	0.87521	0.0007663
Daily fecundity	0.96783	$2.727 \cdot 10^{-13}$
Daily fecundity STD	0.96159	0.4238
Daily fecundity significant	0.88623	$1.58 \cdot 10^{-07}$
Time to death	0.99524	0.006241
Time to death STD	0.96781	0.5673
Time to death significant	0.87586	$9.06 \cdot 10^{-09}$

Table 5: **Shapiro-Wilk-Test**. For each phenotype we summarized the test statistic and the corresponding p-value. The p-values for all significant contextual tests (significant after adjusting p-values with the BH-multiple testing correction procedure) 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 tests where standard, whereas if 'STD' is not appended, it is for the contextual tests. If 'significant' is appended, it means that the contextual tests we consider are the ones that reached statistical significance ($p < 0.05$) after BH-multiple testing correction.

	CFU	**Development**	**Daily fecundity**	**Time to death**
CFU	-	**-0.18** ($p < 0.0005$)	**0.18** ($p < 0.0005$)	**-0.13** ($p < 0.0005$)
Development	**-0.97** ($p < 0.0005$)	-	**-0.37** ($p < 0.0005$)	-0.09 ($p = 0.0037$)
Daily fecundity	**0.54** ($p < 0.0005$)	**-0.91** ($p < 0.0005$)	-	**-0.09** ($p = 0.0048$)
Time to death	**-0.78** ($p < 0.0005$)	-0.10 (p= 0.8729)	**-0.84** ($p < 0.0005$)	-

Table 6: **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 5 to 187 pairwise complete comparisons). Similarly, above the diagonal we compute correlations for all interactions regardless of statistical significance (936 data points).

8 Discussion of Interaction Tests

Overall, when we utilize all of the interactions tests, including the contextual tests, we see more significant interactions for the different phenotypes. Consistently, we see that the same interactions 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 call 'standard tests') and compared them with the results of 910 (different) 'contextual' interactions. We found that for certain phenotypes the standard interactions approximate well the distribution of the other more involved contextual tests. 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 tests 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 mostly small. Moreover, analyzing smaller sets of particularly expressive and biologically interpretable interactions would not only be computationally more efficient, but also would facilitate the comparisons of higher order tests arising in different biological contexts (for example infected fruit flies, or fruit flies treated with antibiotics). The approach we propose here to define higher order interactions, as well as the computations we carried out, easily extends to settings with more species. Thus, our methodology can provide a way to reduce the experimental burden of examining interactions in the microbiome involving fewer experiments.

We also note that our computations are based on discrete data points. However, 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, and it would be interesting to develop a quantitative statistical framework better accounting for possible sensible noise in the data set.

A challenge in the microbiome is to develop a quantitative framework that applies both to the detailed interactions between single species as well as to complex assemblages containing hundreds of species. We believe that our approach is simple enough to be scalable to higher dimensions, and by identifying groups of species that behave as single loci (as detected in specific contextual tests, which are lower dimensional projections of the n-cube), we can reduce the dimensionality of complex assemblages through this combinatorial framework.

9 Averaging model for prediction of high-diversity traits with low-diversity measurements (Figure 2)

9.1 Single-species mixing model

In Figure 2 of the main text, the measured traits appear to converge as bacterial diversity increases. To determine whether this convergence could be attributed to a mixing effect, we constructed a presence-absence "averaging model," in which the trait of a bacterial combination is predicted to be the average of the single-bacteria traits. Formally, we assume that a trait f is a function of the types of bacteria in the microbiome S, and that this microbiome is composed of bacteria s_i for $i \in 1, \ldots, 5$ such that the microbiome can be decomposed into these individual species. For example, $S = 10110$ would contain the species s_1, s_3, and s_4. As in the rest of the Math Supplement, the components $i = 1, \ldots, 5$ respectively correspond to the bacteria LP, LB, AP, AT, and AO. The diversity N of a bacterial combination is given by the cardinality of S, $N = |S|$.

Then, the averaging model predicts that

$$f(S) = \frac{1}{N} \sum_{s_i \in S} f(s_i). \tag{7}$$

This allows us to predict the traits of higher-diversity bacterial combinations as the average of single-species traits (e.g. f(10001) = 1/2 (f(10000) + f(00001))). With this model we predict fly fecundity, time to death,

development time, and bacterial load, and compare the predicted values to the measured experimental values. We display the difference between the predicted and measured values in Figure 2 of the main text.

To evaluate whether this averaging model captures the observed trait measurements, we asked how often the prediction over or underestimated the measured value. If there was a significant tendency to over or underestimate, this would indicate that the mixing model is not sufficient to predict the data. We found that this model significantly under-predicted daily fecundity (20 of 26 negative, p=9e-3, binomial test) and bacterial load (26 of 26 negative, p=3e-8, binomial test), and significantly over-predicted time to death (22 of 26 positive, p=5e-4, binomial test) and development time (25 of 26 positive, p=8e-7, binomial test).

9.2 Pairwise species mixing model

Next, we generalized the averaging model of the previous subsection by averaging the traits of pairs of species, rather than the traits of individual species. We decompose a microbiome S into its constitutive pairs of bacteria r_{ij}, so that if a microbiome consists of the N bacteria, then for all $N(N-1)/2$ pairs of those bacteria s_i and s_j that exist in the microbiome, the element r_{ij} will be in S. Then, the pairwise averaging model predicts that a trait $f(S)$ will be

$$f(S) = \frac{2}{N(N-1)} \sum_{r_{ij} \in S} f(r_{ij}). \tag{8}$$

For example, the pairwise averaging model would predict $f(11001) = \frac{1}{3}(f(11000) + f(10001) + f(01001))$. In Figure 2 of the main text we compare the predictive ability of the pairwise averaging model to that of the single-species averaging model.

9.3 Comparison of single-species and pairwise mixing model predictions

In Figure 2 we display the difference between experimentally measured traits and the mixing model predictions (for the single-species and pairwise cases). We plot 95% confidence intervals for each difference between experiment and prediction. If the confidence interval of this difference does not include 0, then that prediction significantly deviates from the observed measurement. By counting how many trait measurements were or were not captured by the 95% confidence intervals of each model, we achieve a measure of the accuracy of each model. In order to compare the two models, we only consider their predictive ability for diversity 3 or larger.

We found that, across all traits, the single-species mixing model 95% confidence interval captured 28 out of the 64 measured traits (44%), while the pairwise species mixing model captured 50 out of 64 measurements (78%), which indicates that the pairwise model was significantly better at predicting trait observations (p=5e-5, Fisher's exact test, n=64). For individual traits, the two models performed similarly in predicting daily fecundity (single-species predicted 14 of 16 experiments; pairwise species predicted 15 of 16; p=1, Fisher's exact test, n=16) and time to death (single-species predicted 9 of 16 experiments; pairwise species predicted 9 of 16; p=1, Fisher's exact test, n=16), but the pairwise model outperformed the single species model in predicting development time (single-species predicted 4 of 16 experiments; pairwise species predicted 15 of 16; p=8e-5, Fisher's exact test, n=16) and bacterial load (single-species predicted 1 of 16 experiments; pairwise species predicted 11 of 16; p=3e-4, Fisher's exact test, n=16).

9.4 Comparison of single-species and pairwise mixing model errors

We found that for the experiments of diversity 3, 4, and 5, the pairwise model predicts the average error of the daily fecundity to be .322 eggs (SE=.027 eggs), time to death to be 2.587 days (SE=.141 days), development time to be .140 days (SE=0.012 days), and bacterial load to be 123409 CFUs (SE=13646 CFUs). The single-species mixing model predicts average errors of the daily fecundity to be .436 eggs (SE=.304 eggs), time to death to be 3.396 days (SE=1.956 days), development time to be .488 days (SE=0.154 days), and bacterial load to be 322118 CFUs (SE=102579 CFUs). Therefore, we found that the pairwise averaging model better captured the measured traits than the single-species model (daily fecundity p=.149, time to death p=.120, development time p=4e-8, bacterial load p=1e-6; Welch's t-test, n=16).

9.5 Coefficient of variation of traits at increasing diversity

Finally, we examined how the variation in traits changes at increasing diversities. We found that the standard errors of the daily fecundity and bacterial load are significantly positively correlated with increasing diversity (fecundity p=0.01, bacterial load p=3e-3, Wald test) and the standard error of time to death is significantly negatively correlated with increasing diversity (p=0.01, Wald test).

The coefficient of variation of the net bacterial load had a significant negative correlation with increasing diversity (p=0.02, Wald test), as indicated in Figures S15 and S16. However, the coefficient of variation of the individual species CFU counts is relatively constant over increasing diversity, as shown in Figure S15. This indicates that the composition of higher-diversity microbiomes consists of bacterial species that are less variable, and this in turn suggests that higher-order interactions serve a stabilizing role for the microbiome.

10 Pairwise species interactions (Figure 6)

Each of the 32 fully combinatorial experiments has 24 biological replicates, and for each biological replicate we collected the CFU abundance data by averaging three technical replicates. The raw data are displayed in Figure S6. We segmented this collection of CFU counts according to its bacterial diversity (ranging from 1 to 5), and studied how microbiome properties changed for different diversities. For a given bacterial combination, not all of the introduced bacteria were detected in every replicate, due to the limit of detection of our method. In the following measurements, we exclude replicates in which one or more species was undetected. By excluding these replicates, we intend to capture the deterministic aspect of interactions between bacteria, rather than the stochastic aspect that corresponds to the variability in colonization. However, this exclusion also reduces the available samples for analysis— in the most extreme case only 4 biological replicates remained (Figure S15D,E).

We computed the interaction strength between pairs of bacteria following Paine's method [10] and measured the asymmetry of these interactions. We also used a generalized Lotka-Volterra model to fit bacterial interactions from the CFU data, and found that the inferred interactions qualitatively matched those derived from Paine's method. Lastly, we calculated correlations between pairs of bacteria.

10.1 Bacterial interactions determined by Paine's method (Figure 6B,C, S15A,B)

Paine [10] presented a model-free calculation of interaction strength, which we implemented to probe bacterial interaction strength at low diversity (1 and 2 species) and high diversity (4 and 5 species). Note that this method for 3 and 4 species diversity is less-straightforward to implement and is omitted here for simplicity. Let $(\mathrm{y}^{+j})_i$ and $(\mathrm{y}^{-j})_i$ be the abundance of species i in the presence and in the absence of the j community. Then, to measure the effect of a single species j on another species i, Paine measured [10]

$$\frac{(\mathrm{y}^{+j})_i - (\mathrm{y}^{-j})_i}{(\mathrm{y}^{-j})_i}. \tag{9}$$

This value is bounded below by -1 (introduction of j eliminates i), negative values indicate i is inhibited by j, and positive values indicate i is increased by j. We consider a rescaling of this value that has a symmetric range, which we call M_{ij}, that is given by

$$M_{ij} = \log_2\left(1 + \frac{(\mathrm{y}^{+j})_i - (\mathrm{y}^{-j})_i}{(\mathrm{y}^{-j})_i}\right). \tag{10}$$

We compute this value at low ($1 \to 2$) and at high ($4 \to 5$) diversities, where we define the diversity of an experiment as the number of bacterial species that are in the food. For example, to compute M_{12} at low diversities, we consider experiments 10000 and 11000; to compute M_{12} at high diversities we consider experiments 10111 and 11111. Since there are many biological replicates for each experiment, we can bootstrap our samples following the method of Efron and Tibshirani [14] in order to compute the mean and standard error for each interaction value M_{ij}. We use the mean interaction values to populate the interaction matrix M, which we display as a directed graph in the low (Figure 6B) and high (Figure 6C) diversity cases. In the tables below, we also report the standard deviation of the distribution for each interaction (that

is, we show the mean(M_{ij}) $\pm$ standard error(M_{ij})). The interaction matrices for high and low diversity interactions are

$$M^{\text{high}} = \begin{pmatrix} 0.00 \pm 0.00 & \text{-}0.02 \pm 0.53 & \text{-}0.03 \pm 0.58 & \text{-}0.56 \pm 0.50 & \text{-}0.72 \pm 0.50 \\ \text{-}2.48 \pm 0.88 & 0.00 \pm 0.00 & \text{-}0.37 \pm 1.08 & \text{-}1.59 \pm 0.98 & \text{-}0.42 \pm 1.13 \\ \text{-}0.93 \pm 0.81 & 0.10 \pm 0.61 & 0.00 \pm 0.00 & \text{-}0.44 \pm 0.70 & \text{-}1.98 \pm 0.94 \\ \text{-}0.68 \pm 0.75 & 0.12 \pm 0.84 & \text{-}0.41 \pm 0.78 & 0.00 \pm 0.00 & \text{-}0.36 \pm 1.00 \\ \text{-}0.14 \pm 0.63 & 0.91 \pm 0.91 & \text{-}0.18 \pm 0.70 & \text{-}0.38 \pm 0.64 & 0.00 \pm 0.00 \end{pmatrix} \tag{11}$$

and

$$M^{\text{low}} = \begin{pmatrix} 0.00 \pm 0.00 & 0.59 \pm 0.45 & 1.23 \pm 0.41 & 0.69 \pm 0.47 & 0.93 \pm 0.40 \\ \text{-}1.14 \pm 0.54 & 0.00 \pm 0.00 & \text{-}0.51 \pm 0.51 & 0.06 \pm 0.33 & 0.19 \pm 0.36 \\ 1.09 \pm 0.69 & 1.00 \pm 0.65 & 0.00 \pm 0.00 & \text{-}0.28 \pm 0.74 & \text{-}0.56 \pm 0.63 \\ \text{-}0.37 \pm 0.62 & 0.24 \pm 0.62 & 0.51 \pm 0.84 & 0.00 \pm 0.00 & 0.13 \pm 0.51 \\ 0.14 \pm 0.27 & 0.08 \pm 0.30 & 0.28 \pm 0.39 & 0.34 \pm 0.33 & 0.00 \pm 0.00 \end{pmatrix}. \tag{12}$$

The interactions as defined in Eq. (10) need not be— and generally are not— symmetric. We compute their asymmetry with a metric used by Bascompte et al. [13],

$$AS(i,j) = \frac{|M_{ij} - M_{ji}|}{\max(|M_{ij}|, |M_{ji}|)}. \tag{13}$$

This metric ranges from 0 (perfectly symmetric) to 2 (exact opposites). We consider the mean asymmetry of all 10 pairs. For the low diversity case this mean asymmetry is 1.04 (SD = 0.13), and for the high diversity case this mean asymmetry is 0.77 (SD = 0.08). To estimate the standard deviation, we repeatedly permuted the underlying interaction matrix M and created a distribution of permuted mean asymmetry values and used that standard deviation.

10.2 Bacterial interactions fit by a generalized Lotka-Volterra model (Figure S15D,E)

We infer the species interactions by assuming that the system obeys the generalized Lotka-Volterra equations,

$$\frac{d}{d\text{t}} x_i = x_i \left(\mu_i + \sum_{j=1}^{N} M_{ij} x_j \right) + \sum_{i}^{T} \delta(t - t_i^*) v_i, \tag{14}$$

with growth rate μ, interaction matrix M, and pulsed "feeding" of v_i at times $t_1^*, \ldots, t_T^*$.

Previous experiments have shown that when exposed to a steady supply of bacteria-infused food the fly gut approaches equilibrium within 5 days [3], and in this experiment the flies have been feeding for 10 days (see Materials and Methods). Therefore, we assume that the CFU counts of each experiment are measured at equilibrium, and we assume that the median of each combination's CFU counts is the steady state solution to Eq. (14). We additionally assume that the microbiome returns to equilibrium quickly after feeding, so that we may neglect the $\delta(t - t_i^*)v_i$ term in Eq. (14).

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

$$0 = \mu_i + \sum_{j=1}^{N} M_{ij} \tilde{x}_j. \tag{15}$$

If there are m_i experiments of diversity i, then there are $\sum_{i \in D} m_i i$ equations that must be simultaneously satisfied for diversities D. To match the previous interaction calculations, we consider a low-diversity group $D_{\text{low}} = \{1,\ 2\}$ and a high-diversity group $D_{\text{high}} = \{3,\ 4,\ 5\}$, which separates low-order interactions (2-species) from high-order (3-, 4-, and 5-species). For each group, we can rewrite the linear equations of

Eq. (15) in matrix form as $0 = A\vec{y}$, where A is made up of the $\tilde{x}_i$ and is of the form

$$A = \begin{pmatrix} \tilde{x}_1^{10000} & 0 & \cdots & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & \tilde{x}_2^{01000} & \cdots & 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{x}_1^{11000} & \tilde{x}_2^{11000} & \cdots & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \cdots & \tilde{x}_1^{11000} & \tilde{x}_2^{11000} & \cdots & 0 & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \tag{16}$$

and $\vec{y} = [M_{11}, M_{12}, \ldots, M_{21}, M_{22}, \ldots, M_{55}, \mu_1, \ldots, \mu_5]^T$. 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} (so that we are always solving for M_{ij}/μ_i). We solve this system of equations for $\vec{y}$ with linear least-squares. The solution to this least-squares problem is the interaction matrix M, which we plot as a food web in Figure S15D,E.

We tested our inference of the interaction matrix by considering how close the steady states of M were to the experimentally measured medians. At steady state, each combination C corresponds to a set of $|C|$ linear equations, as in Eq. (14), which we write as

$$M_C \vec{x}_C = -\vec{I}_C, \tag{17}$$

where M_C a subset of M pertaining to the bacteria in C, and $\vec{I}_C$ is a vector of ones of length $|C|$. For example, for a combination $C = \{1, 3\}$, we have

$$\begin{aligned} M_{11}x_1 + M_{13}x_3 &= -1 \\ M_{31}x_1 + M_{33}x_3 &= -1. \end{aligned} \tag{18}$$

Therefore, we can solve for the steady state of each combination $\vec{x}_c$ as

$$\vec{x}_C = -M_C^{-1}\vec{I}_C. \tag{19}$$

We compare this predicted steady state $\vec{x}_C$ with the experimentally measured steady state $\vec{x}_C^{\,\text{exp}}$ over all combinations by considering the error

$$\varepsilon := \frac{|\Delta \text{ SS CFUs }|}{|\text{ SS CFUs }|} = \frac{\sqrt{\sum_{C\in\{0,1\}^5}(\vec{x}_C - \vec{x}_C^{\,\text{exp}})^2}}{\sqrt{\sum_{C\in\{0,1\}^5}(\vec{x}_C^{\,\text{exp}})^2}}. \tag{20}$$

The interaction matrix M we fit with the least-squares method has an error ε of 0.322.

We construct a distribution for ε by permuting the entries of M many times, and for each permutation calculating the error ε. From this, we find that the error from the unpermuted interaction matrix M ($\varepsilon = 0.322$) is generally smaller than the permuted matrix errors (median = 5.69, standard deviation = 3.47). Therefore, our least-squares fitting method constructs an interaction matrix that reflects the experimental median CFU counts better than permuted alternatives ($p = 0.01$, comparison with errors of 10000 randomly permuted interaction matrices).

10.3 Prediction of high-diversity bacterial abundances from low-diversity data

Next, we use the gLV model parameterized on low-diversity (1 and 2-species) combinations of bacteria to predict the CFU abundances of high-diversity (3, 4, and 5-species) experiments. We compare these predictions to a simple mean-field model parameterized on the low-diversity experimental data. This mean-field model is ignorant of any microbial interactions, and for a given high-diversity bacterial combination, predicts that the bacteria have abundances equal to their "mean abundance" in the low-diversity experimental data, where a bacterias mean abundance is its average CFU count over all of the low-diversity trials in which it was present.

Then, to attain the gLV model CFU predictions, we simulate the gLV equations for each of the 16 high-diversity combinations. Since the eigenvalues of the low-diversity interaction matrix M are all negative,

each of these simulations will attain a unique steady state irrespective of initial condition. For convenience, we start each simulation with an initial condition equal to the mean-field models prediction. After this process, we have predictions for the CFU counts for each high-diversity experiment for both the gLV and the mean-field models. After comparing these predictions to the median of the true experimental CFU counts, we find that the mean-squared error of the mean-field model was more accurate for 9 combinations, and the gLV model was more accurate for 7 combinations. Applying the binomial test yields that the gLV model parameterized on low-diversity data is not significantly predictive of high-diversity bacterial abundances (p=0.803, n=16).

10.4 Pairwise correlations (Figure 6A)

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 set of experiments of diversity N that contain 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.

10.5 Determining statistical trends

We first examined whether the interaction values determined with Paine's method became more negative at higher diversities. The quantity $M_{ij}^{\text{high}} - M_{ij}^{\text{low}}$ (where each M_{ij} is the is the mean of the bootstrapped distribution, as described above) is negative for 18 out of 20 entries. We compare this to the null hypothesis that the interaction values are the unchanged, which would predict 10 out of 20 to be negative. Therefore, we found that interactions became more negative at higher diversities (binomial test, p = 2e-4).

Next, we studied how pairwise correlations change at increasing diversity. 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. Through the Kendall rank correlation coefficient, the ranking for each matrix element is compared to the strictly increasing vector [1, 2, 3, 4], 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). 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-value for the correlation matrices is 0.0323, indicating that there is a significant trend in the values of the correlation matrices to decrease. These results are robust to how we threshold CFU counts below the detection limit— if we assume that undetected species have an abundance of 1000 CFUs, corresponding to the limit of detection, the resulting one-sided p-value remains 0.0323.

11 Supplements

11.1 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.

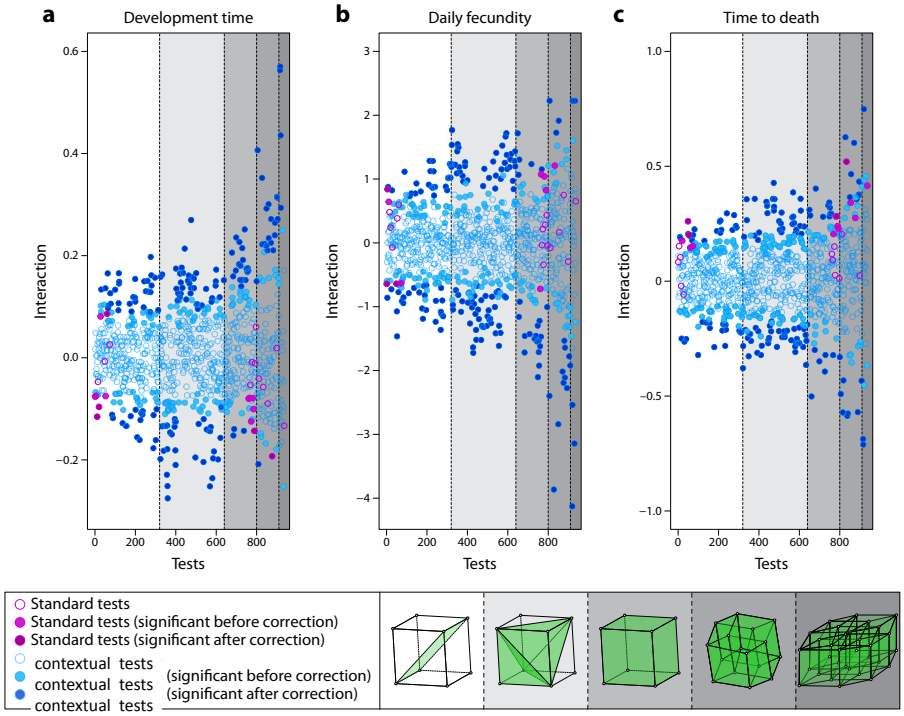


Figure 2: **Scatter plot of all tested interactions on** $\log_2$ **transformed data** (standard in purple and contextual interactions in blue). Interactions for (a) development time, (b) daily fecundity, (c) time to death. Filled circles indicate significant interactions, open circles represent non-significant 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.

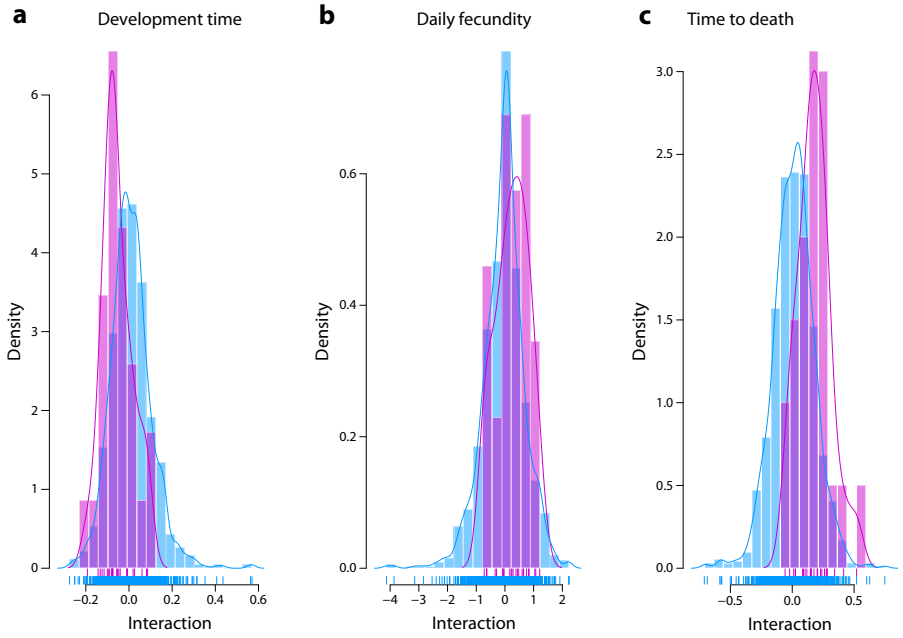


Figure 3: **Density plot of all tested interactions (standard in purple and contextual in blue) for the $\log_2$ transformed data.** Interactions for (a) development time, (b) daily fecundity, (c) time to death.

11.2 Spurious epistasis in the microbiome

The invariance of significance under transformation effectively deals with the problem of spurious epistasis, which was the subject of previous authors' work regarding genetic interactions (see [11]). In particular, Sailer and Harms estimated the nonlinear scales of arbitrary genotype-phenotype maps and then linearize these maps in order to remove the effects. Because the interactions we detect are invariant under transformation, the removal of non-linearity (itself a transformation) cannot affect the outcome of our calculations. Furthermore, the spurious epistasis dealt with in [11] is so because it does not represent true genetic interactions. However, none of the interactions that we seek to detect are genetic interactions – they are microbiome interactions, which are the products of whole organism physiologies. We apply the present framework because the simplest model is that these interactions should be additive (and many of them are). Any non-additivity becomes a potentially interesting interaction, but the mechanisms of these interactions are not due to simple inactivation of genes, and the rules by which such interactions occur are not understood. Thus, it is not appropriate to address so-called 'spurious interactions' in the current work.

11.3 Simulated data

In Figures 4 and 5 we represent the results of the standard and contextual interactions analyzed in this paper for data sampled from the standard normal distribution (on the left) and sampled uniformly at random from $[0, 1]$ (on the right). Figures 4 and 5 can be used to determine difference between two types of random values between 0 and 1 associated to the 32 bacterial combinations and the results obtained in Figure 4 (main text) on the actual experimentally measured data. Here we do not include a study of significant interactions, since these sampled data come without a sampled standard error value, being that the data themselves are randomly generated.

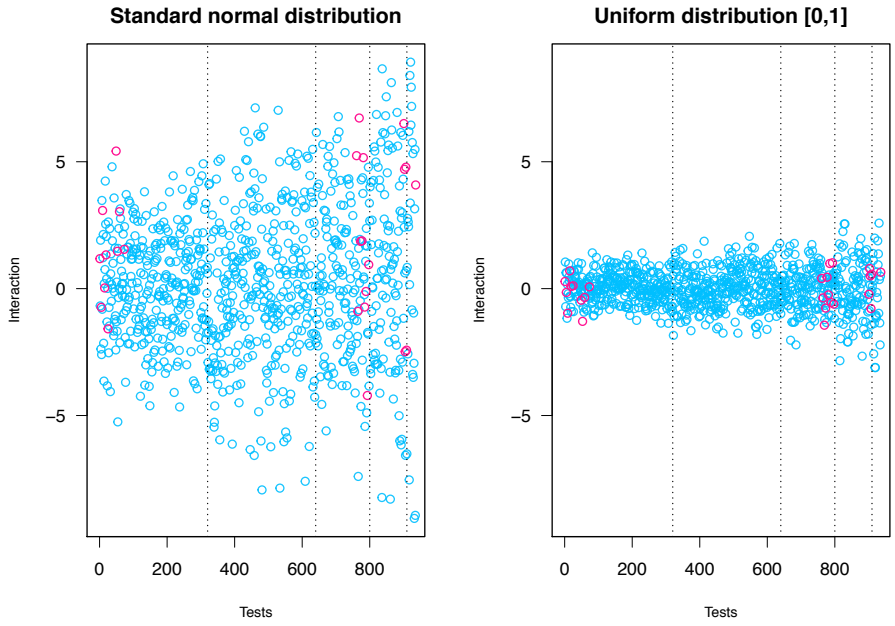


Figure 4: **Scatter plot of all tested interactions on data sampled from the standard normal distribution and from the uniform distribution on** $[0, 1]$ (standard in purple and contextual interactions in blue).

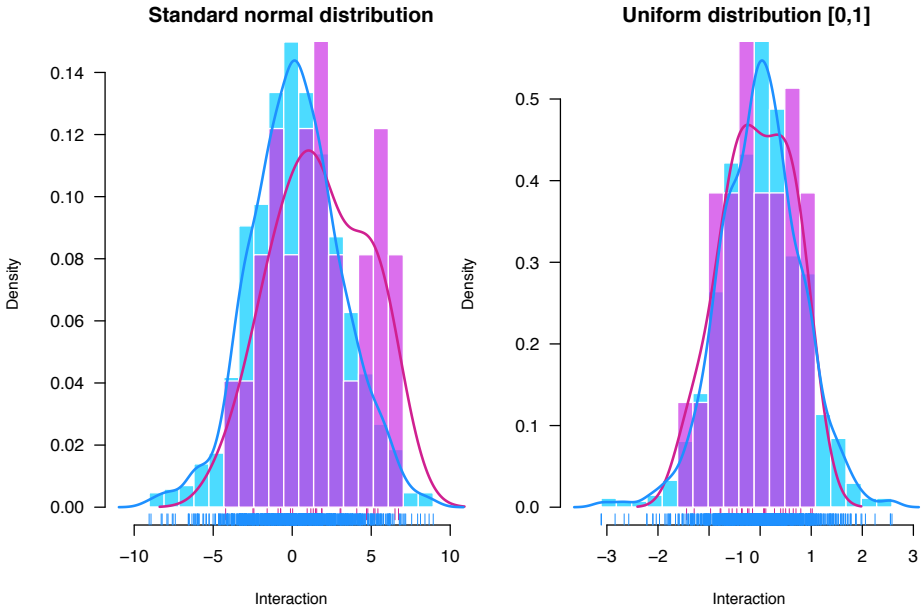


Figure 5: **Density plot of all tested interactions on data sampled from the standard normal distribution and from the uniform distribution on** $[0, 1]$**.**

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] B. Obadia, ZT. Guvener, V. Zhang, JA. Ceja-Navarro, EL. Brodie, WW. Ja, WB. Ludington, *Probabilistic Invasion Underlies Natural Gut Microbiome Stability*, Current Biology, 27, 1999-2014 (2017)

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

[5] J. A. De Loera, J. Rambau, F. Santos *Triangulations*, Algorithms and Computation in Mathematics, 25, Structures for algorithms and applications, Springer-Verlag, Berlin, xiv+535 (2010)

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

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

[8] N. Korasidis, L. Lamberti, code available at https://github.com/cbg-ethz/epistasis-formulas

[9] 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)

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

[11] R. Sailer, M. Harms, *Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps*, Genetics, (2017)

[12] Weinreich, D. M., Lan, Y., Wylie, C. S., and Heckendorn, R. B. *Should evolutionary geneticists worry about higher-order epistasis?*, Current Opinion in Genetics and Development, 23(6), 700707 (2013)

[13] Bascompte, J., Jordano, P., and Olesen, J. M., *Asymmetric Coevolutionary Networks Facilitate Biodiversity Maintenance*, Science 312.5772 (2006)

[14] Efron, B., Tibshirani, R., *Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy*, Statistical Science, vol. 1 no. 1 54-75 (1986)