nature microbiology

Resource

https://doi.org/10.1038/s41564-025-01963-3

Gut metagenomes reveal interactions between dietary restriction, ageing and the microbiome in genetically diverse mice

Received: 4 July 2024

Accepted: 19 February 2025

Published online: 31 March 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

The gut microbiome changes with age and has been proposed to mediate the benefit of lifespan-extending interventions such as dietary restriction. However, the causes and consequences of microbiome ageing and the potential of such interventions remain unclear. Here we analysed 2,997 metagenomes collected longitudinally from 913 deeply phenotyped, genetically diverse mice to investigate interactions between the microbiome, ageing, dietary restriction (caloric restriction and fasting), host genetics and a range of health parameters. Among the numerous age-associated microbiome changes that we find in this cohort, increased microbiome uniqueness is the most consistent parameter across a second longitudinal mouse experiment that we performed on inbred mice and a compendium of 4,101 human metagenomes. Furthermore, cohousing experiments show that age-associated microbiome changes may be caused by an accumulation of stochastic environmental exposures (neutral theory) rather than by the influence of an ageing host (selection theory). Unexpectedly, the majority of taxonomic and functional microbiome features show small but significant heritability, and the amount of variation explained by host genetics is similar to ageing and dietary restriction. We also find that more intense dietary interventions lead to larger microbiome changes and that dietary restriction does not rejuvenate the microbiome. Lastly, we find that the microbiome is associated with multiple health parameters, including body composition, immune components and frailty, but not lifespan. Overall, this study sheds light on the factors influencing microbiome ageing and aspects of host physiology modulated by the microbiome.

Dietary restriction (DR) improves health and extends the lifespan in diverse organisms[1,2]. However, its efficacy is variable and influenced by numerous factors, including the type of DR and the organism's genetic background[3–7]. To determine whether different types of DR extend the lifespan in a genetically heterogeneous population, we randomized 960 diversity outbred (DO) mice to fasting and caloric restriction (CR) regimes and tracked their health and lifespan with extensive phenotyping. The design of this Dietary Restriction in Diversity Outbred mice (DRiDO) study was described previously[8].

A major goal of the DRiDO study was to identify predictors of longevity. One candidate for such a predictor is the gastrointestinal microbiome, which has recently been suggested to modulate ageing[9–11] as well as responses to dietary restriction[12–15]. To investigate the relationship between the gut microbiome and lifespan, we performed shotgun metagenomic sequencing on longitudinally collected stool samples, resulting in 2,997 metagenomic profiles from 913 DO mice. Using this large dataset, we were able to address several fundamental questions.

e-mail: thaiss@stanford.edu

First, how does the microbiome age? Numerous studies in mice[16–18] and humans[19–22] have reported age-associated microbiome changes, but these changes are inconsistent across cohorts[23]. Two community features reported to increase with age are α-diversity[19,23–25] and uniqueness[26,27]. Whether these are universal properties of an ageing microbiome remains unknown. Furthermore, it remains unclear to what degree age-associated microbiome changes are caused by the ageing host.

Second, to what extent does host genetics influence the microbiome? The prevailing notion is that the environment, especially diet, has a much greater contribution to the gut microbiome than host genetics[28,29]. At the same time, human studies have identified significant microbiome heritability and quantitative trait loci (QTL) for microbiome features[30–34]. Moreover, a recent study in baboons found that nearly all microbiome features were heritable but that identifying this heritability required large sample sizes and the use of longitudinal data[35]. Therefore, it remains unclear whether the influence of genetics on the microbiome is perhaps larger than previously appreciated.

Third, what aspects of host ageing are modulated by the microbiome? Mice in the DRiDO study were deeply phenotyped, allowing us to ask whether the microbiome influences aspects of host physiology over the lifespan. The relationship between the microbiome and body composition is well established[36–38], but the DRiDO study was uniquely suited to discovering other host phenotypes influenced by the microbiome.

In addition to generating the metagenomic dataset in DO mice, we generated a second longitudinal microbiome dataset in genetically homogeneous mice, performed two validation experiments to investigate the mechanism by which host age influences the microbiome, integrated our dataset with thousands of human metagenomic samples and analysed hundreds of longitudinal host phenotypes collected as part of the DRiDO study. We begin by describing the generation of the DRiDO microbiome dataset.

Results

Metagenomic sequencing of DO mice

We enroled 960 female DO mice (corresponding to 6 breeding generations) into the DRiDO study[8] (Methods). Mice were enroled in 12 sequential cohorts ($n = 80$ mice per cohort, two cohorts per breeding generation). At 6 months of age ($n = 937$ mice survived to 6 months), mice were randomized to one of five dietary interventions (Fig. 1a and Methods): ad libitum food consumption (AL, control group), fasting 1 day week^{-1} (1D), fasting two consecutive days per week (2D), consuming 20% fewer calories every day (20) or consuming 40% fewer calories every day (40). Mice were extensively phenotyped over their lifespans (Supplementary Table 1). All dietary restriction interventions extended their lifespan (log-rank test, $P < 2.2 \times 10^{-16}$), but there was substantial interindividual variability (Fig. 1b). The largest extension in lifespan was achieved by the 40% CR group (36% increase in median lifespan versus AL).

To characterize the gut microbiome, we collected stool samples approximately every 6 months and performed metagenomic sequencing on extracted DNA (Fig. 1c). We sequenced 3,586 stool samples, 62 positive controls and 71 negative controls (Extended Data Fig. 1a–c). Different library preparations of the same DNA and repeat sequencing of the same libraries produced highly similar microbiome profiles (Extended Data Fig. 1d,e).

In addition, we leveraged[39] the fact that every mouse in our study was genetically unique and that each stool sample contained some host DNA (~9% of reads) to exclude samples in which the genotype of the host and stool sample did not definitively match (Extended Data Fig. 2a–d, Supplementary Note 1 and Supplementary Tables 3 and 4). Our pipeline for identifying sample mix-ups allowed us to detect and remedy errors that occurred during data generation (Extended Data Fig. 2e), including an animal swap (Extended Data Fig. 2f,g).

Samples were also discarded if they had inconsistent metadata, insufficient reads or an unusually high proportion of host reads, or if they appeared to be outliers (Extended Data Fig. 3a and Methods). Our final quality-controlled cohort consisted of 2,997 metagenomic profiles from 913 mice, with a median of three samples per mouse, five timepoints with at least 360 samples and at least 550 samples per dietary group.

We performed taxonomic classification using Kraken2 (ref. 40) and the Mouse Gastrointestinal Bacterial Catalogue[41] (MGBC) as reference, and we performed functional classification with HUMAnN3 (ref. 42). We used the taxonomic results from Kraken2 and MGBC instead of those available from HUMAnN3 (MetaPhlAn) for two reasons: (1) Kraken2 with MGBC classified more reads (Extended Data Fig. 3b) and (2) the fraction of characterized (that is, named) taxa was higher (55% versus 14% for genera, 12% versus 9% for species). The two methods showed good concordance (Extended Data Fig. 3c). We primarily present Kraken2 taxonomic results, except when verifying key findings and in cross-dataset comparisons. For functional results, we present MetaCyc[43] pathways and further distinguish between 'community-wide' and 'specialized' pathways, with specialized pathways defined as being highly correlated with just a few genera (Extended Data Fig. 3d).

The microbiome changes with age

First, we assessed the effect of ageing on the microbiome. Two-dimensional ordination plots of both genera (Fig. 1d) and pathways (Extended Data Fig. 3e) suggested that host age influences overall microbiome composition and function (permutational multivariate analysis of variance (PERMANOVA) pseudo-$F = 67.0$, $P \leq 0.001$, df = 1 for genera; pseudo-$F = 44.7$, $P \leq 0.001$, df = 1 for pathways). To assess the association between age and individual microbiome features, we fit a linear mixed model separately for each feature (Methods). This model accounted for age, dietary group, host genetics (via a kinship matrix), mouse identity and technical factors including date of stool collection, cohort, cage and DNA extraction batch. Notably, as different cohorts of DO mice were enroled into the study sequentially (Extended Data Fig. 4a), we were able to disentangle the effects of age and sampling timepoint (Extended Data Fig. 4b,c and Supplementary Note 2).

The majority of genera (61 of 100 tested) were associated (conditional Wald test, adjusted $P < 0.01$) with host age (Fig. 2a and Supplementary Table 5). Genera that increased with age include *Bifidobacterium* (Fig. 2b), *Turicibacter*, *Ligilactobacillus*, *Enterococcus_D* and *Escherichia*, while genera that decreased with age include poorly characterized microorganisms such as ASF356 (family Anaerotignaceae), CAG-475 (phylum Bacillota) and CAG-273 (family Clostridiaceae).

The only community feature (of 7 tested) associated with host age was uniqueness—defined as a microbiome sample's β-diversity (or distance) to its nearest neighbour[27] (Fig. 2c; adjusted $P = 5.4 \times 10^{-6}$). We confirmed that this trend was not due to the number of mice per cage decreasing with age (Extended Data Fig. 5a). We found that α-diversity (as measured by the Shannon and Simpson indexes) appeared to increase with age, but the trend was not significant (Extended Data Fig. 5b).

Repeating our analysis with species-level data produced a similar fraction of age-associated features (Extended Data Fig. 5c and Supplementary Table 6; 61% of genus-level features versus 63% of species-level features) and reproduced the association of uniqueness with age (Extended Data Fig. 5d). In addition, repeating our analysis with MetaPhlAn taxonomic results produced similar age coefficients (Extended Data Fig. 5e and Supplementary Table 7; $\rho = 0.70$, $P = 1.97 \times 10^{-4}$).

Many microbial pathways (139 of 272 tested) were affected by host age (Extended Data Fig. 5f and Supplementary Table 8), including glycolysis (PWY-1042), a community-wide pathway that decreased with age (Extended Data Fig. 5g,h) and L-lysine biosynthesis

https://doi.org/10.1038/s41564-025-01963-3

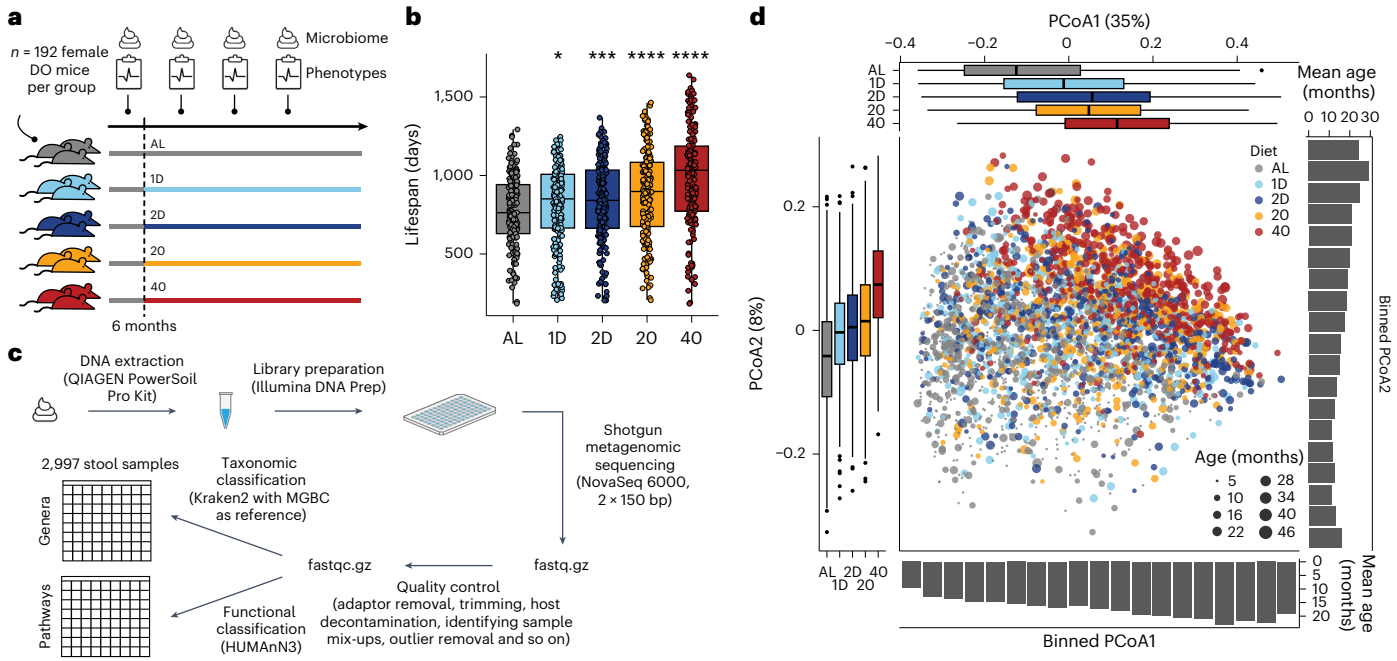


Fig. 1 | Overview of DRiDO study and microbiome dataset. a, At 6 months of age, genetically diverse mice started one of five dietary interventions. They were extensively phenotyped and stool was collected for microbiome profiling. **b**, Lifespan per dietary group. This analysis includes $n = 924$ mice: $n = 937$ mice were alive at the start of dietary restriction at 6 months, and $n = 13$ mice were omitted from analysis owing to accidental death during technician handling. *P* values were calculated with pairwise log-rank tests against the AL group and adjusted with the Benjamini–Hochberg procedure. *P* value symbols are defined as follows: $^{*}P < 0.05$; $^{***}P < 0.001$; $^{****}P < 0.0001$. **c**, Microbiome data generation consisted of extracting DNA from stool samples, preparing libraries, performing shotgun metagenomic sequencing, performing quality control, and finally, taxonomic and functional classification. After all quality control procedures, the cohort consisted of 2,997 stool samples. **d**, PCoA plot of $n = 2,997$ quality-controlled metagenomes. Ordination based on Bray–Curtis distances of genus-level relative abundances. The colours denote the dietary groups, and the sizes of the circles denote the mouse ages at the time of stool collection. Box plots along the sides show PCoA1 (top) and PCoA2 (left) coordinates per dietary group. Bar plots along the sides show the mean age of stool samples within each bin of PCoA1 (bottom) and PCoA2 (right) coordinates. PCoA1 and PCoA2 explain 35% and 8% of overall variance, respectively. For box plots in **b** and **d**, boxes extend from the 25th to 75th percentiles, whiskers extend to 1.5 times the interquartile range and the centre line is the median.

(DAPLYSINESYN-PWY), a specialized pathway that increased with age and was strongly correlated with *Bifidobacterium* (Extended Data Fig. 5i,j). Uniqueness based on functional profiles also increased with age (Extended Data Fig. 5k).

As an alternative way to assess the impact of host age on the microbiome, we asked whether microbiome information could be used to predict host age[17,22,44–46] (Methods). We found that a random forest regressor could predict host age—considering just AL mice (Fig. 2d; mean absolute error (MAE) = 16.3 weeks), all mice (Extended Data Fig. 5l; MAE = 17.4 weeks), species-level data (Fig. 5m; MAE = 13.8 weeks) or functional data (Extended Data Fig. 5n; MAE = 21.9 weeks). The features most important for age prediction—such as the genus *Ligilactobacillus* and the glycolysis pathway (PWY-1042)—overlap with the features most strongly associated with age (Fig. 2e and Extended Data Fig. 5o). However, the MAE of age prediction was relatively high, indicating that additional factors influence microbiome composition and function.

Universality of age-associated microbiome changes

Next, we asked whether any of the age-associated changes that we identified constituted a conserved microbiome signature of ageing. To this end, we compared our dataset with two other ageing microbiome datasets: a longitudinal mouse microbiome study that we conducted in male C57BL/6 mice and a publicly available human metagenomic sequencing database[47] (Fig. 3a). To enable more direct comparisons, we processed the mouse datasets with HUMAnN (because the human samples had been processed with HUMAnN), and we considered only AL mice from the DO dataset to avoid the confounding variable of dietary restriction. We fit linear mixed models separately for each dataset to identify age-associated taxonomic and functional features (Supplementary Tables 9–14 and Methods).

Within each dataset, many genera and pathways were associated with age (Fig. 3b and Extended Data Fig. 6a; 60%, 61% and 57% of genera tested in DO AL mice, C57BL/6 mice and humans, respectively; 52%, 70% and 62% of pathways), but among taxonomic features that could be compared across datasets, there was little concordance in age-associated changes (Fig. 3c; $\rho = 0.02$, −0.18 and 0.16; $P = 0.76$, 0.42 and 0.45 for pairwise comparisons). Just one taxonomic feature increased with age in all three datasets: uniqueness (Fig. 3d). We verified that uniqueness tended to increase with age in most of the studies comprising our human cohort (Extended Data Fig. 6b). *Blautia* was negatively associated with age across all datasets, but upon closer inspection, the association in humans was inconsistent (Extended Data Fig. 6c). Some genera—such as *Paramuribaculum*, *Muribaculum* and *Adlercreutzia*—increased with age in both mouse cohorts but not in humans. No α-diversity metrics were associated with age in all three datasets, and in humans, the relationship between α-diversity and age was highly variable by study (Extended Data Fig. 6d). Taken together, the most consistent taxonomic signature of ageing that we could identify was an increase in uniqueness.

As with genera, there was poor overall concordance of age-associated pathway changes across datasets (Extended Data Fig. 6e; $\rho = -0.13$, −0.01 and 0.03; $P = 0.06$, 0.88 and 0.68 for pairwise comparisons), but several individual pathways showed consistent changes: inosine-5′-phosphate (IMP) biosynthesis (PWY-7234) increased with age, while thiamine diphosphate formation (PWY-7357), thiamine

https://doi.org/10.1038/s41564-025-01963-3

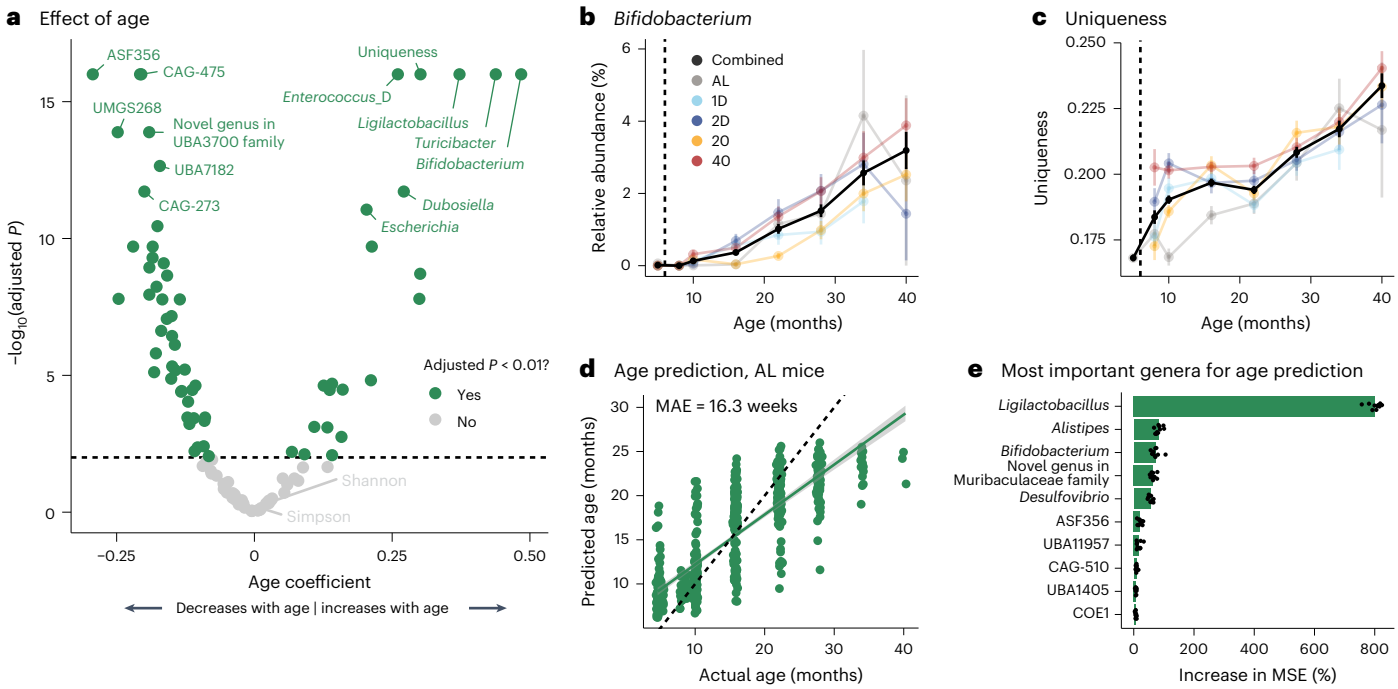


Fig. 2 | Host age influences the microbiome. a, Effect of age on genera. Age coefficients and standard errors were calculated with a linear mixed model (model 1). *P* values were calculated with a conditional Wald test and adjusted with the Benjamini–Hochberg procedure. **b**, *Bifidobacterium* increases with age ($n = 2{,}988$ metagenomes with age ≤40 months). The vertical dashed line at 6 months represents the start of dietary restriction. **c**, Uniqueness increases with age ($n = 2{,}988$ metagenomes with age ≤40 months). **d**, Host age prediction using 573 genus-level metagenomes from AL mice. The green line represents the line of best fit, and the grey shading represents the 95% confidence interval (linear regression). The black dashed line at $y = x$ represents perfect accuracy. **e**, Top 10 most important genera for age prediction. Each dot is one of 10 cross-validation folds. The *x* axis shows the percentage increase in MSE when that particular genus is excluded from a tree within the random forest regressor. In **b** and **c**, data are presented as mean ± standard error of the mean (s.e.m.).

diphosphate salvage (PWY-6897), inositol degradation (PWY-7237) and flavin biosynthesis (RIBOSYN2-PWY; Extended Data Fig. 6f) decreased with age. With the exception of IMP biosynthesis, these are all community-wide pathways (Extended Data Fig. 6g). These results suggest that a functional signature of microbiome ageing may be characterized by decreased production of cofactors such as thiamine (vitamin B_1) and riboflavin (vitamin B_2) and increased production of IMP, a precursor for de novo purine biosynthesis.

Microbiome ageing may be explained by neutral theory

What causes the microbiome to change with age? We investigated this question through the lens of selection theory and neutral theory[48–50]. We hypothesized that the apparent 'age' of the microbiome is dictated by selective pressure related to host age (selection theory) or exposure to stochastic events that accumulate with time (neutral theory). To test these hypotheses, we needed to disentangle host age from microbiome age. We achieved this by cohousing young (8 weeks) and old (19–20 months) C57BL/6 mice for 1 month and monitoring their microbiomes (with 16S sequencing) after 2, 4, 6 and 8 weeks of separation (Fig. 3e). Cohousing caused the microbiomes of young mice to resemble those of old mice (Fig. 3f). In other words, it had decoupled host age and microbiome age in young mice.

After separation, the microbiomes of young mice never reverted to a young state: they remained different from the microbiomes of non-cohoused young mice (Fig. 3g). Furthermore, a classifier trained on baseline microbiome samples always incorrectly identified both cohoused and previously cohoused young mice as old (Fig. 3h). These results indicate that a young host does not dictate the apparent age of the microbiome. The effect of cohousing was much less pronounced in old mice (Fig. 3f,g), and the classifier almost always correctly identified old mice, even during cohousing (Fig. 3h).

We observed again that uniqueness was higher in old compared with young mice (Fig. 3i; *t*-test, $t = -7.8$, df = 138.9, $P = 1.7 \times 10^{-12}$). Notably, young mice cohoused with old mice acquired higher microbiome uniqueness compared with non-cohoused young mice (*t*-test, $t = -5.0$, df = 89.9, $P = 2.7 \times 10^{-6}$), and this was not reversed upon separation from old mice (*t*-test, $t = -5.6$, df = 99.0, $P = 2.0 \times 10^{-7}$), showing that uniqueness is not strictly determined by host age.

We next investigated why the old microbiome appeared to be dominantly transferred upon cohousing. We reasoned that this was due to either microbiome-intrinsic properties or differences in host properties, for example, if old mice were less coprophagic or if their intestinal colonization niches were less permissive of new microorganisms. We controlled for host differences by repeating the cohousing experiment with young (3–5 months) germ-free mice that received faecal microbiome transplants (FMTs) from young (3 months) or old (17 months) mice (Extended Data Fig. 6h). Two weeks after the FMT, young and old recipients had distinct microbiomes (Extended Data Fig. 6i). Notably, the old FMT recipients were again affected by cohousing less than the young recipients (Extended Data Fig. 6i), indicating that resistance to cohousing is caused by the old microbiome, rather than by host differences between young and old mice.

Taken together, our observation that a young host does not transform an old microbiome into a young one argues against selection theory, which would predict that the age of the host dictates the apparent age of the microbiome. An alternative explanation for why the microbiome changes with age is thus that it accumulates stochastic events over time (neutral theory), consistent with increasing uniqueness.

Host genetics shape the gut microbiome

Next, we examined how other experimental variables affect the microbiome. The DRiDO study's use of genetically diverse and genotyped mice

https://doi.org/10.1038/s41564-025-01963-3

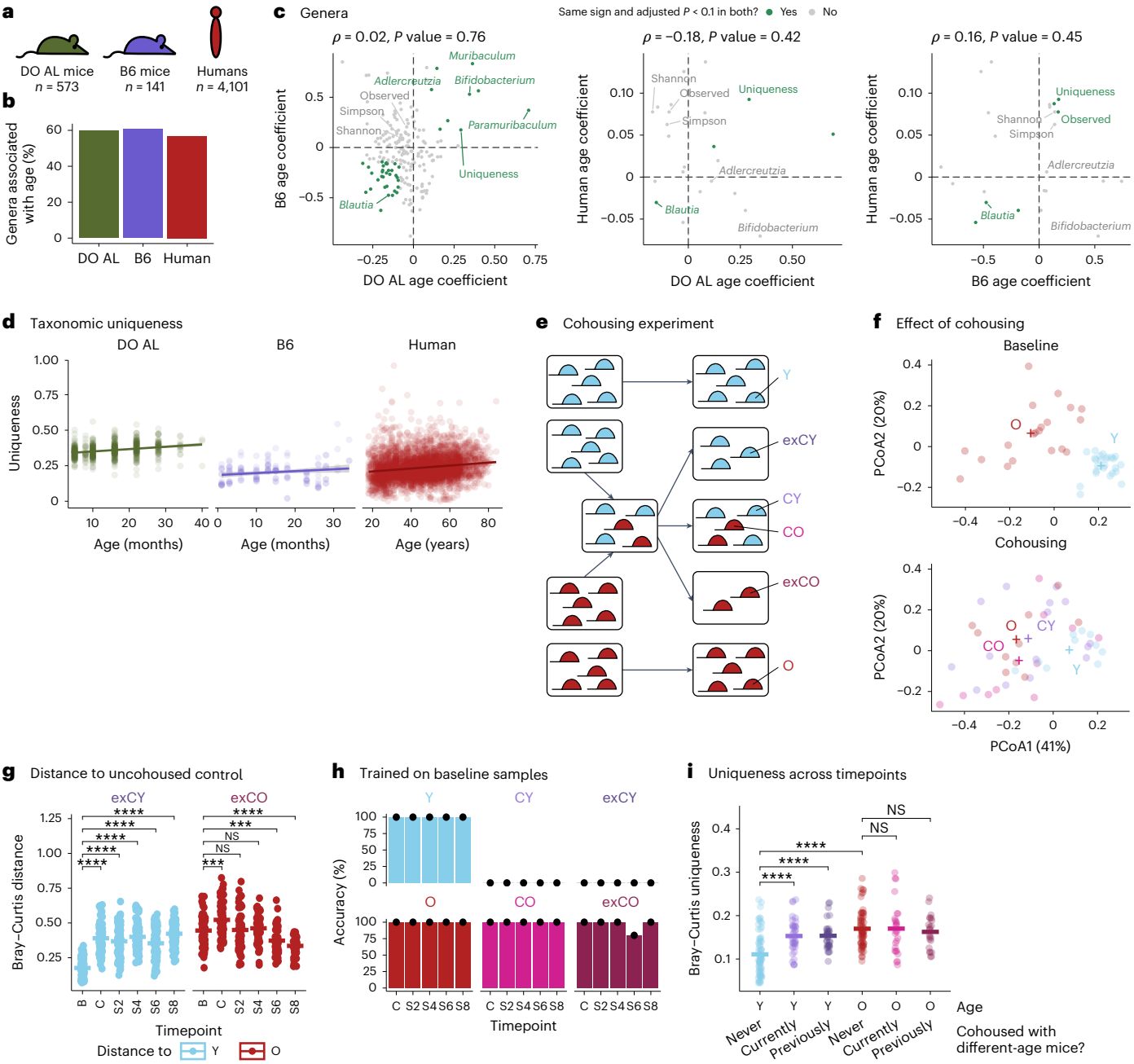


Fig. 3 | Universality of age-associated microbiome changes. a, We compared 573 samples from DO AL mice with 141 samples from a different mouse ageing cohort ('B6') and with 4,101 human gut microbiome samples. **b**, Percentage of genera associated with age (Benjamini–Hochberg-adjusted $P < 0.1$) based on linear mixed models (models 5–7) within each dataset. **c**, Comparison of age-associated taxonomic changes across datasets. Each pairwise comparison shows all features that passed prevalence filtration in both datasets. Spearman correlation and corresponding P value are shown above each plot. Features associated with age and with the same sign in the pairwise comparison are shown in green. **d**, Taxonomic uniqueness increases with age in all three datasets. Each panel includes the line of best fit and 95% confidence interval. **e**, Schematic of the cohousing and separation experiments. Y, young always housed with young; O, old always housed with old; CY, young housed with old; CO, old housed with young; exCY, formerly CY that were separated from old; exCO, formerly CO that were separated from young. **f**, PCoA (genus-level Bray–Curtis distances) of samples at baseline and 1 month of cohousing. The plus sign denotes the group centroid. **g**, Bray–Curtis distances ($n = 863$) between previously cohoused mice (exCY, exCO) and non-cohoused controls (Y, O). **h**, Random forest classifier trained on baseline samples and evaluated on cohousing and separation samples. Accuracy is the percentage of samples within each group correctly classified as young or old. **i**, Uniqueness split by age and cohousing status ($n = 264$ samples). B, baseline; C, cohousing; S2, 2 weeks of separation; S4, 4 weeks of separation; and so on. In **g** and **i**, the significance of group differences was evaluated with a t-test, and P value symbols are defined as follows: NS, $P \geq 0.05$; ***$P < 0.001$; ****$P < 0.0001$.

enabled us to evaluate the contribution of host genetics to microbiome composition and function. The first indication that host genetics might have a substantial influence on the microbiome was the observation that (genetically diverse) DO mice had higher uniqueness than (genetically homogeneous) C57BL/6 mice (Fig. 3d; t-test, $t = 23.2$, df = 174.2, $P < 2.2 \times 10^{-16}$), suggesting that genetic diversity led to more interindividual microbiome variation.

We found that the majority of genera (65%, mean heritability of heritable features = 0.11) and pathways (52%, mean = 0.08) had significant, though modest, heritability (Fig. 4a, Extended Data Fig. 7a,

https://doi.org/10.1038/s41564-025-01963-3

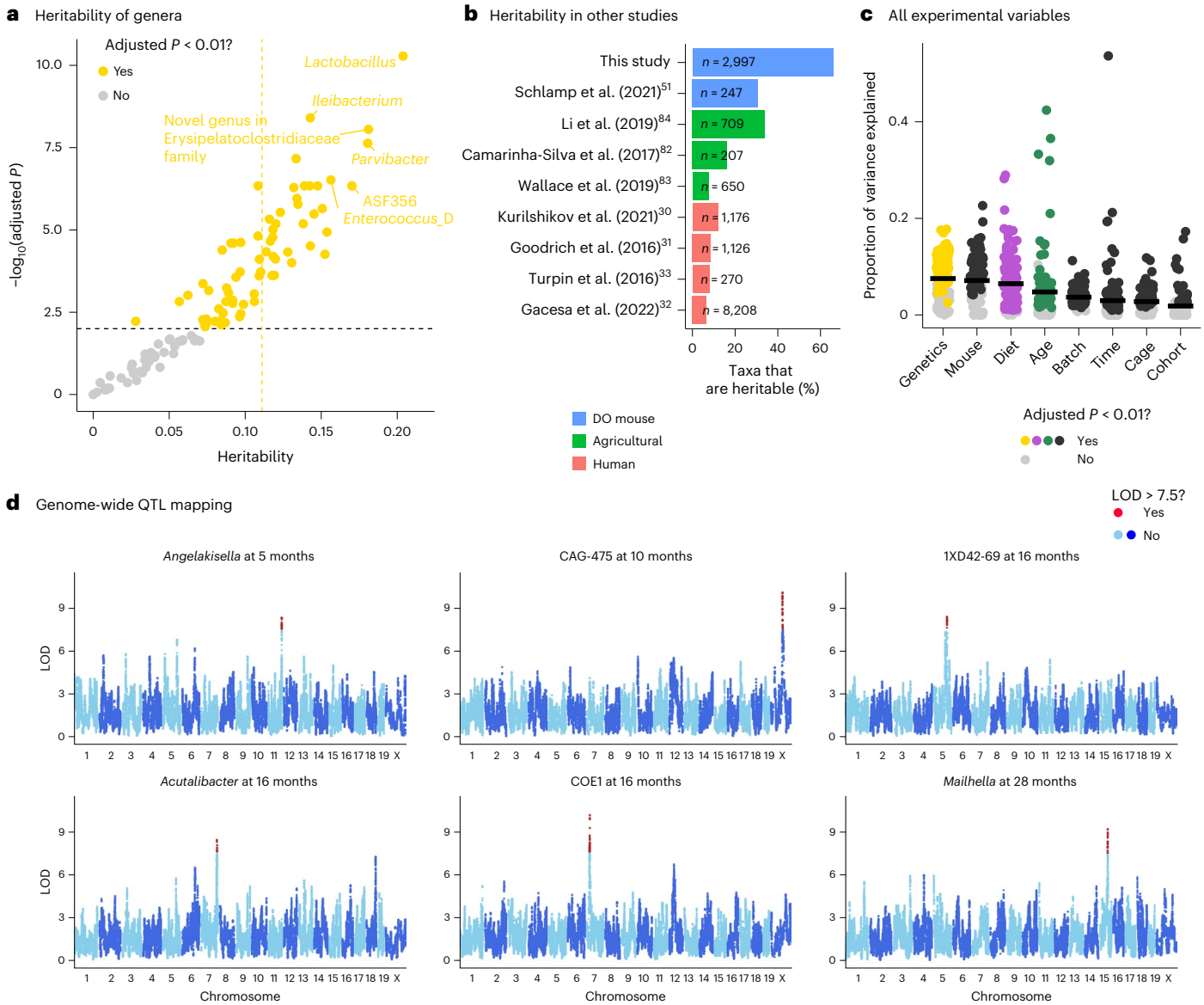


Fig. 4 | Genetic influence on the microbiome. a, Heritability of $n = 107$ genus-level features as calculated by a linear mixed model (model 1). *P* values were calculated with a likelihood ratio test and adjusted with the Benjamini–Hochberg procedure. The yellow vertical dashed line shows mean heritability for heritable features. **b**, Percentage of heritable taxa (as reported by the authors) in other studies[30,33,51,82,84]. The number of samples per study is indicated. The colour of each bar indicates whether the study was performed in DO mice, agricultural animals or humans. **c**, Proportion of variance explained (PVE) by all experimental variables for 107 genus-level features (model 10). *P* values were calculated using a likelihood ratio test and adjusted with the Benjamini–Hochberg procedure. The horizontal lines show the mean PVE. **d**, Genome-wide results for the six age-specific significant QTL with Benjamini–Hochberg-adjusted $P < 0.1$ (*P* values calculated by permutation). Markers with LOD greater than 7.5 are coloured red.

Supplementary Tables 3 and 6, and Methods). The most heritable taxa were *Lactobacillus*, *Ileibacterium* and a previously undescribed genus in the Erysipelatoclostridiaceae family. The most heritable microbial pathway was lactose and galactose degradation (LACTOSECAT-PWY), a specialized pathway strongly correlated with *Lactobacillus* (Extended Data Fig. 7b). Heritable community-wide pathways include de novo queuosine biosynthesis (PWY-6700) and fatty acid biosynthesis initiation (PWY66-429). The percentage of heritable species-level features (49%) was similar to the percentage of heritable genus-level features (64%; Extended Data Fig. 7c), and using MetaPhlAn taxonomic results produced generally concordant heritability estimates (Extended Data Fig. 7d; $\rho = 0.58$, $P = 3.28 \times 10^{-3}$).

As human studies often report a minor role of genetics in shaping the gut microbiome[28,32], we confirmed our surprising result with several additional approaches. First, we made sure that this was not a peculiarity of the software we used for calculating heritability, by exactly reproducing our heritability estimates with an alternative software package (Extended Data Fig. 7e). Second, we compared our results with an independent dataset of DO mice[51] and found that, of genera that could be compared across studies, the most heritable genus in both datasets was *Lactobacillus* (Extended Data Fig. 7f and Methods). Third, we calculated heritability separately per age (Methods) to determine whether our widespread heritability was a consequence of using longitudinal data, as previously suggested[35]. Many more features were heritable when using longitudinal data ($n = 66$ genera) compared with cross-sectional data ($n = 10, 4, 4, 10$ and 1 genera for 5-, 10-, 16-, 22- and 28-month-old samples, respectively), but downsampling to a similar number of samples ($n = 516$) as in cross-sectional data completely

erased heritability (Extended Data Fig. 7g; zero heritable genera). In other words, a large number of samples was critical for detecting heritability, not the use of longitudinal data per se.

Another potential explanation for why we observed higher heritability than typically reported in human studies relates to differences in environmental variability. Heritability is a relative measure; it is defined as $V_G/(V_G + V_E)$, where V_G is the amount of variance explained by additive genetic effects, and V_E is all other sources of variance. Therefore, a larger V_E will decrease heritability even if V_G is unchanged. If we assume that human observational studies have a higher V_E (that is, more unexplained environmental variability) than studies involving agricultural animals and an even higher V_E than studies in laboratory animals, then differences in V_E may explain why heritability estimates are generally lower in humans than in agricultural animals and lower still than in DO mice (Fig. 4b).

How does the magnitude of this genetic effect compare with the effects of other experimental variables? To answer this question, we fit a linear mixed model in which all variables were treated as random effects (Supplementary Tables 15 and 16 and Methods). We found that the proportion of variance explained by genetics (mean 7.5% for genera, 5.5% for pathways) was similar to that of dietary restriction (mean 6.5% and 3.6%) and ageing (mean 4.7% and 3.3%; Fig. 4c and Extended Data Fig. 7h). The majority of variance remained unexplained (mean 63% for genera, 76% for pathways), emphasizing that the microbiome is strongly influenced by factors that are yet unaccounted for. Technical factors such as sampling timepoint, cage, cohort of DO breeding and DNA extraction batch explained smaller, but notable, amounts of variance, emphasizing the importance of retaining these technical factors in our linear modelling.

Lastly, because many microbiome features showed heritability, we performed genome-wide QTL mapping to find loci that influence microbiome composition (Methods). We tested 107 features at five different ages and identified just six significant QTL (Fig. 4d and Supplementary Table 2). The disconnect between widespread microbiome heritability but few significant QTL probably reflects the fact that microbial abundance is a complex, polygenic trait. Furthermore, these six QTL were significant at only one age (Extended Data Fig. 7i), suggesting that the genetic influence on the microbiome may be temporally variable.

Effects of dietary restriction

In addition to host age and genetics, dietary restriction also had a strong influence on the microbiome (Fig. 1d and Fig. 4c). Indeed, we found that nearly all microbiome features (100 of 107 genus-level features tested, 244 of 273 pathway-level features tested) were affected by DR (Fig. 5a, Extended Data Fig. 8a and Supplementary Tables 5 and 8). The genus most increased by DR was UMGS1815 (order Oscillospirales; Fig. 5b), and we recovered the commonly observed phenomenon[18,52–54] of DR increasing the abundance of *Lactobacillus* and closely related genera (Fig. 5c). Microbial functions strongly increased by DR include the specialized pathway lysine biosynthesis II (PWY-2941, highly correlated with *Ligilactobacillus*; Extended Data Fig. 8b,c) and the specialized pathway O-antigen building block biosynthesis (OANTIGEN-PWY, highly correlated with *Lactobacillus*), while functions decreased by DR include the community-wide pathways of urea cycle (PWY-4984; Extended Data Fig. 8d,e) and citrulline biosynthesis.

We observed that 40% CR had the strongest overall effect on the microbiome, followed by 2D fasting and 20% CR, and finally by 1D fasting (Fig. 5d and Extended Data Fig. 8f), suggesting that the intensity of dietary restriction parallels the extent of microbiome changes. Fasting and CR had similar effects on the microbiome, as evidenced by high correlations between dietary restriction coefficients calculated by the linear mixed model (Fig. 5e and Extended Data Fig. 8g). Furthermore, few genera were affected only by CR or fasting (Extended Data Fig. 8f). Two exceptions are *Emergencia*, which was decreased only by CR (Fig. 5g), and *Roseburia*, which was decreased by fasting and increased by CR (Fig. 5h). We were unable to find pathways differentially affected by CR or fasting (Extended Data Fig. 8h). These findings suggest that CR and fasting have similar overall effects on the microbiome, and that more intense dietary interventions cause greater microbiome shifts.

As an orthogonal way to investigate how the gut microbiome is influenced by DR, we asked whether a machine-learning algorithm would be able to predict whether a mouse was on DR based on its microbiome profile. We trained a random forest classifier (separately at each age) to predict binary DR status (Methods). As expected, the classifier had no predictive accuracy at 5 months (that is, before the start of DR; Fig. 5i and Extended Data Fig. 8i; mean area under the ROC curve (AUC) = 0.47 for genera, 0.50 for pathways). After DR initiation, the classifier could predict whether a mouse was on DR (mean AUC = 0.81–0.92 for genera, 0.71–0.80 for pathways). The features most important for prediction include those strongly associated with DR, such as the genera UMGS1815 and *Ligilactobacillus* and the lysine biosynthesis II and urea cycle pathways (Extended Data Fig. 8j,k). We also asked whether a random forest classifier could perform the more difficult task of predicting which dietary intervention a mouse was assigned to. Again, accuracy was no better than chance (20%, or randomly guessing one of the five dietary groups) at 5 months and better than chance at later ages (Fig. 5j and Extended Data Fig. 8l). Furthermore, accuracy was highest for the AL, 2D and 40% dietary groups (Fig. 5k and Extended Data Fig. 8m), consistent with more intense dietary interventions creating more distinct microbiome states.

We also assessed the impact of DR using species-level data and MetaPhlAn taxonomic results. The fraction of species affected by DR (89%) was similar to that of genera (93%; Extended Data Fig. 8n). DR coefficients based on MetaPhlAn data were generally concordant (Extended Data Fig. 8o; $\rho = 0.50–0.70$), with more intense dietary interventions (2D, 40% CR) having higher correlations.

Finally, we investigated the hypothesis suggested by previous work that DR induces a more youthful-appearing microbiome[52,55,56]. To test this notion, we trained a random forest regressor on AL samples and predicted the host age of DR samples. If DR produced a more youthful microbiome state, the predicted age of DR samples would be lower than that of AL samples. Surprisingly, we found that DR samples had higher (t-test, $P < 0.01$) predicted ages than AL samples (Fig. 5l and Extended Data Fig. 9a). Consistently, a regressor trained on 40% CR samples predicted lower ages for AL samples than for the other DR samples (Extended Data Fig. 9b). Furthermore, we identified genera that were affected by age and DR in the same direction, such as UBA11957 (Fig. 5m) and *Ligilactobacillus* (Fig. 5n), and two-dimensional ordination showed that ageing and DR shifted the microbiome in the same direction (Extended Data Fig. 9c). In summary, our data do not suggest that DR 'rejuvenates' the microbiome to a more youthful state.

The microbiome associates with host health, not lifespan

Having characterized the factors that influence the microbiome, we next asked whether the microbiome modulates any of the host phenotypes measured in the DRiDO study. We tested genera and pathways for association with 197 phenotypic traits across 13 assays, while controlling for age, dietary group and mouse (Fig. 6a, Supplementary Tables 17 and 18, and Methods). The proportion of associations (adjusted $P < 0.01$) was similar for genera (0.9%) and pathways (0.7%; Extended Data Fig. 10). We observed associations with phenotypes measured in the body weight, body composition (dual-energy X-ray absorptiometry, DEXA), metabolic cage, frailty and flow cytometry assays (Fig. 6b). No associations were observed with phenotypes measured in the echocardiogram, glucose, grip strength, void (bladder function), rotarod or wheel running assays.

Focusing on phenotypes related to body mass and composition, we identified that genera positively associated with body mass included *Alistipes*, *COE1* (Lachnospiraceae family), *Lachnospira* and

https://doi.org/10.1038/s41564-025-01963-3

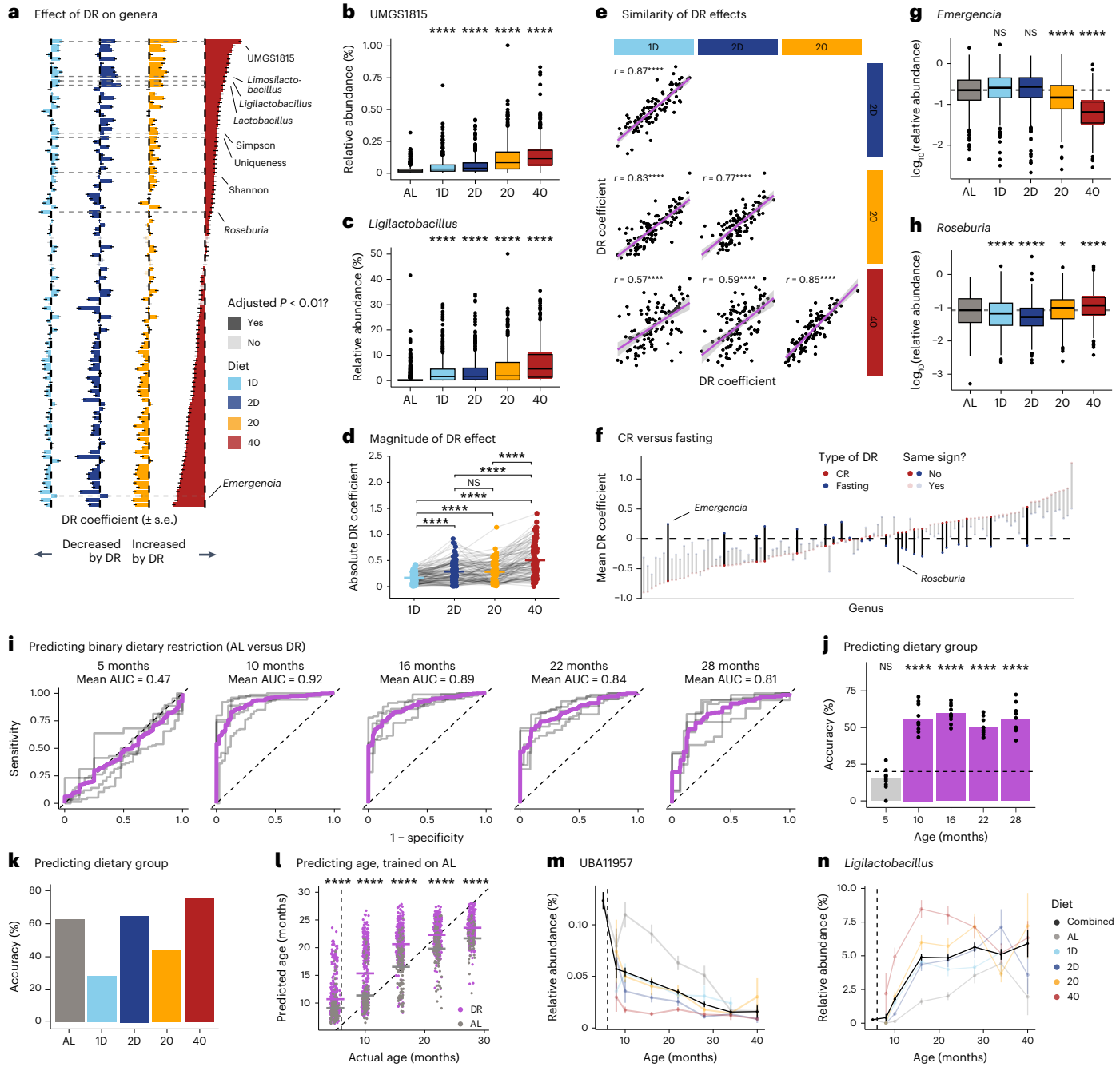


Fig. 5 | Effects of dietary restriction on the microbiome. a, Effect of DR on 107 genus-level features. DR coefficients and standard errors were calculated with a linear mixed model (model 1). *P* values were calculated with a conditional Wald test and adjusted with the Benjamini–Hochberg procedure. The horizontal dashed grey lines are visual aids to help compare dietary groups. **b**, UMGS1815 was increased by DR ($n = 2{,}988$ metagenomes). **c**, *Ligilactobacillus* was increased by DR ($n = 2{,}988$ metagenomes). **d**, Absolute magnitude of DR coefficients ($n = 107$ genus-level features). The grey lines connect the same genus in different dietary groups. The horizontal bars show the mean. Statistical significance was evaluated by a paired *t*-test. **e**, Comparison of DR coefficients. The Pearson correlation and *P* value are indicated above each scatter plot. Lines of best fit and 95% confidence intervals (linear regression) are shown in purple. **f**, Mean CR (red) versus mean fasting (blue) coefficients are connected by vertical lines. Genera with opposite signs are opaque, while genera with the same sign are transparent. Dashed horizontal line at 0. **g**, *Emergencia* is decreased only by CR ($n = 2{,}988$ metagenomes). The AL group median is designated by a horizontal dashed grey line. **h**, *Roseburia* is decreased by fasting and increased by CR ($n = 2{,}988$ metagenomes). **i**, Receiver operating characteristic (ROC) curves for binary DR prediction. Each grey line is the ROC curve for one of five cross-validation folds. The purple line is the mean ROC curve. The diagonal dashed line represents no predictive accuracy. **j**, Predicting the dietary group before (grey) and after (purple) DR initiation. Each dot represents the prediction accuracy in 1 of 10 cross-validation folds. The horizontal dashed line at 20% represents expected accuracy by chance. Statistical significance was evaluated by a one-sided *t*-test (testing whether the mean accuracy is greater than 20%). **k**, Prediction accuracy stratified by dietary group. Only predictions after the start of DR were considered. **l**, Age prediction with a random forest regressor trained on AL samples ($n = 2{,}618$ metagenomes from 5, 10, 16, 22 and 28 months). The vertical dashed line at 6 months represents DR initiation; the diagonal dashed line represents perfect prediction. Statistical significance was evaluated by a *t*-test between AL and DR predictions at each age. **m**, UBA11957 decreases with both age and DR ($n = 2{,}988$ metagenomes). **n**, *Ligilactobacillus* increases with both age and DR ($n = 2{,}988$ metagenomes). In **b**, **c**, **g** and **h**, statistical significance was evaluated by a *t*-test against the AL group. For box plots in **b**, **c**, **g** and **h**, boxes extend from the 25th to 75th percentiles, whiskers extend to 1.5 times the interquartile range and the centre line is the median. In **m** and **n**, data are presented as mean ± s.e.m. In **b**–**e**, **g**, **h**, **j** and **l**, *P* value symbols are defined as follows: NS, $P \geq 0.05$; *$P < 0.05$; ****$P < 0.0001$.

https://doi.org/10.1038/s41564-025-01963-3

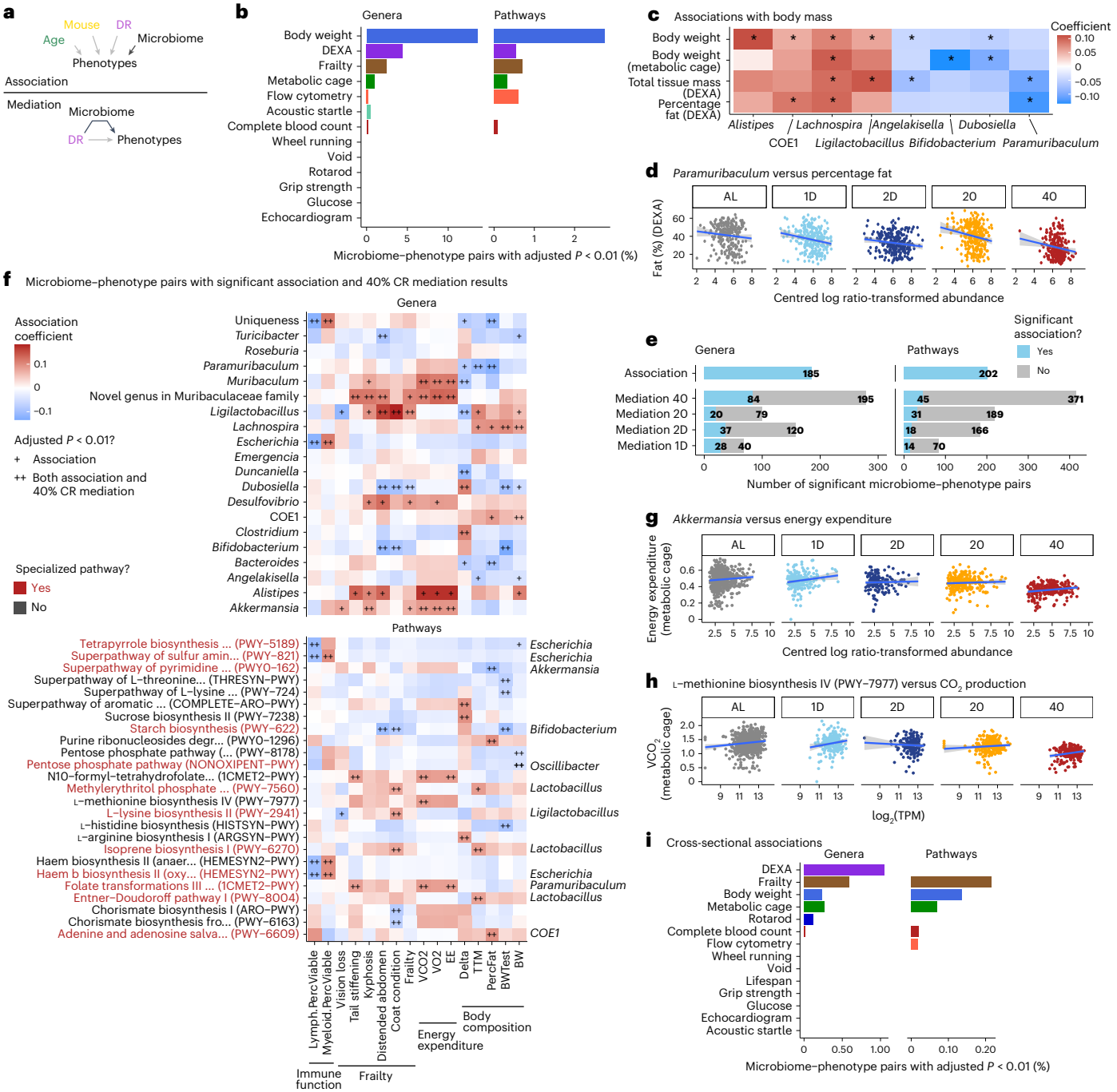


Fig. 6 | Microbiome–phenotype associations. a, Association and mediation analyses were used to identify host phenotypes influenced by the microbiome. For association analysis, a linear mixed model was fit for every microbiome feature–phenotype pair, with age, DR and mouse as covariates (model 11). For mediation analysis, we tested each microbiome feature–phenotype pair to see which effects of DR are mediated by the microbiome (models 13 and 14). **b**, Percentage of significant (Benjamini–Hochberg-adjusted $P < 0.01$) microbiome–phenotype associations per phenotypic assay. *P* values were calculated using a likelihood ratio test in which the null model omitted the microbiome term. The denominator for the percentage is the number of microbiome–phenotype pairs tested within each assay. **c**, Select associations (calculated as in **b**) between genera and body-mass-related phenotypes. Positive associations are red; negative associations are blue. *Adjusted $P < 0.01$. **d**, *Paramuriculum* is associated with percentage fat, as measured by DEXA. **e**, Overlap of microbiome–phenotype pairs with significant (adjusted $P < 0.01$) association (model 11) and mediation (models 13 and 14) results. **f**, Heatmap of select microbiome–phenotype pairs significant by just association analysis or by both association and 40% CR mediation analysis. Phenotypes are grouped by phenotypic domain. Positive associations are red; negative associations are blue. Pathways in red are specialized pathways, and their most similar genus is shown on the right side of the heatmap. Lymph.PercViable, percentage of viable cells that are lymphocytes; Myeloid.PercViable, percentage of viable cells that are myeloid; VCO2, volume of carbon dioxide produced; VO2, volume of oxygen produced; EE, energy expenditure; Delta, change in body weight; TTM, total tissue mass; PercFat, percentage fat; BWTest, body weight (metabolic cage); BW, body weight. **g**, *Akkermansia* is associated with energy expenditure. **h**, Methionine biosynthesis (PWY-7977) is associated with the volume of carbon dioxide produced. **i**, Percentage of significant (adjusted $P < 0.01$) cross-sectional microbiome–phenotype associations (model 12) per phenotypic assay. *P* values were calculated using a likelihood ratio test. The denominator for the percentage includes associations across all five ages tested. In **d**, **g** and **h**, the blue lines represent the line of best fit and 95% confidence interval (linear regression).

Ligilactobacillus, while genera negatively associated with body mass included *Angelakisella*, *Bifidobacterium*, *Dubosiella* and *Paramuribaculum* (Fig. 6c,d).

To investigate whether the microbiome was involved in the impact of DR on host phenotypes, we used mediation analysis (Fig. 6a, Supplementary Tables 19 and 20, and Methods). Several hundred microbiome–phenotype pairs were significant (mediation effect adjusted $P < 0.01$), and a subset of these overlapped with significant associations (Fig. 6e). The mediation effect was small (mean proportion mediated = 19% for both genera and pathways), consistent with the microbiome modulating, rather than driving, the effects of DR on host physiology.

We focused on our highest-confidence microbiome–phenotype pairs, those having both an association and 40% CR mediation result (Fig. 6f). For example, *Muribaculum* and *Akkermansia* (Fig. 6g) are positively associated with energy expenditure. Indeed, the association between *Akkermansia* and increased energy expenditure has been reported previously[57]. Furthermore, pathway–phenotype associations propose mechanisms for the taxonomic associations. For example, the methionine biosynthesis pathway (PWY-7977) is associated with increased carbon dioxide production suggesting that methionine may be involved in the associations with energy expenditure.

Lastly, we asked whether the microbiome is associated with lifespan. Because lifespan is not a longitudinal measurement, we performed association analysis separately per age (Supplementary Tables 21 and 22, and Methods). Consistent with the longitudinal analysis (Fig. 6b), we identified numerous associations with frailty and body composition (Fig. 6i) but no associations with lifespan, indicating that several parameters of host health are affected by the microbiome, but that lifespan may not be a microbiome-associated trait.

Discussion

We generated nearly 3,000 metagenomic profiles from more than 900 genetically diverse mice. The metagenomics data are longitudinal; have passed thorough and conservative quality control, including a pipeline to confirm that stool samples unambiguously matched their corresponding host genomes (Extended Data Fig. 2 and Supplementary Note 1); and are paired with host genomes and hundreds of longitudinally collected host phenotypes. In addition to the raw sequencing files, we have made available summarized data tables, code and an example analysis tutorial (https://github.com/levlitichev/DRiDO_microbiome) to facilitate use of this dataset by the community. Our work generated four major insights.

First, we find that the most consistent age-associated microbiome change was an increase in uniqueness, that is, accumulated interindividual distance in taxonomic space, rather than changes in specific taxa. Together with our finding that young mice do not exert control over the 'age' of the microbiome (arguing against selection theory), we speculate that age-associated microbiome changes are caused by an accumulation of stochastic changes (in support of neutral theory). This model of microbiome ageing bears resemblance to the recent suggestion that increased stochasticity may be the signal used by methylation clocks to successfully predict an animal's age[58].

Second, we find that host genetics explain similar amounts of microbiome variation as ageing and dietary restriction, and that most microbiome features show significant but modest heritability. We demonstrate that detecting this genetic influence requires not only large sample sizes but also a controlled laboratory environment, such that the effect of host genetics is not obscured by unexplained environmental variability. Notably, even in a controlled study such as DRiDO, the overall effect of unexplained environmental variability is larger than the effect of all other measured variables combined (Fig. 4c).

Third, we make several noteworthy observations about how DR affects the microbiome: namely, that fasting and CR do not induce notably different microbiome changes, more intense dietary interventions lead to more distinct microbiome changes, and dietary restriction does not appear to produce a more youthful-appearing microbiome.

Fourth, the microbiome is associated with many aspects of host health—in particular, body composition, immune function and frailty—but it is not associated with longevity. As DR extends the lifespan without rejuvenating the microbiome, the microbiome appears to contribute to the impact of DR on health but not on longevity.

In summary, this work advances our understanding of the factors shaping the microbiome over the lifespan and elucidates which aspects of host physiology and ageing are influenced by the microbiome.

Methods

Mouse experiments were approved by the Institutional Animal Care and Use Committees at The Jackson Laboratory (protocol number 06005) and the Perelman School of Medicine at the University of Pennsylvania (protocol number 806361).

DRiDO study

The DRiDO study was previously described in depth[8]. Briefly, we enroled 960 female DO mice (corresponding to breeding generations 22–24 and 26–28) into the study from March 2016 until November 2017. All DO mice are genetically unique, and the parents of a given generation were from the previous generation[59]. Just one female mouse was used from each litter, so no mice in the study were siblings. Mice were enroled in 12 sequential cohorts ($n = 80$ mice per cohort, two cohorts per generation). Only female mice were used to prevent aggressive competition for limited food resources[60]. The sample size of 960 was chosen to detect a 10% change in lifespan based on historical lifespan data from DO mice.

At 6 months of age, the surviving $n = 937$ mice began one of five dietary interventions: ad libitum, 20% CR, 40% CR, 1 day week^{-1} fasting or two consecutive days per week fasting. All mice were fed the same standard mouse chow diet (5K0G, LabDiet), but the timing and amount varied by dietary intervention. Dietary restriction was initiated at 6 months to evaluate the consequences of adult-onset, rather than lifelong, DR. Mice were housed up to eight animals per cage, with all cage-mates assigned to the same dietary intervention. The 20% CR mice received 2.75 g of food per day, the 40% CR mice received 2.06 g d^{-1}, the 1D mice were fasted from Wednesday 3 p.m. until Thursday 3 p.m. and the 2D mice were fasted from Wednesday 3 p.m. until Friday 3 p.m. The CR mice received a triple feeding on Friday afternoon. They consumed this food quickly, causing the 20% CR and 40% CR mice to undergo approximately 1 day or 2 days of fasting, respectively, over the weekend.

Body weight was measured weekly, while a variety of other phenotypes were collected every 6 months or yearly[8] (Supplementary Table 1). The experiment was conducted at The Jackson Laboratory, and animal procedures were approved by the Animal Care and Use Committee at The Jackson Laboratory (protocol number 06005). Data collection and analysis were not performed blind to the conditions of the experiments.

Stool collection

Mice were scruffed, and one stool pellet was collected fresh. The pellet was added to RNAlater and stored at −80 °C.

DNA extraction

Stool pellets were removed from RNAlater with clean tweezers. DNA was extracted with the QIAGEN DNeasy PowerSoil Pro Kit (catalogue number 47016) following manufacturer instructions with one modification: pellets were homogenized for 1 min using the MP Biomedicals FastPrep-24 Classic Bead Beating Grinder and Lysis System, rather than using a vortex adaptor. DNA was eluted in 50 µl of C6. DNA concentration was measured using a NanoDrop spectrophotometer. DNA samples with concentrations <10 ng µl^{-1} were excluded from further data generation. DNA was extracted

across 124 batches of 24 or 48 samples. We included at least one negative control (nothing added to the lysis tube at the start of the protocol) and one positive control every ~100 samples. We used two different positive controls: ATCC 10 Strain Even Mix Whole Cell Material (catalogue number MSA-2003) or ZymoBIOMICS Gut Microbiome Standard (catalogue number D6331). These controls contain cells from 10 and 21 microorganisms, respectively, with a variety of relative abundances.

Library preparation

Extracted DNA was prepared for sequencing using the Illumina DNA Prep Kit (catalogue number 20060059). We added 7.5 µl of input DNA diluted to 3–22 ng μl^{-1}. We modified the protocol to use one-fourth of the manufacturer-recommended reagent volumes. Samples were barcoded using IDT for Illumina DNA/RNA UD Indexes Sets A and B (catalogue numbers 20027213 and 20027214), except for a handful of samples that were barcoded with Nextera DNA CD Indexes (catalogue number 20018708). Samples were eluted in 35 µl of elution buffer. DNA concentration was measured using the Qubit dsDNA High Sensitivity Assay Kit (catalogue number Q32851). If the library preparation was unsuccessful (that is, concentration <1 ng μl^{-1}), we attempted to redo the library starting from the input DNA. We examined a random subset of samples from each batch on a TapeStation 4200 instrument using a High Sensitivity D1000 ScreenTape (catalogue number 5067-5584). The library preparation was performed in 96-well plates across 70 batches of 48 or 96 samples each. We included at least one negative control (7.5 µl of dilution buffer added to a well at the start of the protocol) and one positive control every ~100 samples. The positive control was the ATCC 10 Strain Staggered Mix Genomic Material (catalogue number MSA-1001), which contains DNA from 10 bacteria at a variety of relative abundances.

Pooling and sequencing

Libraries were pooled and sequenced on a NovaSeq 6000 with paired ends (2 × 150 bp). Sequencing was performed across eight sequencing runs and two NovaSeq 6000 machines: one at Calico Life Sciences LLC and one at the University of Pennsylvania.

Data preprocessing

We performed quality control of our sequencing reads using the Snakemake pipeline Sunbeam[61] (v2.1.1). More specifically, we removed adaptors with cutadapt[62] (v3.1, forward and reverse adaptors = CTGTCTCTTATACACATCT), trimmed low-quality bases and discarded low-quality reads with trimmomatic[63] (v0.39, ILLUMINACLIP:trimmomatic/adapters/NexteraPE-PE.fa:2:30:10:8:true LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36), discarded reads with many repetitive sequences with komplexity (https://github.com/eclarke/komplexity) and removed host reads (mean 9% of input reads) using bwa[64] (v0.7.17) against the mm10 genome.

Quality control

We collected 4,214 stool samples from 944 DO mice (16 mice died before the first collection). Of these 4,214 stool samples, 3,586 stool samples (85%) were sequenced. The mean sequencing depth was 14.1 M read pairs per sample. We were unable to sequence 628 stool samples because either we could not successfully extract DNA (the DNA concentration had to be >10 ng μl^{-1}) or we could not prepare a library from the DNA (the library concentration had to be >1 ng μl^{-1}) despite multiple attempts. After accounting for positive and negative controls and some stool samples being sequenced more than once, we sequenced 4,352 samples.

Samples were discarded for any of the following reasons (in this order):

(1) The sample did not definitively come from the expected mouse ('Identifying sample mix-ups'), $n = 775$
(2) We discovered that the date of sample collection was after the animal's date of death, so we could not be certain that the stool sample was correctly identified, $n = 6$
(3) The sample received too few reads (<750,000 read pairs for stool samples, <100,000 read pairs for positive controls), $n = 29$
(4) The proportion of reads assigned to the mouse genome was suspiciously high (>50%), $n = 53$
(5) The sample was very different from all other samples ('Discarding outliers'), $n = 13$

After these quality-control steps, we were left with 3,473 samples, corresponding to 2,997 stool samples (71% of the original 4,214) from 913 DO mice.

Identifying sample mix-ups

Because every DO mouse was genotyped, we could compare the small fraction (~9%) of host reads in each stool sample to each mouse genome to confirm that every stool sample came from the expected mouse. We did this by adapting a previously published pipeline[5]. The idea is to count the number of times that host reads are discordant with the host genotype. After running this pipeline on all samples, we discarded 775 samples and renamed 111 samples. Notably, 136 of the discarded samples were due to one mistake (an index collision between two library batches), and 225 of the discarded samples may not have been mix-ups but the result of coprophagy. The small number of renamed samples involved situations in which we could confidently identify the cause of the mistake. For complete details, please see Supplementary Note 1.

Discarding outliers

We calculated all pairwise sample distances to identify potential outliers (Extended Data Fig. 3a). The largest distances were enriched for 13 samples: the 45,348 pairwise distances greater than 0.9 involved 3,310 samples, but just 13 samples were involved in 48.3% of these distances. These 13 samples were considered outliers and omitted from analysis.

Taxonomic and functional classification

After quality control, we performed taxonomic and functional classification with two distinct approaches. First, we used HUMAnN3 (ref. 42) for taxonomic (MetaPhlAn4, ref. 65) and functional profiling with the following databases: mpa_vOct22_CHOCOPhlAnSGB_202212, full_chocophlan.v201901_v31 and uniref90_annotated_v201901b. We used all default parameters and added '--unclassified-estimation' to estimate the proportion of unclassified reads. We also performed taxonomic profiling with a second approach: Kraken2 (ref. 40, v2.1.2) with the MGBC[41] and default parameters. Because the Kraken2 plus MGBC approach classified more reads (Extended Data Fig. 3b) and returned fewer uncharacterized taxa, we present Kraken taxonomic results except in three places: positive control stacked bar plots (because MGBC does not contain all microorganisms present in the positive controls; Extended Data Fig. 1c), when verifying key findings and in comparisons to the other datasets ('Comparison to B6 and human datasets').

We identified 376 genera and 482 species using MetaPhlAn, 252 genera and 1,093 species with Kraken and 422 MetaCyc[43] pathways with HUMAnN3. Because so many microorganisms in the mouse gut microbiome are uncharacterized, we focused on genera rather than species for better interpretability. We verify certain key findings using species-level taxonomic results.

To account for differences in sequencing depth, we calculated genus-level relative abundances (exclusive of unclassified reads), and we divided pathway reads-per-kilobase (RPK) abundances by the sample sum (excluding UNMAPPED and UNINTEGRATED) and multiplied by 1 million. This is equivalent to the transcripts-per-million (TPM) normalization used in RNA sequencing[66], so we adopt this nomenclature even though we are not referring to transcripts.

Samples produced from the same stool sample were aggregated together. For Kraken taxonomic results, we aggregated by summing absolute counts. For HUMAnN results, we aggregated by taking the mean genus-level relative abundance and mean pathway TPM abundance.

We distinguished between 'community-wide' and 'specialized' pathways on the basis of their similarity to taxonomic features. Specialized pathways were defined as having a small number (1–4) of Pearson correlations with genera above 0.5. Correlations were calculated using centred log ratio-transformed abundances for genera and $\log_2$(TPM) abundances for pathways.

Data normalization before linear modelling

For downstream linear modelling, we used centred log ratio-transformed, genus-level abundances and pathway $\log_2$(TPM) abundances after replacing zeroes. Zeroes were replaced with each feature's minimum non-zero value divided by two.

We excluded low prevalence features. For Kraken taxonomic results, we considered the 100 most abundant genera (based on total counts across all samples), which accounted for >99% of all genus-level counts. For HUMAnN, we retained pathways with at least 100 TPM in at least 10% of samples (273 of 422 pathways).

We included several community features in our linear modelling: three genus-level measures of α-diversity (Shannon index, Simpson index and Chao1 index), the first three principal coordinates (based on Bray–Curtis distance of genus-level relative abundances), taxonomic uniqueness (also based on genus-level Bray–Curtis distance) and functional uniqueness (based on Euclidean distance of $\log_2$(TPM) pathway abundances). Uniqueness is defined as the distance (or β-diversity) of a microbiome sample to its nearest neighbour[27].

Finally, before use in a linear model, features were scaled so that estimated coefficients were comparable across features.

PCoA and PCA

Principal coordinate analysis (PCoA) plots were based on Bray–Curtis distances of genus-level relative abundances. The principal component (PCA) plot was based on Euclidean distance of pathway $\log_2$(TPM) abundances.

PERMANOVA

We performed PERMANOVA using the adonis2 function from the vegan[67] R package with the following parameters: formula = dist ~ age + DR, by = "margin", permutations = 999. Age is a continuous variable encoding age in months; DR is a categorical variable encoding each of the five dietary interventions (samples collected before dietary randomization are considered AL). We ran PERMANOVA on all 2,997 samples. For genera, we used Bray–Curtis distance on relative abundances. For pathways, we used Euclidean distance on $\log_2$(TPM) abundances.

Linear mixed model per microbiome feature

To assess the influence of age, DR and genetics on each microbiome feature (y_mb), we fit the following linear mixed model:

(1) y_mb ~ age + DR + (1|quarter) + (1|mouse) + (1|cohort) + (1|batch) + (1|cage) + (1|genetics)

where y_mb is defined above ('Data normalization before linear modelling'), age (in weeks, scaled) is a fixed effect, DR is a fixed effect with five levels (AL, 1D, 2D, 20, 40), (1|quarter) is a random intercept corresponding to the yearly quarter of stool collection (for example, 1 January 2019 to 31 March 2019; $n = 18$ quarters), (1|mouse) is a random intercept, (1|cohort) is a random intercept corresponding to $n = 12$ DO breeding cohorts, (1|batch) is a random intercept corresponding to $n = 124$ microbiome DNA extraction batches, (1|cage) is a random intercept corresponding to the cage in which a mouse was housed for the entirety of its life ($n = 120$), and (1|genetics) is a random effect corresponding to additive genetic effects, which we encode by providing the kinship matrix. In ecology, this model is referred to as the repeated measures animal model[68]. The (1|mouse) random effect accounts for 'repeatability' or 'permanent environment effects', while (1|genetics) accounts strictly for additive genetic effects.

We used the previously published[69] kinship matrix for this cohort of DO mice. Importantly, we multiplied the kinship matrix (in which the diagonal was approximately 0.5) by 2 before using it in the linear mixed model. This step is necessary because the genetic covariance matrix in a linear mixed model must contain coefficients of relationships, which are twice the kinship coefficients[70].

The significance of each fixed effect was evaluated using a conditional Wald test. The significance of each random effect was evaluated using a likelihood ratio test where the null model omitted the random effect. P values were adjusted with the Benjamini–Hochberg procedure, separately for taxonomic and functional features. Adjusted $P < 0.01$ was considered significant.

Model 1 was fit using ASReml-R[71] (v4.1.0.716).

Effect of sampling timepoint

Besides implicitly accounting for temporal batch effects in the DNA extraction batch and DO mouse breeding cohort, we also assessed the importance of explicitly accounting for sampling timepoint in our model (Extended Data Fig. 4) by running three additional models. We used a longitudinal model that accounted for sampling timepoint ('time') as a fixed effect, rather than a random intercept:

(2) y_mb ~ age + DR + time + (1|mouse) + (1|cohort) + (1|batch) + (1|cage) + (1|genetics)

where time is a fixed effect corresponding to the date of stool collection, encoded as the number of days (scaled) from the first stool collection on 31 October 2016.

We ran a longitudinal model that did not include sampling timepoint:

(3) y_mb ~ age + DR + (1|mouse) + (1|cohort) + (1|batch) + (1|cage) + (1|genetics)

And we used cross-sectional models in which all samples were collected around the same time:

(4) y_mb ~ age + DR + (1|cohort) + (1|batch) + (1|cage) + (1|genetics)

Model 4 omitted the random intercept for mouse identity because the data were no longer longitudinal. We ran this cross-sectional model on three data slices:

- Slice 1 = 573 samples collected from 20 November 2017 to 5 March 2018 (mean age = 11.1 months, 25th percentile = 4.8 months, 75th percentile = 15.9 months)
- Slice 2 = 424 samples collected from 30 April 2018 to 13 August 2018 (mean age = 17.2 months, 25th percentile = 10.1 months, 75th percentile = 22.1 months)
- Slice 3 = 381 samples collected from 22 October 2018 to 11 February 2019 (mean age = 20.5 months, 25th percentile = 15.9 months, 75th percentile = 22.1 months)

Models 2–4 were fit using ASReml-R[71] (v4.1.0.716).

Age prediction

We used a random forest regressor (randomForest R package v4.7–1.1 with default parameters) to predict the age of a mouse based on its microbiome profile. We performed age prediction in three different contexts: (1) predicting the age of DO mice, (2) predicting the age of DR mice using a regressor trained on AL mice and vice versa, and (3) predicting age in the cohousing experiment.

To predict the age of DO mice (Fig. 2 and Extended Data Fig. 5), we performed 10-fold cross-validation, that is, 90% of samples were used for training, and predictions were made on the remaining 10% of samples. Age was treated as a continuous variable. We performed age prediction considering just AL mice (n = 573, in which we avoid the confounding variable of dietary restriction) or all samples (n = 2,997). We used relative abundances for the top 100 most abundant genera, relative abundances for the top 200 most abundant species or $\log_2$(TPM) abundances for the 272 pathways that passed prevalence filtration.

To evaluate whether DR rejuvenates the microbiome (Fig. 5 and Extended Data Fig. 9), we trained a random forest regressor on all AL samples and predicted age for all DR samples. We used out-of-bag predictions when reporting AL accuracy. Conversely, we trained a regressor on all 40% CR samples and evaluated it on all other samples, including the other DR groups. Age was treated as a continuous variable, and we again considered 100 genera or 272 pathways.

To predict the age of mice in the cohousing experiment, a random forest classifier was trained on all baseline samples and tested on all other samples. Age was treated as a binary variable (young or old). We considered relative abundances for all 125 genera ('Cohousing experiment').

We evaluated feature importance using the 'importance' function from the randomForest package. For regression, we quantified importance using the percentage increase in mean squared error (MSE) and, for classification, we used the mean decrease in accuracy.

Longitudinal B6 mouse cohort

A total of 15 4-week-old male C57BL/6 ('B6') mice were ordered from The Jackson Laboratory. The mice were housed five per cage across three cages. Stool pellets were collected upon arrival and then every 3 months. The pellets were collected fresh into empty 1.7-ml tubes and frozen at −80 °C. Animal procedures were approved by the Institutional Animal Care and Use Committee at the Perelman School of Medicine, University of Pennsylvania (protocol number 806361).

DNA was extracted with the Qiagen DNeasy PowerSoil Kit (catalogue number 12888-100), libraries were prepared with the Nextera DNA Flex Library Prep Kit (catalogue number 20018705) and libraries were sequenced on a NextSeq 550 machine with 75-bp single-end sequencing. Data were analysed the same way as the DO samples: pre-processed with Sunbeam and then taxonomically and functionally classified with HUMAnN.

Human metagenomic data

Human metagenomic sequencing data were obtained using the curatedMetagenomicData[47] package (v3.6.2). We filtered for stool samples from individuals meeting the following criteria: age ≥18 years, 'healthy' or 'control', and no current antibiotic use. Furthermore, we included only studies with ≥50 individuals meeting these criteria and an age interquartile range ≥5 (to make sure each study had a diversity of ages). This resulted in 4,101 individuals from 20 studies. We obtained genus-level relative abundances and pathway RPK abundances, which we normalized to $\log_2$(relative abundances) and $\log_2$(TPM) values after replacing zeros, as described above.

Comparison to B6 and human datasets

We considered only samples from AL mice when comparing the DO cohort with the B6 cohort and human samples. We fit the following linear models to identify age-associated microbiome features in each dataset:

(5) DO AL mice: y_mb ~ age + (1|mouse) + (1|cohort) + (1|batch) + (1|cage)

(6) B6 mice: y_mb ~ age + (1|cage)

(7) Humans: y_mb ~ age + (1|study)

where (1|study) is a random intercept corresponding to one of 20 studies comprising the human cohort. See 'Linear mixed model per microbiome feature' for details about the other terms. Models 5, 6 and 7 were fit with MaAsLin2 (ref. 72, v1.12.0).

Because the human dataset had been processed with HUMAnN, we used MetaPhlAn taxonomic results for the DO AL mice, instead of the Kraken taxonomic results. We used $\log_2$-transformed genus-level relative abundances after zero replacement. For prevalence filtration, we retained genera with at least 0.001% relative abundance in at least 10% of samples (DO AL: 248 of 311, B6: 262 of 310, humans: 90 of 331) and pathways with at least 100 TPM in at least 10% of samples (DO AL: 262 of 399, B6: 233 of 329, humans: 358 of 573).

We also tested several measures of α-diversity (Shannon index, Simpson index, inverse Simpson index and richness) and uniqueness. For DO AL mice, uniqueness was recalculated considering just AL mice. For humans, uniqueness was calculated separately within each of 20 studies.

P values for each age coefficient were calculated with MaAsLin2 (based on a t-statistic using Satterthwaite's method for denominator degrees of freedom[73]). P values were adjusted with the Benjamini–Hochberg procedure, separately per dataset and separately for taxonomic and functional features. Due to the additional burden of needing to be consistent across datasets, the adjusted P value threshold was increased to 0.1 for this analysis.

For select microbiome features, we also fit the following basic linear model separately within each human study to assess consistency across studies (Extended Data Fig. 6b–d):

(8) Within each human study: y_mb ~ age

The 20 P values (one from each human study) were adjusted with the Benjamini–Hochberg procedure.

Cohousing experiment

We obtained n = 25 8-week-old female C57BL/6 mice from The Jackson Laboratory and n = 20 19- and 20-month-old female C57BL/6 mice from the National Institute on Aging. Before the start of the experiment, the mice were housed five per cage with other mice of the same age. The mice were allowed to acclimate for at least 1 week before the start of the experiment. During cohousing, three young mice were housed with two old mice for 1 month. Control young and old mice remained housed with other mice of the same age. After 1 month, two of the cohousing cages were separated by age and two cohousing cages remained cohoused. Stool pellets were collected at baseline, after 1 month of cohousing, and after 2, 4, 6 and 8 weeks of separation. Animal procedures were approved by the Institutional Animal Care and Use Committee at the Perelman School of Medicine, University of Pennsylvania (protocol number 806361).

We performed 16S sequencing of these stool samples. DNA was extracted using the Qiagen DNeasy PowerSoil Pro Kit (catalogue number 47016). We amplified the V1/V2 variable region using KAPA HiFi HotStart ReadyMix (Roche, catalogue number KK2602) and the F27/R338 primer pair (F27: 5′-AGAGTTTGATCCTGGCTCAG-3′, R338: 5′-TGCTGCCTCCCGTAGGAGT-3′). We performed a bead-based clean-up of the pooled libraries using AMPure XP SPRI beads (catalogue number A63881). Libraries were paired-end (2 × 250 bp) sequenced across two MiSeq runs. The first run contained samples collected at baseline, at 4 weeks of cohousing and at 4 weeks of separation. The second run contained samples collected at 2, 6 and 8 weeks of separation. Data were processed using QIIME2 (ref. 74, v2023.2.0). Reads were demultiplexed[75] and denoised with DADA2 (ref. 76). The following parameters were supplied to DADA2 for trimming demultiplexed reads: '--p-trunc-len-f 240 --p-trunc-len-r 220' for the first run and '--p-trunc-len-f 250 --p-trunc-len- 165' for the second run. Reads from the two runs were then merged and

taxonomically classified against the SILVA 138 database[77,78] using a naive Bayes classifier[79,80].

Cohousing experiment in germ-free mice

We collected stool pellets from $n = 24$ young (3 months) and $n = 18$ old (17 months) female C57BL/6 mice. These donor mice had been housed four to five per cage. Pellets from mice in the same cage were combined in a single 5-ml tube with 5–10 pellets per tube. Pellets were resuspended in 2 ml of PBS and vortexed, and then 200 µl of this faecal slurry was gavaged into recipient germ-free (GF) mice. The recipients were $n = 44$ young (3–5 months) GF female C57BL/6 mice from the Penn Gnotobiotic Mouse Facility. Recipients were housed five per cage for 2 weeks in isocages to allow the FMT to engraft. Recipients in the same isocage received input from the same donor cage.

Two weeks after the FMT, we cohoused recipients of young microbiomes (CY_{GF}) with recipients of old microbiomes (CO_{GF}). Cohoused cages contained three CY_{GF} and two CO_{GF} mice. Non-cohoused control cages contained five young FMT recipients (Y_{GF}) or four to five old FMT recipients (O_{GF}). After 2 weeks and 1 month of cohousing, we collected stool pellets from all recipients.

We performed 16S sequencing of these stool samples as above. DNA was extracted using the Qiagen DNeasy PowerSoil Pro Kit, we amplified the V1/V2 variable region using the F27/R338 primer pair, we performed a bead-based clean-up of the pooled libraries using AMPure XP SPRI beads, and we sequenced libraries (2 × 250 bp) across two MiSeq runs. The first run contained samples collected at baseline (2 weeks after the FMT) and at 1 month of cohousing. The second run contained samples collected at 2 weeks of cohousing. Again, data were processed with QIIME2. The following parameters were supplied to DADA2 for trimming demultiplexed reads: '--p-trim-left-f 14 --p-trunc-len-f 249 --p-trunc-len-r 231' for the first run and '--p-trunc-len-f 249 --p-trunc-len-r 231' for the second run. Reads from the two runs were then merged and taxonomically classified against the SILVA 138 database.

Heritability

Using model 1, heritability was calculated as the variance assigned to the (1|genetics) random effect divided by total variance. The standard error of heritability was estimated using the vpredict function from ASReml. A feature was considered heritable if it had an adjusted $P < 0.01$ (likelihood ratio test, followed by Benjamini–Hochberg for P value adjustment). We also calculated heritability (model 1) using lme4qtl[81](v0.2.2) to confirm identical results with a different software package (Extended Data Fig. 7e).

We also calculated heritability with the following cross-sectional model, separately at each of 5, 10, 16, 22 and 28 months (Extended Data Fig. 7g):

(9) y_mb ~ DR + (1|cohort) + (1|batch) + (1|cage) + (1|genetics)

The DR term was omitted from the 5-month analysis because this was before dietary randomization. Model 9 was fit with ASReml. For this cross-sectional analysis, heritability was computed using all samples at a given age. We also fit model 1 after downsampling to 110 mice to obtain a similar number of samples as in cross-sectional data (Extended Data Fig. 7g).

Comparison to heritability estimates from other studies

For Extended Data Fig. 7f, we compared our heritability estimates to those reported in Supplementary Table 3A (ref. 51). This table contained 27 operational taxonomic units (OTUs) that could be classified to the level of genus. Of these 27, only 8 overlapped with the genera that we tested for heritability. For Fig. 4b, we plotted the proportion of significantly heritable taxa (based on the authors' definitions) from eight other studies[30–33,51,82–84].

Comparing all experimental variables

To compare the effects of all experimental variables to each other, we used a modified version of model 1 in which age, DR and sampling timepoint ('quarter') were encoded as random intercepts:

(10) y_mb ~ (1|age) + (1|DR) + (1|quarter) + (1|mouse) + (1|cohort) + (1|batch) + (1|cage) + (1|genetics)

This allowed us to compare the variance explained by each variable. Age was provided as a categorical variable (nine levels). P values were adjusted separately for each experimental variable. Model 10 was fit with ASReml.

QTL mapping

We performed QTL mapping as described previously[85]. Briefly, we tested the association between each SNP marker against each of 107 microbiome features (top 100 genera plus 7 community features) at each of 5 ages: 5, 10, 16, 22 and 28 months. Dietary group and DO mouse cohort were included as covariates, except at 5 months (before the start of DR) when only cohort was included as a covariate. QTL mapping was performed with R/qtl2 (ref. 86) using previously published leave-one-chromosome-out kinship matrices and genotype probabilities[85]. P values were calculated based on permutation and adjusted using the Benjamini–Hochberg procedure.

Prediction of dietary group

We used random forest classifiers (randomForest R package v4.7–1.1 with default parameters) to predict binary dietary restriction status (AL or DR) or dietary restriction group (AL, 1D, 2D, 20% or 40%). We performed prediction separately per age. Within each age, we performed 5-fold (for predicting binary dietary restriction) or 10-fold (for predicting dietary group) cross-validation while stratifying by cage so that no samples from the same cage were present in both the training and testing sets. We used the same 100 genera and 272 pathways as above.

Microbiome–phenotype associations

To identity associations between microbiome features and host phenotypes, we tested all microbiome–phenotype pairs with the following model:

(11) y_pheno ~ y_mb + age + DR + (1|mouse)

where y_pheno is a phenotypic trait such as body weight, age and DR are fixed effects, and (1|mouse) is a random intercept. For each phenotype, we selected the measurement closest in time to each microbiome sample and included only microbiome–phenotype pairs obtained within 100 days of each other. The significance of each microbiome–phenotype association was calculated using a likelihood ratio test where the null model omitted y_mb. We tested 100 genera and 252 pathways against 197 phenotypic traits measured across 13 assays.

Model 11 could not be used to test for association for lifespan (a non-longitudinal measurement), so we fit the following cross-sectional model at each of 5, 10, 16, 22 and 28 months:

(12) y_pheno ~ y_mb + DR

The DR term was omitted from the 5-month analysis because this was before dietary randomization. The significance of each association was calculated using a likelihood ratio test where the null model omitted y_mb. Model 11 was fit with lme4 (ref. 87, v1.1-33) and model 12 was fit with the lm function in R.

Mediation

Mediation analysis estimates the proportion of treatment T's effect on outcome Y that is mediated by mediator M (see Supplementary Discussion for limitations of this approach). In our study, treatment T

is dietary restriction, outcome *Y* is a phenotypic trait and mediator *M* is a microbiome feature. We performed mediation analysis using the model-based approach within the mediate R package[88] (v4.5.0). We fit the following two models:

(13) Mediator model: y_mb ~ DR_X + age + (1|mouse)

(14) Outcome model: y_pheno ~ y_mb + DR_X + age + (1|mouse)

where DR_X corresponds to one of the four dietary interventions. Mediation analysis was performed separately for each of the DR groups. We report the average causal mediation effect (ACME) estimate and *P* value. *P* values were adjusted with Benjamini–Hochberg separately per dietary group. Microbiome–phenotype pairs with an ACME-adjusted $P < 0.01$ were considered significant. We tested the same features as for association analysis. When reporting the mean proportion mediated, we consider only significant mediation results where the direct effect and mediation effect estimates have the same sign.

Statistics

P values were adjusted using the Benjamini–Hochberg procedure, separately for taxonomic and functional features and for each model specification. For example, 107 *P* values were adjusted after running model 1 on genus-level taxonomic features. Significance was defined as an adjusted $P < 0.01$, unless explicitly stated otherwise. For pairwise comparisons, *t*-tests were used. All statistical tests were two sided except for one-sided *t*-tests in Fig. 5j and Extended Data Fig. 8l. Tests were unpaired except in Fig. 5d and Extended Data Fig. 8f. For box plots, boxes extend from the 25th to 75th percentiles, whiskers extend to 1.5 times the interquartile range and the centre line is the median. Error bars show one standard error, except in Extended Data Fig. 3c where error bars show one standard deviation. *P* value symbols are defined as follows: NS, $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$. *P* values below 2.2×10^{-16} (the epsilon of a double-precision float) are reported as '$<2.2 \times 10^{-16}$'. Downstream analysis and plotting was performed in RStudio (R v4.2) using the tidyverse[89] (v2.0.0), phyloseq[90] (v1.42.0) and vegan[67] (v2.6.4) packages. Final figures were created with Adobe Illustrator.

Reporting summary

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

Data availability

We generated the following fastq files, which are available in Sequence Read Archive: DRiDO metagenomic sequencing, PRJNA1054518; longitudinal B6 metagenomic sequencing, PRJNA1073968; 16S sequencing of the cohousing experiment, PRJNA1072097; and 16S sequencing of cohousing in the germ-free mice experiment, PRJNA1128683. Processed data are available via GitHub at https://github.com/levlitichev/DRiDO_microbiome. Host phenotypes collected as part of the DRiDO study are available via Figshare at https://doi.org/10.6084/m9.figshare.24600255.v1. The genetic kinship matrix and genotype probabilities are available via Figshare at https://doi.org/10.6084/m9.figshare.13190735. Other databases used in this study include the mm10 genome (http://igenomes.illumina.com.s3-website-us-east-1.amazonaws.com/Mus_musculus/UCSC/mm10/Mus_musculus_UCSC_mm10.tar.gz), MetaPhlan (http://cmprod1.cibio.unitn.it/biobakery4/metaphlan_databases/) and HUMAnN databases (http://huttenhower.sph.harvard.edu/humann_data/chocophlan/full_chocophlan.v201901_v31.tar.gz, http://huttenhower.sph.harvard.edu/humann_data/uniprot/uniref_annotated/uniref90_annotated_v201901b_full.tar.gz), MGBC Kraken2 database (https://github.com/BenBeresfordJones/MGBC), curatedMetagenomicData (https://doi.org/10.18129/B9.bioc.curatedMetagenomicData) and SILVA 138 (https://www.arb-silva.de/documentation/release-138/).

Code availability

All figures in this paper may be reproduced using code (and processed data) available via GitHub at https://github.com/levlitichev/DRiDO_microbiome. This GitHub repository also contains an example analysis tutorial.

References

1. Green, C. L., Lamming, D. W. & Fontana, L. Molecular mechanisms of dietary restriction promoting health and longevity. *Nat. Rev. Mol. Cell Biol.* **23**, 56–73 (2022).
2. Fontana, L. & Partridge, L. Promoting health and longevity through diet: from model organisms to humans. *Cell* **161**, 106–118 (2015).
3. Liao, C.-Y., Rikke, B. A., Johnson, T. E., Diaz, V. & Nelson, J. F. Genetic variation in the murine lifespan response to dietary restriction: from life extension to life shortening. *Aging Cell* **9**, 92–95 (2010).
4. Mitchell, S. J. et al. Effects of sex, strain, and energy intake on hallmarks of aging in mice. *Cell Metab.* **23**, 1093–1112 (2016).
5. Pak, H. H. et al. Fasting drives the metabolic, molecular and geroprotective effects of a calorie-restricted diet in mice. *Nat. Metab.* **3**, 1327–1341 (2021).
6. Mitchell, S. J. et al. Daily fasting improves health and survival in male mice independent of diet composition and calories. *Cell Metab.* **29**, 221–228.e3 (2019).
7. Acosta-Rodríguez, V. et al. Circadian alignment of early onset caloric restriction promotes longevity in male C57BL/6J mice. *Science* **376**, 1192–1202 (2022).
8. Di Francesco, A. et al. Dietary restriction impacts health and lifespan of genetically diverse mice. *Nature* **634**, 684–692 (2024).
9. Bárcena, C. et al. Healthspan and lifespan extension by fecal microbiota transplantation into progeroid mice. *Nat. Med.* **25**, 1234–1242 (2019).
10. Smith, P. et al. Regulation of life span by the gut microbiota in the short-lived African turquoise killifish. *eLife* **6**, e27014 (2017).
11. Kim, K. H. et al. Gut microbiota of the young ameliorates physical fitness of the aged in mice. *Microbiome* **10**, 238 (2022).
12. Fabbiano, S. et al. Functional gut microbiota remodeling contributes to the caloric restriction-induced metabolic improvements. *Cell Metab.* **28**, 907–921.e7 (2018).
13. Jie, Z. et al. The baseline gut microbiota directs dieting-induced weight loss trajectories. *Gastroenterology* **160**, 2029–2042.e16 (2021).
14. Diener, C. et al. Baseline gut metagenomic functional gene signature associated with variable weight loss responses following a healthy lifestyle intervention in humans. *mSystems* **6**, e00964-21 (2021).
15. Nielsen, R. L. et al. Data integration for prediction of weight loss in randomized controlled dietary trials. *Sci. Rep.* **10**, 20103 (2020).
16. Langille, M. G. et al. Microbial shifts in the aging mouse gut. *Microbiome* **2**, 50 (2014).
17. Low, A., Soh, M., Miyake, S. & Seedorf, H. Host age prediction from fecal microbiota composition in male C57BL/6J mice. *Microbiol. Spectr.* **10**, e00735-22 (2022).
18. Zhang, C. et al. Structural modulation of gut microbiota in life-long calorie-restricted mice. *Nat. Commun.* **4**, 2163 (2013).
19. Odamaki, T. et al. Age-related changes in gut microbiota composition from newborn to centenarian: a cross-sectional study. *BMC Microbiol.* **16**, 90 (2016).
20. Biagi, E. et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS ONE* **5**, e10667 (2010).
21. Johansen, J. et al. Centenarians have a diverse gut virome with the potential to modulate metabolism and promote healthy lifespan. *Nat. Microbiol.* **8**, 1064–1078 (2023).

22. Zhang, X. et al. Sex- and age-related trajectories of the adult human gut microbiota shared across populations of different ethnicities. *Nat. Aging* **1**, 87–100 (2021).
23. Badal, V. D. et al. The gut microbiome, aging, and longevity: a systematic review. *Nutrients* **12**, 3759 (2020).
24. Martino, C. et al. Microbiota succession throughout life from the cradle to the grave. *Nat. Rev. Microbiol.* **20**, 707–720 (2022).
25. Rothschild, D. et al. An atlas of robust microbiome associations with phenotypic traits based on large-scale cohorts from two continents. *PLoS ONE* **17**, e0265756 (2022).
26. Ghosh, T. S., Shanahan, F. & O'Toole, P. W. Toward an improved definition of a healthy microbiome for healthy aging. *Nat. Aging* **2**, 1054–1069 (2022).
27. Wilmanski, T. et al. Gut microbiome pattern reflects healthy ageing and predicts survival in humans. *Nat. Metab.* **3**, 274–286 (2021).
28. Rothschild, D. et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* **555**, 210–215 (2018).
29. Carmody, R. N. et al. Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17**, 72–84 (2015).
30. Kurilshikov, A. et al. Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat. Genet.* **53**, 156–165 (2021).
31. Goodrich, J. K. et al. Genetic determinants of the gut microbiome in UK twins. *Cell Host Microbe* **19**, 731–743 (2016).
32. Gacesa, R. et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732–739 (2022).
33. Turpin, W. et al. Association of host genome with intestinal microbial composition in a large healthy cohort. *Nat. Genet.* **48**, 1413–1417 (2016).
34. Bonder, M. J. et al. The effect of host genetics on the gut microbiome. *Nat. Genet.* **48**, 1407–1412 (2016).
35. Grieneisen, L. et al. Gut microbiome heritability is nearly universal but environmentally contingent. *Science* **373**, 181–186 (2021).
36. Thaiss, C. A. et al. Persistent microbiome alterations modulate the rate of post-dieting weight regain. *Nature* **540**, 544–551 (2016).
37. Sonnenburg, J. L. & Bäckhed, F. Diet–microbiota interactions as moderators of human metabolism. *Nature* **535**, 56–64 (2016).
38. Turnbaugh, P. J. et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
39. Lobo, A. K. et al. Identification of sample mix-ups and mixtures in microbiome data in diversity outbred mice. *G3* **11**, jkab308 (2021).
40. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
41. Beresford-Jones, B. S. et al. The Mouse Gastrointestinal Bacteria Catalogue enables translation between the mouse and human gut microbiotas via functional mapping. *Cell Host Microbe* **30**, 124–138.e8 (2022).
42. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife* **10**, e65088 (2021).
43. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **42**, D459–D471 (2014).
44. Huang, S. et al. Human skin, oral, and gut microbiomes predict chronological age. *mSystems* **5**, e00630-19 (2020).
45. Galkin, F. et al. Human gut microbiome aging clock based on taxonomic profiling and deep learning. *iScience* **23**, 101199 (2020).
46. Chen, Y. et al. Human gut microbiome aging clocks based on taxonomic and functional signatures through multi-view learning. *Gut Microbes* **14**, 2025016 (2022).
47. Pasolli, E. et al. Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **14**, 1023–1024 (2017).
48. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)* (Princeton Univ. Press, 2001).
49. Sherrill-Mix, S. et al. Allometry and ecology of the bilaterian gut microbiome. *mBio* https://doi.org/10.1128/mbio.00319-18 (2018).
50. Sieber, M. et al. Neutrality in the metaorganism. *PLoS Biol.* **17**, e3000298 (2019).
51. Schlamp, F. et al. High-resolution QTL mapping with diversity outbred mice identifies genetic variants that impact gut microbiome composition. Preprint at *bioRxiv* https://doi.org/10.1101/722744 (2021).
52. Kok, D. E. G. et al. Lifelong calorie restriction affects indicators of colonic health in aging C57Bl/6J mice. *J. Nutr. Biochem.* **56**, 152–164 (2018).
53. Pan, F. et al. Predominant gut *Lactobacillus murinus* strain mediates anti-inflammaging effects in calorie-restricted mice. *Microbiome* **6**, 54 (2018).
54. Fraumene, C. et al. Caloric restriction promotes rapid expansion and long-lasting increase of *Lactobacillus* in the rat fecal microbiota. *Gut Microbes* **9**, 104–114 (2017).
55. Kurup, K. et al. Calorie restriction prevents age-related changes in the intestinal microbiota. *Aging* **13**, 6298–6329 (2021).
56. Zeng, T. et al. Short-term dietary restriction in old mice rejuvenates the aging-induced structural imbalance of gut microbiota. *Biogerontology* **20**, 837–848 (2019).
57. Depommier, C. et al. Pasteurized *Akkermansia muciniphila* increases whole-body energy expenditure and fecal energy excretion in diet-induced obese mice. *Gut Microbes* **11**, 1231–1245 (2020).
58. Meyer, D. H. & Schumacher, B. Aging clocks based on accumulating stochastic variation. *Nat. Aging* **4**, 871–885 (2024).
59. Svenson, K. L. et al. High-resolution genetic mapping using the mouse diversity outbred population. *Genetics* **190**, 437–447 (2012).
60. Perea, C. et al. Caloric restriction in group-housed mice: littermate and sex influence on behavioral and hormonal data. *Front. Vet. Sci.* **8**, 639187 (2021).
61. Clarke, E. L. et al. Sunbeam: an extensible pipeline for analyzing metagenomic sequencing experiments. *Microbiome* **7**, 46 (2019).
62. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **17**, 10–12 (2011).
63. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
64. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
65. Blanco-Míguez, A. et al. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat. Biotechnol.* **41**, 1633–1644 (2023).
66. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* **131**, 281–285 (2012).
67. Oksanen, J. et al. vegan: Community Ecology Package. (2022).
68. Wilson, A. J. et al. An ecologist's guide to the animal model. *J. Anim. Ecol.* **79**, 13–26 (2010).
69. Wright, K. M. et al. Age and diet shape the genetic architecture of body weight in diversity outbred mice. *eLife* **11**, e64329 (2022).
70. Almasy, L. & Blangero, J. Variance component methods for analysis of complex phenotypes. *Cold Spring. Harb. Protoc.* **2010**, pdb.top77 (2010).
71. Gilmour, A. R. et al. *ASReml User Guide Release 4.1 Structural Specification* (VSN, 2015).
72. Mallick, H. et al. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput. Biol.* **17**, e1009442 (2021).

73. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* **82**, 1–26 (2017).
74. Bolyen, E. et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
75. Hamady, M. & Knight, R. Microbial community profiling for human microbiome projects: tools, techniques, and challenges. *Genome Res.* **19**, 1141–1152 (2009).
76. Callahan, B. J. et al. DADA2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
77. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2013).
78. Robeson, I. I. et al. RESCRIPt: reproducible sequence taxonomy reference database management. *PLoS Comput. Biol.* **17**, e1009581 (2021).
79. Bokulich, N. A. et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
80. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
81. Ziyatdinov, A. et al. lme4qtl: linear mixed models with flexible covariance structure for genetic studies of related individuals. *BMC Bioinformatics* **19**, 68 (2018).
82. Camarinha-Silva, A. et al. Host genome influence on gut microbial composition and microbial prediction of complex traits in pigs. *Genetics* **206**, 1637–1644 (2017).
83. Wallace, R. J. et al. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci. Adv.* **5**, eaav8391 (2019).
84. Li, F. et al. Host genetics influence the rumen microbiota and heritable rumen microbial features associate with feed efficiency in cattle. *Microbiome* **7**, 92 (2019).
85. Zhang, G. et al. Intermittent fasting and caloric restriction interact with genetics to shape physiological health in mice. *Genetics* **220**, iyab157 (2022).
86. Broman, K. W. et al. R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics* **211**, 495–502 (2019).
87. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
88. Tingley, D., Yamamoto, T., Hirose, K., Keele, L. & Imai, K. mediation: R package for causal mediation analysis. *J. Stat. Softw.* **59**, 1–38 (2014).
89. Wickham, H. et al. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
90. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).

Acknowledgements

We thank the members of the Thaiss and Levy laboratories for valuable discussions and feedback. For their assistance with sequencing, we thank M. Roy and T. Vijay at the genomics core at Calico Life Sciences LLC, J. Schug at the Penn Genomics and Sequencing Core Facility (RRID:SCR_024999) and A. M. Misic at Illumina, Inc. For their assistance with animal husbandry, we thank University Laboratory Animal Resources (ULAR) at the University of Pennsylvania, D. Kobuley and M. Albright for dedicated germ-free husbandry, and the animal husbandry team at The Jackson Laboratory. This work was in part supported by Calico Life Sciences LLC. L.L. was supported by the Blavatnik Family Fellowship in Biomedical Research and T32HG000046. E.R.D. was supported by NIH R35GM146980. G.A.C. was supported by NIH P30AG038070. M. Li was supported by NIH R01GM125301, R01EY030192, R01EY031209, R01HL113147 and R01HL150359. C.A.T. is a Pew Biomedical Scholar and a Burroughs Wellcome Fund Investigator in the Pathogenesis of Infectious Diseases, and is supported by NIH DP2-AG-067492, R01-DK-129691, R01-NS-134976 and DP1-DK-140021, the Kenneth Rainin Foundation, a McKnight Brain Research Foundation Innovator Award, the Human Frontier Science Program and the Penn Institute on Aging.

Author contributions

L.L.: investigation, formal analysis, data curation, writing—original draft and writing—review and editing. M. Considine: investigation. J.G.: investigation. V.S.: investigation. T.O.C.: investigation. H.C.D.: investigation. K.M.W.: software and formal analysis. K.R.A.: software and formal analysis. L.D.: validation. M.J.L.: validation. M.T.: investigation. M.R.G.-F.: investigation. A.C.W.: validation. P.L.: validation. J.K.: validation. G.T.U.: validation. R.J.R.: investigation. S.M.: investigation. C.M.: formal analysis. F.D.B.: supervision. A.R.: software. F.H.: formal analysis. Z.C.: formal analysis. G.V.P.: formal analysis. M.M.: formal analysis. A.G.D.: data curation. L.R.: investigation and data curation. C.T.: software. K.B.: supervision. M. Chakraborty: software. A.S.B.: supervision. H.L.: methodology and supervision. I.B.: methodology. E.R.D.: methodology. K.W.B.: methodology. M. Levy: supervision. R.L.C.: conceptualization and funding acquisition. D.B.: conceptualization and funding acquisition. A.F.: conceptualization, supervision and project administration. A.D.F.: supervision and project administration. G.A.C.: formal analysis, data curation and supervision. M. Li: methodology and supervision. C.A.T.: supervision, funding acquisition, writing—original draft and writing—review and editing.

Competing interests

K.M.W., A.R., F.H., Z.C., G.V.P., M.M., R.L.C., D.B., A.F. and A.D.F. are current or former employees of Calico Life Sciences LLC. The other authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41564-025-01963-3.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41564-025-01963-3.

Correspondence and requests for materials should be addressed to Christoph A. Thaiss.

Peer review information *Nature Microbiology* thanks David Hughes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

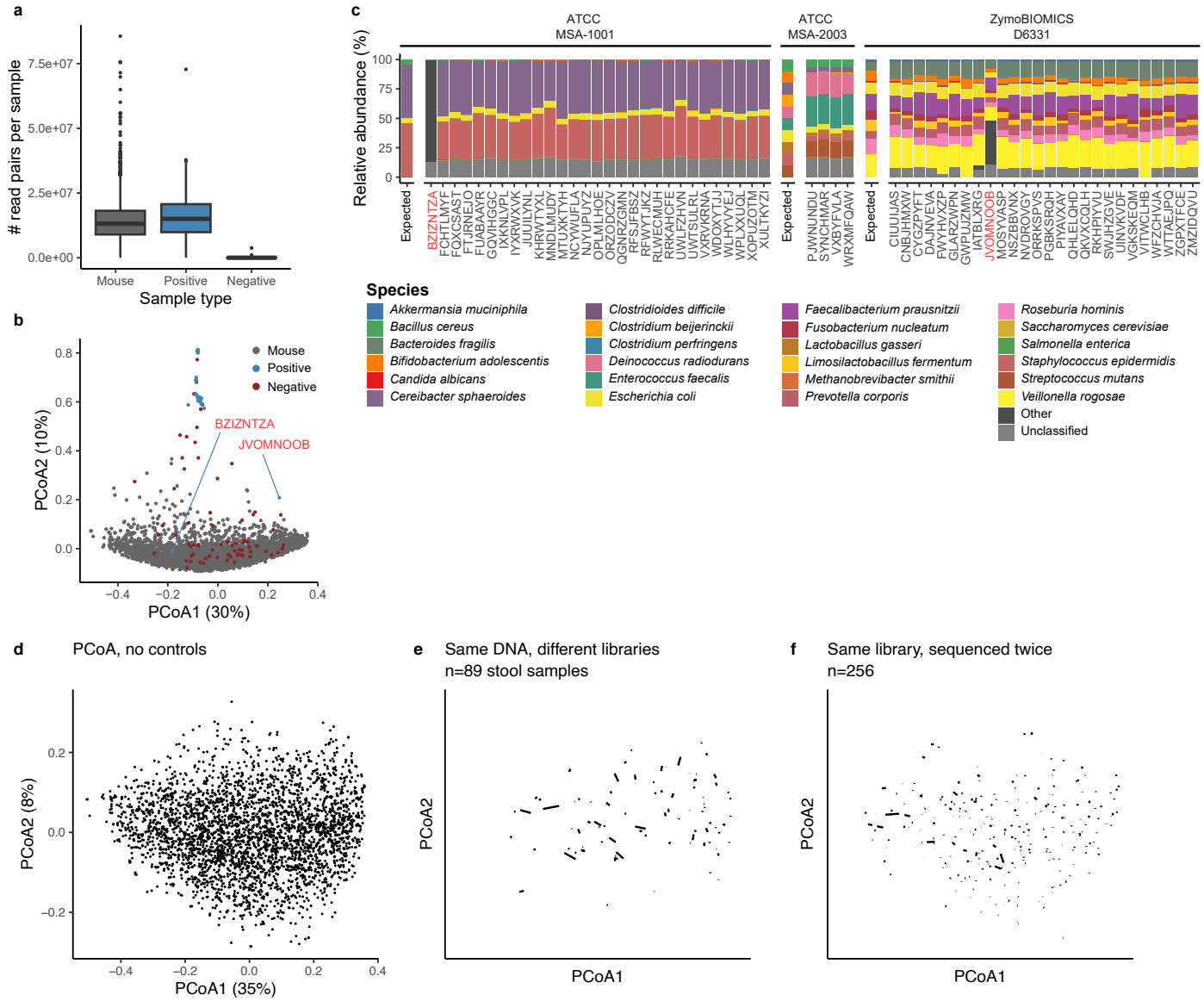
© The Author(s), under exclusive licence to Springer Nature Limited 2025

https://doi.org/10.1038/s41564-025-01963-3

Lev Litichevskiy[1,2], Maya Considine[1], Jasleen Gill[1], Vasuprada Shandar[1], Timothy O. Cox[1], Hélène C. Descamps[1], Kevin M. Wright[3], Kevin R. Amses[1], Lenka Dohnalová[1], Megan J. Liou[1], Monika Tetlak[1], Mario R. Galindo-Fiallos[1], Andrea C. Wong[1], Patrick Lundgren[1], Junwon Kim[1], Giulia T. Uhr[1], Ryan J. Rahman[1], Sydney Mason[1], Carter Merenstein[1], Frederic D. Bushman[1], Anil Raj[3], Fiona Harding[3], Zhenghao Chen[3], G. V. Prateek[3], Martin Mullis[3], Andrew G. Deighan[4], Laura Robinson[4], Ceylan Tanes[5], Kyle Bittinger[5,6], Meenakshi Chakraborty[7], Ami S. Bhatt[7,8], Hongzhe Li[9], Ian Barnett[2], Emily R. Davenport[10,11], Karl W. Broman[12], Maayan Levy[1,13,14], Robert L. Cohen[3], David Botstein[3], Adam Freund[3], Andrea Di Francesco[3,15], Gary A. Churchill[4,15], Mingyao Li[2,15] & Christoph A. Thaiss[1,13,14,15]

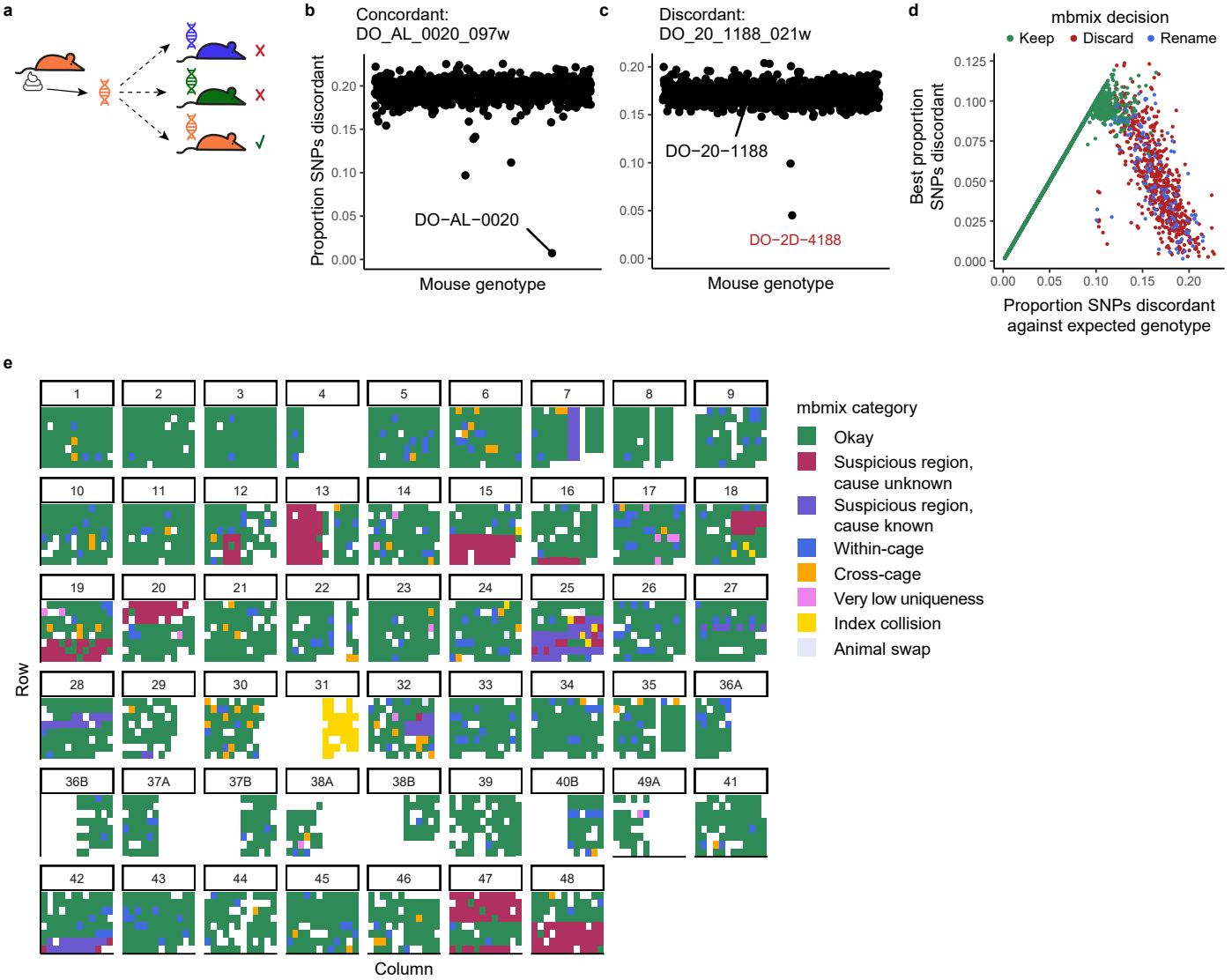
[1]Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [2]Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Calico Life Sciences LLC, South San Francisco, CA, USA. [4]The Jackson Laboratory, Bar Harbor, ME, USA. [5]Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, Perelman School of Medicine, Philadelphia, PA, USA. [6]Division of Pediatrics, Children's Hospital of Philadelphia, Perelman School of Medicine, Philadelphia, PA, USA. [7]Department of Genetics, Stanford University, Stanford, CA, USA. [8]Divisions of Hematology and Blood & Marrow Transplantation, Department of Medicine, Stanford University, Stanford, CA, USA. [9]Center for Clinical Epidemiology and Biostatistics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [10]Department of Biology, Pennsylvania State University, University Park, PA, USA. [11]Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA. [12]Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA. [13]Department of Pathology, Stanford University, Stanford, CA, USA. [14]Arc Institute, Palo Alto, CA, USA. [15]These authors jointly supervised this work: Andrea Di Francesco, Gary A. Churchill, Mingyao Li, Christoph A. Thaiss. e-mail: thaiss@stanford.edu

https://doi.org/10.1038/s41564-025-01963-3



Extended Data Fig. 1 | Positive and negative controls. a, Number of read pairs per sample (prior to aggregation), grouped by sample type (n = 3577 samples prior to aggregation by stool ID). Boxes extend from the 25th to 75th percentiles, whiskers extend to 1.5 times the interquartile range, and the center line is the median. **b**, PCoA of all samples prior to aggregation. Two positive controls (BZIZNTZA and JVOMNOOB, highlighted in red) clustered separately from the other positive controls. PCoA1 and PCoA2 explain 30% and 10% of overall variance, respectively. **c**, Species-level relative abundances (MetaPhlAn4) for positive controls. Two positive controls (BZIZNTZA and JVOMNOOB, highlighted in red) did not display the expected community composition. **d**, PCoA of non-control samples prior to aggregation. **e**, Same ordination as **d**, with lines connecting samples originating from the same DNA. **f**, Same ordination as **d**, with lines connecting samples in which a library was sequenced multiple times.

https://doi.org/10.1038/s41564-025-01963-3

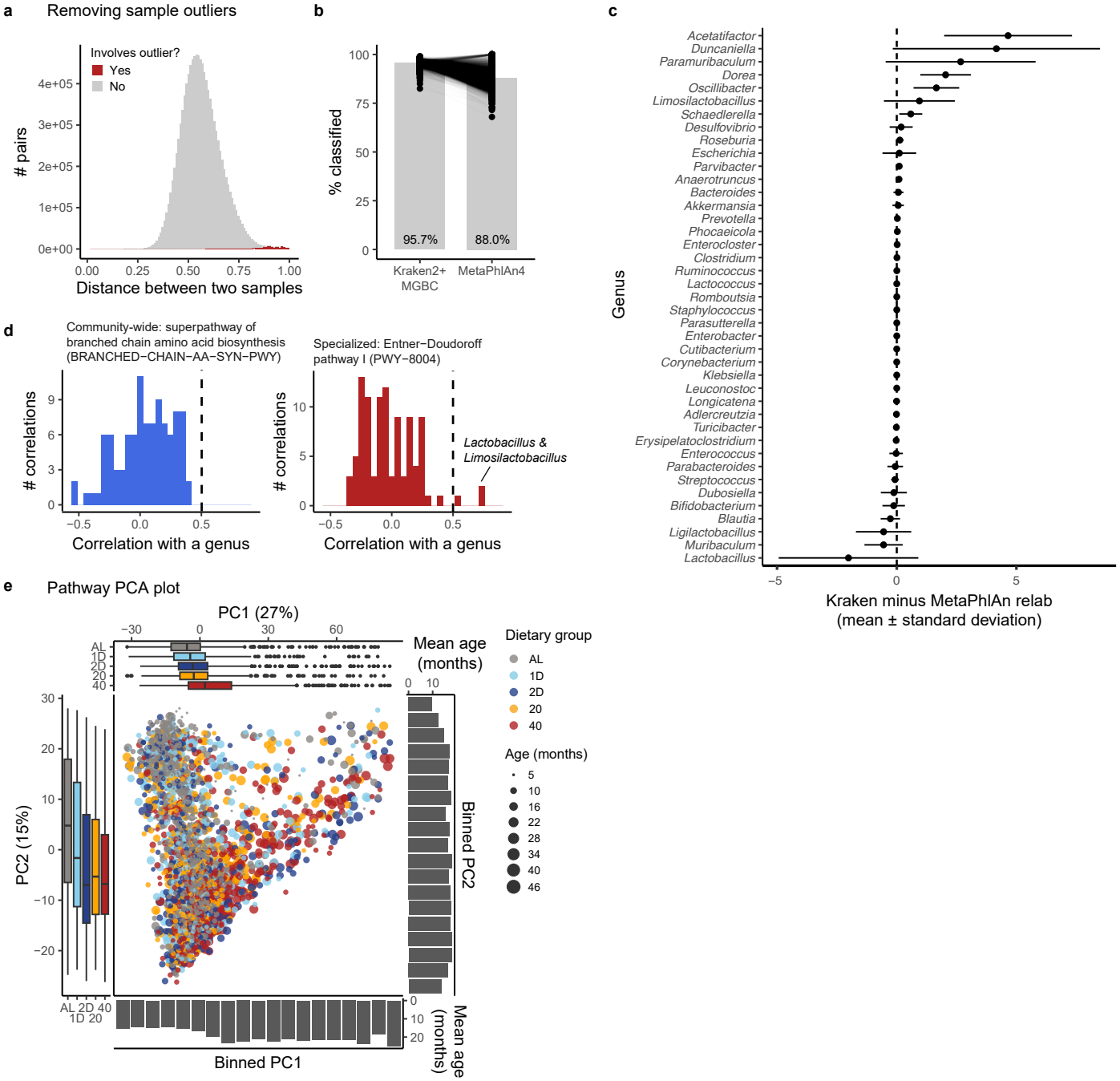


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41564-025-01963-3

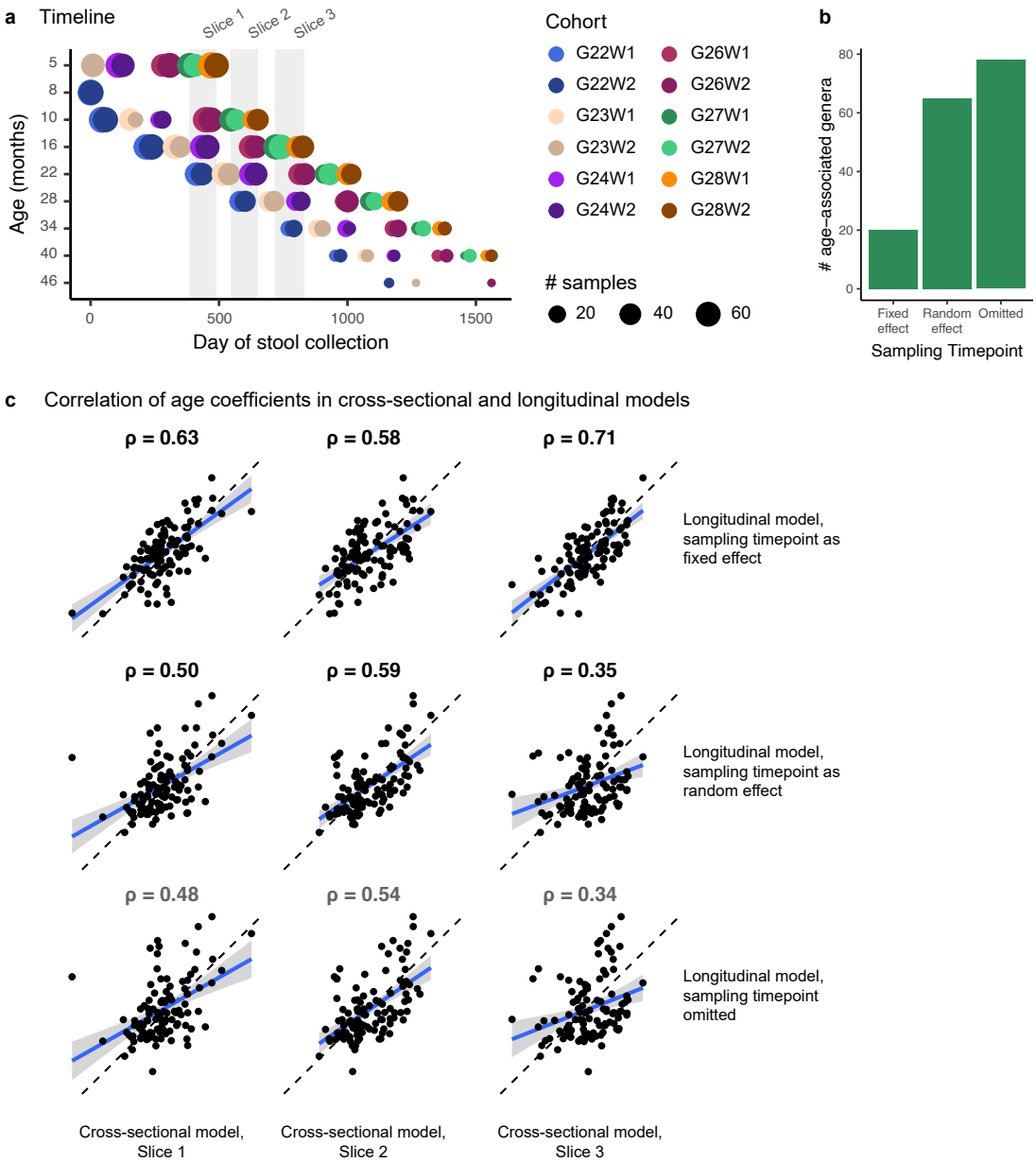
Extended Data Fig. 2 | Identifying sample mix-ups. a, Sample mix-ups were identified by comparing host reads from each microbiome sample against all host genotypes (we term this pipeline "mbmix"). For more details, see Supplementary Note 1. **b**, Example of concordance between a microbiome sample and the expected host genotype. The x-axis is each host genotype, the y-axis is the proportion of single nucleotide polymorphisms (SNPs) that were discordant between the microbiome sample and the host genome. **c**, Example of discordance. Microbiome sample DO_20_1188_021w was supposed to originate from mouse DO-20-1188, but it appears to have come from DO-2D-4188. **d**, Best proportion discordant SNPs versus proportion of discordant SNPs against the expected genotype. The fate of each sample is indicated by its color: kept (green), discarded (red), or renamed (blue). **e**, Plate view of mbmix categorization. Each sub-panel is a "final plate", a 96-well plate of libraries prior to pooling. White regions either didn't contain a sample, contained a sample that obtained no reads (for example, left half of final plate 31), contained a sample whose mouse did not have a genotype, or contained a control sample. **f**, Proportion discordant SNPs for stool samples from mice DO-AL-0097 and DO-AL-0105. Samples from 44 weeks were concordant with the expected mouse genotype. All other samples from mouse DO-AL-0097 appeared to come from mouse DO-AL-0105, except DO_AL_0097_148w, which was inconclusive. The two other samples from mouse DO-AL-0105 appeared to come from DO-AL-0097. For discordant results, the mouse with the lowest proportion of discordant SNPs is colored red. **g**, Body weight for mice DO-AL-0097 and DO-AL-0105. The vertical dashed line at 56 weeks represents the likely time that these mice were swapped in their cages.

https://doi.org/10.1038/s41564-025-01963-3



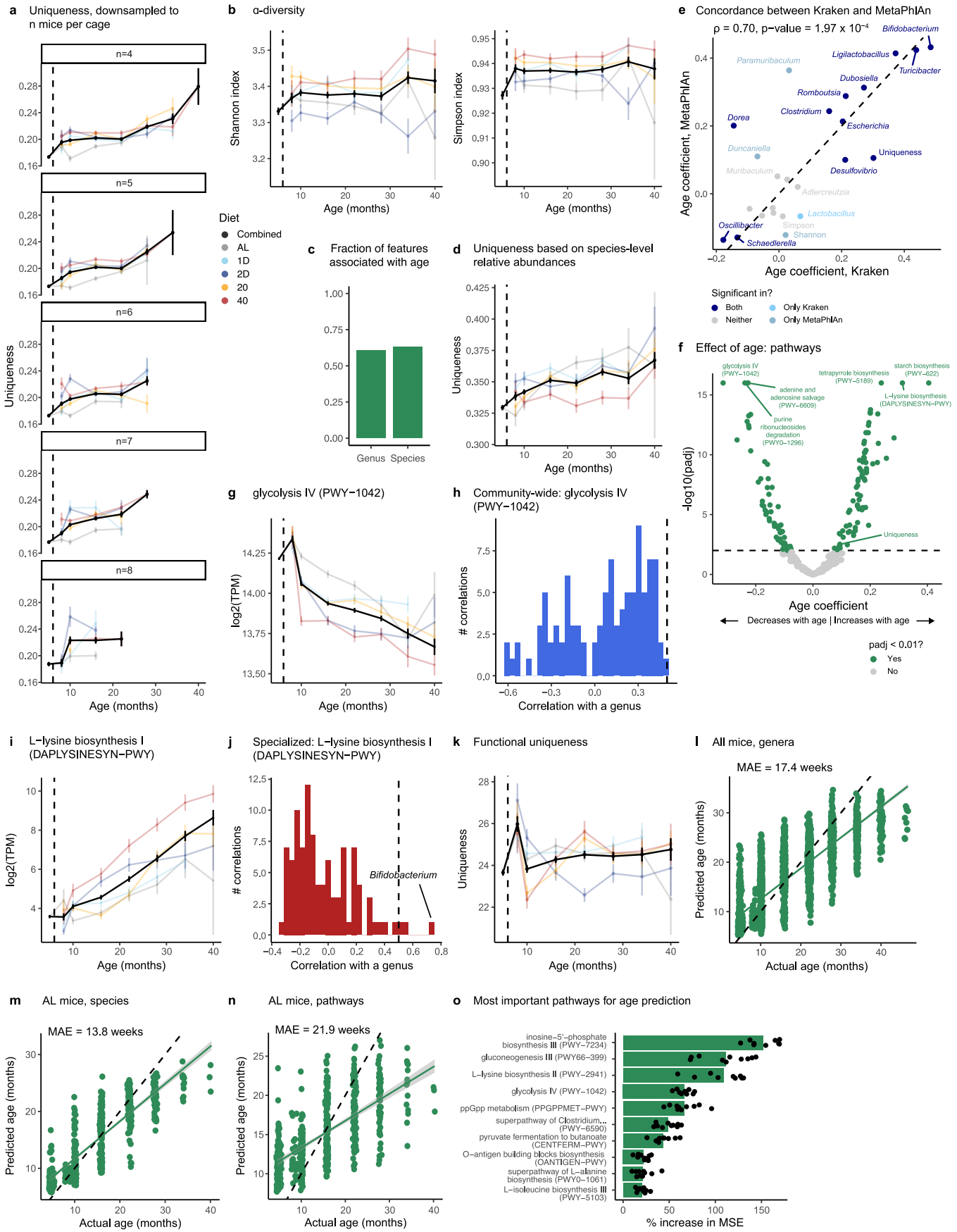
Extended Data Fig. 3 | Quality-control and details related to taxonomic and functional classification. a, Histogram of all pairwise sample distances (Bray-Curtis on relative abundances). Distances involving any of 13 outlier samples are shown in red. **b**, Percentage of reads that could be classified for non-control samples using either Kraken2+MGBC or MetaPhlAn4. Mean percent classified indicated in black text. **c**, Difference between Kraken2+MGBC and MetaPhlAn4 genus-level relative abundances for the 41 genera present in both taxonomic databases. Each horizontal line shows the mean ± standard deviation across all 2997 non-control samples. **d**, Examples of community-wide and specialized pathways. The largest correlations for the specialized pathway (PWY-8004) were with *Lactobacillus* and *Limosilactobacillus*. **e**, PCA plot based on microbial pathways (n = 2997 metagenomes). PC1 and PC2 explain 21% and 8% of overall variance, respectively. For the boxplots, boxes extend from the 25th to 75th percentiles, whiskers extend to 1.5 times the interquartile range, and the center line is the median. For more details, see Fig. 1 legend.

https://doi.org/10.1038/s41564-025-01963-3



Extended Data Fig. 4 | Effect of sampling timepoint. a, Timeline of stool collection. X-axis shows the day of stool collection, with the first day of the overall experiment as day 1. Y-axis indicates the age of a mouse when a stool sample was collected. The color of each circle corresponds to one of 12 DO breeding cohorts that were sequentially entered into the study. The size of each circle corresponds the number of samples collected for each cohort-age combination. Gray rectangles correspond to three cross-sectional data slices used in later analyses. **b**, Number of genera associated (conditional Wald test, Benjamini-Hochberg adjusted p-value < 0.01) with age using linear mixed models that included sampling timepoint as a fixed effect (Model 2), random effect (Model 1), or not at all (Model 3). **c**, Correlations between age coefficients calculated using cross-sectional (columns) and longitudinal (rows) models. Blue line represents the line of best fit and 95% confidence interval (linear regression). Spearman correlation (ρ) indicated above each scatterplot. Black dashed line at y = x represents perfect agreement between two models.

https://doi.org/10.1038/s41564-025-01963-3

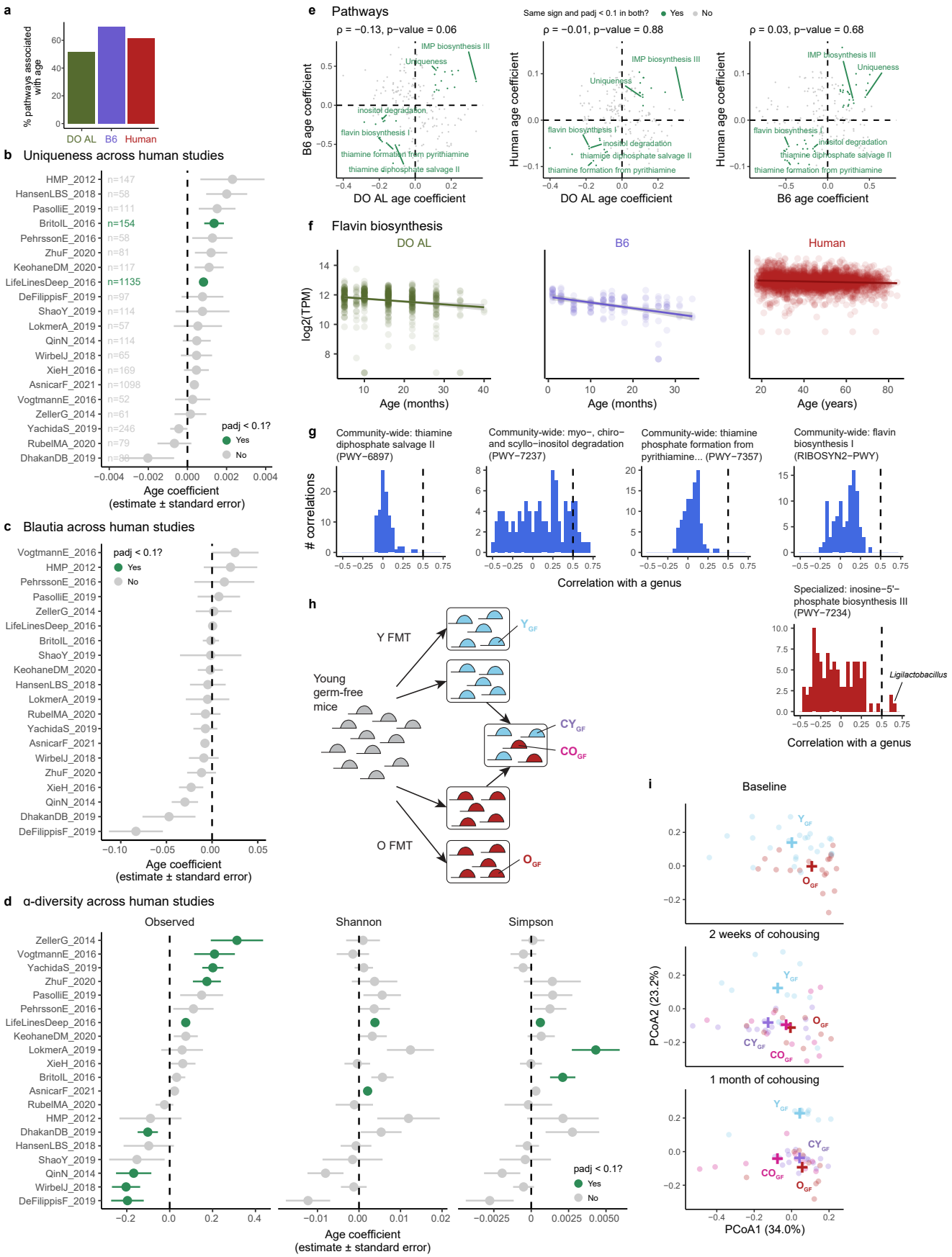


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41564-025-01963-3

Extended Data Fig. 5 | Additional details related to age-associated microbiome changes. a, Uniqueness increases with age even when the number of mice per cage is kept fixed. For various n, cages with at least n mice at that age were considered. If the number of mice was greater than n, then n mice were randomly chosen. Uniqueness was then recomputed on this subset of samples. **b**, α-diversity (as measured by Shannon and Simpson indexes) appears to increase with host age (n = 2988 metagenomes with age ≤ 40 months), but this trend is not significant (Model 1, conditional Wald test, Benjamini-Hochberg adjusted p-value > 0.01). **c**, Fraction of features associated (Model 1, conditional Wald test, Benjamini-Hochberg adjusted p-value < 0.01) with age when using genus-level or species-level data. **d**, Uniqueness increases with age when using species-level relative abundances (n = 2988 metagenomes). **e**, Comparison of age coefficients calculated using Kraken2 or MetaPhlAn taxonomic results. Diagonal dashed line at y = x represents perfect agreement. Spearman correlation (ρ) and p-value are indicated above the plot. **f**, Effect of age on microbial pathways. Age coefficients and standard errors were calculated with Model 1. p-values were calculated with a conditional Wald test and adjusted with the Benjamini-Hochberg procedure. **g**, Glycolysis IV (PWY − 1042) decreases with age (n = 2988 metagenomes). **h**, Correlations between PWY − 1042 and all genera. PWY − 1042 is a community-wide pathway because it has no correlations above 0.5. **i**, L-lysine biosynthesis II (DAPLYSINESYN-PWY) increases with age (n = 2988 metagenomes). **j**, DAPLYSINESYN-PWY is a specialized pathway because it is highly correlated with *Bifidobacterium*. **k**, Functional uniqueness increases with age (n = 2988 metagenomes). **l-n**, Age prediction based on 10-fold cross validation. Green line represents the line of best fit and 95% confidence intervals (linear regression). Black dashed line at y = x represents perfect accuracy. MAE = mean absolute error. **l**, Prediction considering all mice, rather than just AL mice. **m**, Prediction using species-level relative abundances in AL mice. **n**, Prediction using pathway log2(TPM) abundances in AL mice. **o**, Top 10 most important pathways for age prediction (just AL mice, n = 573 metagenomes). Each dot is one of 10 cross-validation folds. X-axis shows the percent increase in mean squared error (MSE) when that particular pathway is excluded from a tree within the random forest regressor. In **a**, **b**, **d**, **g**, **i**, and **k**, data are presented as mean ± SEM.

https://doi.org/10.1038/s41564-025-01963-3

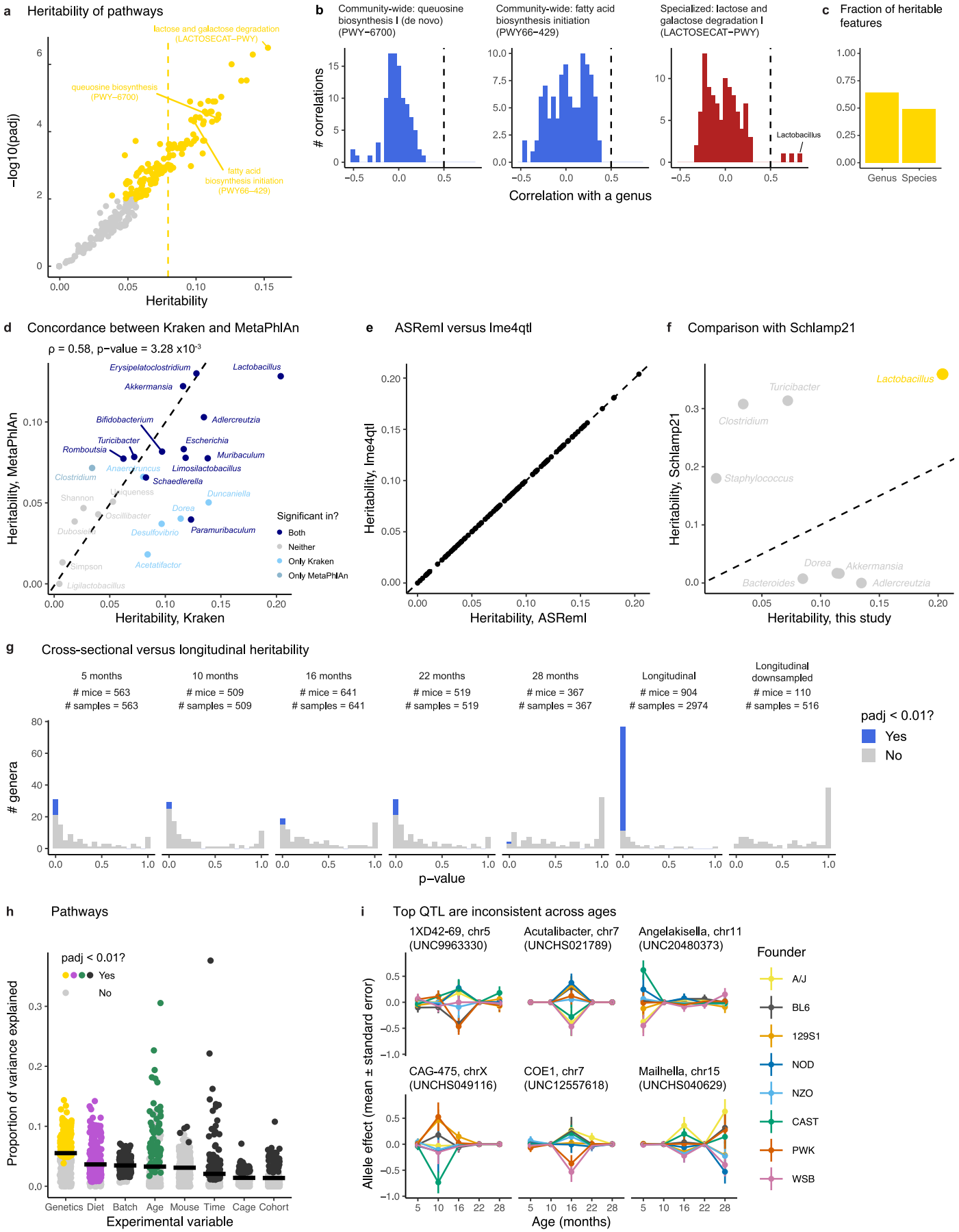


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41564-025-01963-3

Extended Data Fig. 6 | Additional details related to universal age-associated microbiome changes. a, Percentage of pathways associated with age (Models 5-7, Benjamini-Hochberg adjusted p-value < 0.1) within each dataset. **b-d**, Associations with age within human studies. Coefficients, standard errors, and p-values were calculated with Model 8, and p-values were adjusted with the Benjamini-Hochberg procedure. Adjusted p-values < 0.1 are shown in green. The number of individuals per study indicated in **b** is the same as in **c** and **d**. **b**, Uniqueness tends to increase with age in most human studies, including the largest (LifeLinesDeep_2016). **c**, *Blautia* appears to increase with age in some studies and decrease with age in others, and when regressing against age separately per study, no studies have an adjusted p-value < 0.1. **d**, α-diversity versus age, separately per human study. p-values were adjusted separately per metric. **e**, Comparison of age-associated functional changes across datasets. Each pairwise comparison shows all features that passed prevalence filtration in both datasets. Line of best fit and 95% confidence interval shown in gray. Spearman correlation and corresponding p-value shown above each plot. Features associated with age and with the same sign in the pairwise comparison are shown in green. **f**, Flavin biosynthesis I (RIBOSYN2-PWY) decreases with age in all three datasets. Each panel includes the line of best fit and 95% confidence interval (linear regression). **g**, Histograms of pathway-genus correlations. For the specialized pathway (PWY-7234), the largest genus correlation is to *Ligilactobacillus*. **h**, Schematic of cohousing experiment in germ-free mice. Young germ-free mice (gray) received fecal microbiome transplants (FMTs) from young donors (Y FMT) or old donors (O FMT). Mice that received Y FMT are shown in blue, mice that received O FMT are shown in red. Y_{GF} = Y FMT recipients housed with other Y FMT recipients, O_{GF} = O FMT recipients housed with other O FMT recipients, CY_{GF} = Y FMT recipients cohoused with O FMT recipients, CO_{GF} = O FMT recipients cohoused with Y FMT recipients. **i**, PCoA (based on genus-level Bray-Curtis distances) of samples at baseline, after two weeks of cohousing, and after one month of cohousing. Ordination based on all samples shown in this plot. + denotes group centroid. In **b**, **c**, and **d**, data are presented as mean ± SEM.

https://doi.org/10.1038/s41564-025-01963-3

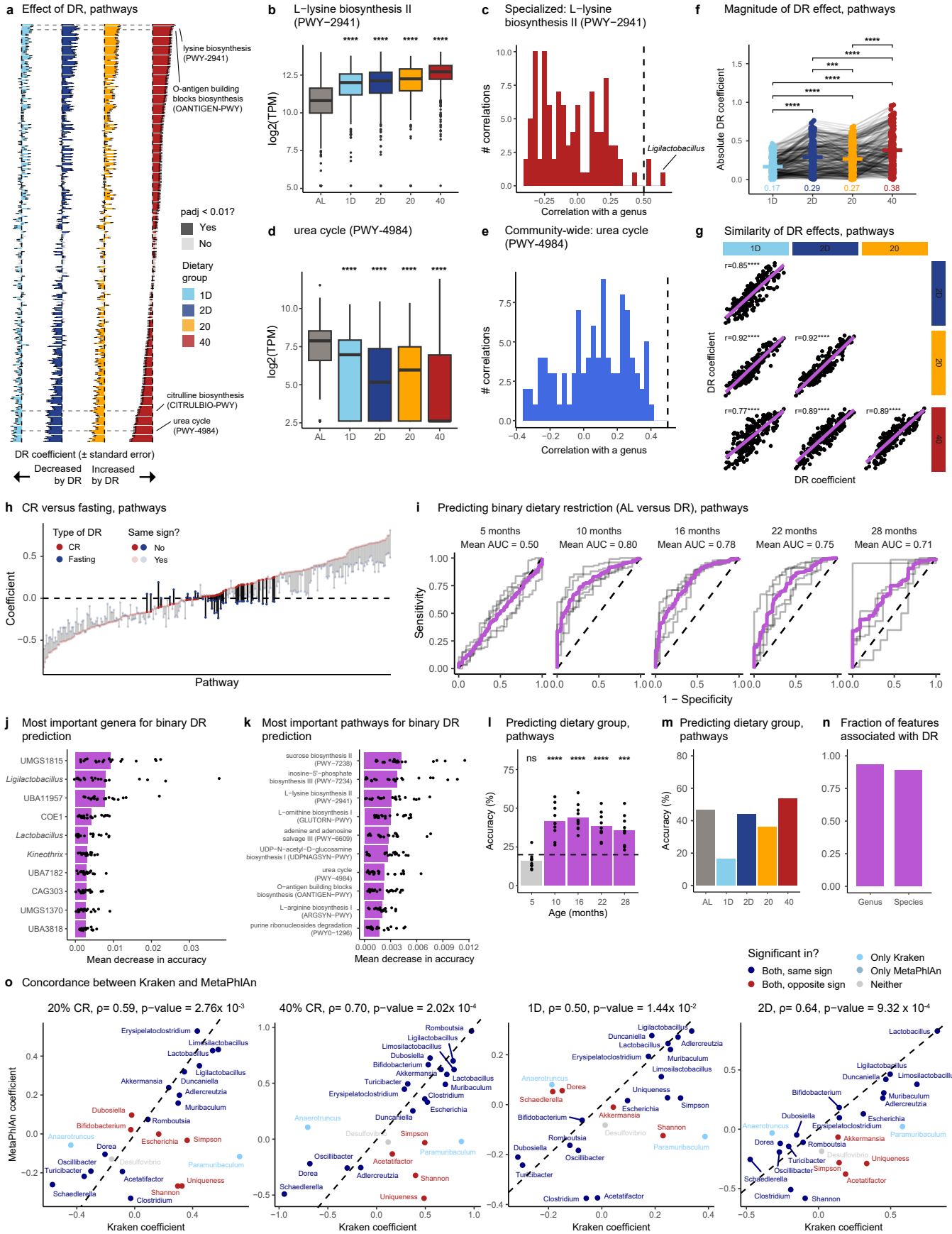


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41564-025-01963-3

Extended Data Fig. 7 | Additional details related to microbiome heritability. **a**, Heritability of pathways. Heritability was calculated with Model 1. p-values were calculated using a likelihood ratio test and adjusted with the Benjamini-Hochberg procedure. Yellow vertical dashed line shows mean heritability for heritable features. **b**, Histograms of pathway-genus correlations. For the specialized pathway (LACTOSECAT-PWY), the largest genus correlation is to *Lactobacillus*. **c**, Fraction of heritable (Model 1, likelihood ratio test, Benjamini-Hochberg adjusted p-value < 0.01) features when using genus-level or species-level data. **d**, Comparison of heritability calculated using Kraken2 or MetaPhlAn taxonomic results. Diagonal dashed line at y = x represents perfect agreement. Spearman correlation (ρ) and p-value are indicated above the plot. **e**, Heritability computed with lme4qtl or ASReml using the same model and data. **f**, Comparison of heritability estimates from a different DO mouse study (Schlamp et al. 2021). Plot shows the eight genera for which heritability was assessed in both datasets. Of these eight, the most heritable taxon in both studies was *Lactobacillus* (highlighted in yellow). **g**, Cross-sectional (Model 9) versus longitudinal (Model 1) versus downsampled longitudinal (Model 9, downsampled to 110 mice) heritability. Heritable genera (Benjamini-Hochberg adjusted p-value < 0.01) are shown in blue. The longitudinal results are the primary heritability results presented throughout the manuscript. **h**, Proportion of variance explained (PVE) by all experimental variables for n = 273 pathways (Model 10, p-values calculated with likelihood ratio test, adjusted with Benjamini-Hochberg). Horizontal lines show the mean PVE. **i**, Allele effects across ages for the top six age-specific QTL (permutation test, adjusted p-value < 0.01). QTL mapping was performed using n = 569, 513, 646, 522, and 368 samples respectively at 5, 10, 16, 22, and 28 months. Data are presented as mean ± SEM. The title above each sub-panel indicates the genus, chromosome, and genotyping marker for the QTL result. Color of each line represents the allele effect for each of eight founders comprising the Diversity Outbred genetic pool.

https://doi.org/10.1038/s41564-025-01963-3

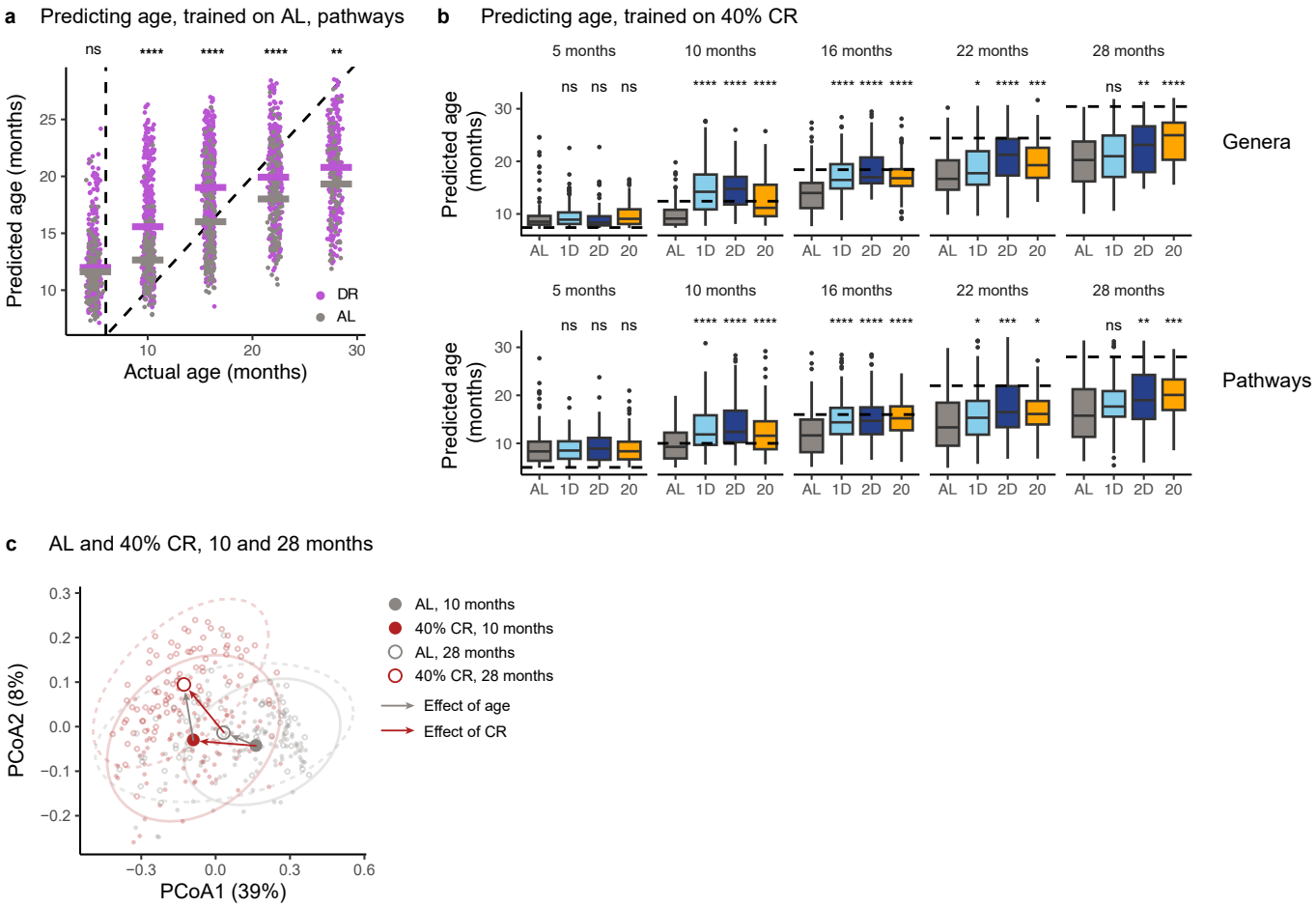


Extended Data Fig. 8 | See next page for caption.

https://doi.org/10.1038/s41564-025-01963-3

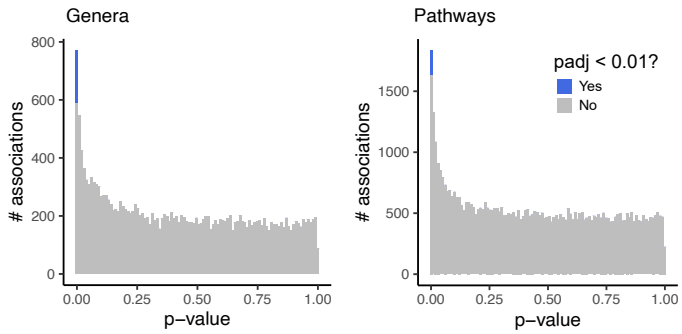
Extended Data Fig. 8 | Additional details related to the effects of dietary restriction. a, Effect of dietary restriction (DR) on 273 pathways. DR coefficients and standard errors were calculated with a linear mixed model (Model 1). p-values were calculated using a conditional Wald test and adjusted with the Benjamini-Hochberg procedure. **b**, The L-lysine biosynthesis II pathway (PWY-2941) is increased by DR (n = 2988 metagenomes with age ≤ 40 months). **c**, PWY-2941 is a specialized pathway, most highly correlated with *Ligilactobacillus*. **d**, The urea cycle pathway (PWY-4984) is decreased by DR (n = 2988 metagenomes). **e**, PWY-4984 is a community-wide pathway with no correlations with genera above 0.5. **f**, Absolute magnitude of DR coefficients for 273 pathways. Gray lines connect the same pathway in different dietary groups. Horizontal bars show the mean. Statistical significance evaluated by a paired t-test. **g**, Comparison of DR coefficients for pathways. Pearson correlation and p-value is indicated above each scatterplot. Lines of best fit and 95% confidence intervals (linear regression) are shown in purple. **h**, Mean CR versus mean fasting coefficients for pathways. Vertical lines highlight the difference in mean CR coefficients (red) versus mean fasting coefficients (blue). Pathways with opposite signs are opaque, while pathways with the same sign are transparent. Dashed horizontal line at 0. **i**, Receiver operating characteristic (ROC) curves for the prediction of binary DR using pathways, separately at each age. Each gray line is the ROC curve for one of 5 cross-validation folds. The purple line is the mean ROC curve. The diagonal dashed line at y = x represents no predictive accuracy. AUC = area under the curve. **j-k**. Top 10 most important genera (**j**) and pathways (**k**) for prediction of binary DR status. Each dot is one of 20 cross-validation folds (4 post-randomization ages x 5 folds per age). X-axis shows the mean decrease in accuracy, that is between trees in the random forest that do include the feature of interest and trees that do not. **l**, Predicting dietary group using pathways before (gray) and after (purple) initiation of DR. Each dot represents prediction accuracy in one of 10 cross-validation folds. Horizontal dashed line at 20% represents expected accuracy by chance. Statistical significance evaluated by a one-sided t-test (testing whether the mean accuracy is greater than 20%). **m**, Prediction accuracy stratified by dietary group using pathways. Only predictions after the start of dietary restriction were considered. **n**, Fraction of features associated (Model 1, conditional Wald test, Benjamini-Hochberg adjusted p-value < 0.01) with DR when using genus-level or species-level data. **o**, Comparison of DR coefficients calculated using Kraken2 or MetaPhlAn taxonomic results. Diagonal dashed line at y = x represents perfect agreement. Spearman correlation (ρ) and p-value are indicated above the plot. For the boxplots in **b** and **d**, boxes extend from the 25th to 75th percentiles, whiskers extend to 1.5 times the interquartile range, and the center line is the median. In **b**, **d**, **f**, **g**, and **l**, p-value symbols are defined as follows: ns: $p \geq 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

https://doi.org/10.1038/s41564-025-01963-3



Extended Data Fig. 9 | Dietary restriction does not rejuvenate the microbiome. **a**, Age prediction with a random forest regressor trained on pathway data from n = 573 AL samples. Vertical dashed line at six months represents start of dietary restriction, diagonal dashed line represents perfect prediction. Statistical significance evaluated by a t-test between AL (gray) and DR (purple) predictions at each age. **b**, Age prediction of a regressor trained on n = 623 40% CR samples and evaluated on all other samples (n = 2374), using genera (top) or pathways (bottom). Horizontal dashed line shows the actual age of samples collected at that timepoint. Boxes extend from the 25th to 75th percentiles, whiskers extend to 1.5 times the interquartile range, and the center line is the median. Statistical significance evaluated with a t-test against the AL group. **c**, PCoA of AL and 40% CR samples from middle-aged (10 months) and old (28 months) samples. Ordination based on just these samples. Group centroids are depicted by the four large points, along with 95% data ellipses. Arrows connect group centroids to depict the effect of age (gray) and the effect of caloric restriction (red). PCoA1 and PCoA2 explain 39% and 8% of overall variance, respectively. In **a** and **b**, p-value symbols are defined as follows: ns: $p \geq 0.05$, *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$, ****: $p < 0.0001$.

https://doi.org/10.1038/s41564-025-01963-3



Extended Data Fig. 10 | Additional details related to microbiome-phenotype associations. Histogram of p-values for associations between phenotypes and genera (left) or pathways (right). Associations were performed using a linear mixed model (Model 11), and p-values were calculated using a likelihood ratio test in which the null model omitted the microbiome term. Associations with a Benjamini-Hochberg adjusted p-value < 0.01 are shown in blue.

nature portfolio

Corresponding author(s): Christoph A. Thaiss

Last updated by author(s): Jan 25, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Fastq files were generated by MiSeq, NextSeq, and NovaSeq sequencers.
Data analysis	Fastq files were processed using Sunbeam (v2.1.1), Kraken2 (v2.1.2), and HUMAnN3. Downstream analysis was performed in RStudio (R v4.2) using the following packages: tidyverse (v2.0.0), phyloseq (1.42.0), vegan (v2.6.4), ASReml (v.4.1.0.716), randomForest (v4.7-1.1), curatedMetagenomicData (v.3.6.2), lme4qtl (v0.2.2), and mediate (v4.5.0). Final figures were created with Adobe Illustrator 2024.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

We generated the following fastq files, which are available on SRA:
• DRiDO metagenomic sequencing: PRJNA105451

• Longitudinal B6 metagenomic sequencing: PRJNA1073968
• 16S sequencing of cohousing experiment: PRJNA1072097
• 16S sequencing of cohousing in germ-free mice experiment: PRJNA1128683

Processed data are available at https://github.com/levlitichev/DRiDO_microbiome. Host phenotypes collected as part of the DRiDO study are available at https://doi.org/10.6084/m9.figshare.24600255.v1. The genetic kinship matrix and genotype probabilities are available at https://doi.org/10.6084/m9.figshare.13190735.

Other databases used in this study include the mm10 genome (http://igenomes.illumina.com.s3-website-us-east-1.amazonaws.com/Mus_musculus/UCSC/mm10/Mus_musculus_UCSC_mm10.tar.gz), MetaPhlan (http://cmprod1.cibio.unitn.it/biobakery4/metaphlan_databases/) and HUMAnN databases (http://huttenhower.sph.harvard.edu/humann_data/chocophlan/full_chocophlan.v201901_v31.tar.gz, http://huttenhower.sph.harvard.edu/humann_data/uniprot/uniref_annotated/uniref90_annotated_v201901b_full.tar.gz), MGBC Kraken2 database (https://github.com/BenBeresfordJones/MGBC), curatedMetagenomicData (DOI: 10.18129/B9.bioc.curatedMetagenomicData), and SILVA 138 (https://www.arb-silva.de/documentation/release-138/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We do not report sex or gender of human participants.
Reporting on race, ethnicity, or other socially relevant groupings	We do not report race, ethnicity, or other socially relevant groupings.
Population characteristics	Our human cohort contains metagenomic data from 4101 humans across 20 different studies in 16 countries. We considered samples from individuals meeting the following criteria: age ≥ 18 years, “healthy” or “control”, and no current antibiotic use.
Recruitment	We used publicly available, de-identified metagenomic data from the curatedMetagenomicData R package.
Ethics oversight	n/a

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the DRiDO study, simulations were used to determine sample size needed to detect 10% change in lifespan based on historical lifespan data from DO mice. For the longitudinal B6 and cohousing experiments, sample sizes were standard for mouse experiments, e.g. 3-5 per experimental condition.
Data exclusions	No data were excluded from analysis.
Replication	No experiments were biologically replicated. Technical replication consisted of preparing different metagenomic libraries from the same DNA and sequencing the same library multiple times. Computational replication included performing taxonomic and functional classification with multiple approaches and fitting linear mixed models with different tools.
Randomization	Diversity Outbred mice were randomized by housing cage at wean (~4 weeks of age). No randomization was performed in the longitudinal B6 experiment. Cohousing and separation was performed by randomization of housing cages.
Blinding	In the DRiDO study, blinding was not possible due to different cages requiring distinct dietary interventions. The dietary interventions also dramatically affected body weight, allowing technicians to tell whether or not a mouse was on dietary restriction. No blinding was performed during the generation of microbiome data due to the high-throughput nature of data generation. In the cohousing experiment, technicians could not be blinded due to young and old mice looking different. In the cohousing experiment in germ-free mice, technicians were blinded to which FMT treatment was received by germ-free mice.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Mus musculus. Diversity Outbred mice from the Jackson Laboratory (J:DO IMSR_JAX:009376) were followed longitudinally starting at 4 weeks of age. For the longitudinal B6 experiment, 4-week-old C57BL/6 mice were ordered from the Jackson Laboratory. For the cohousing experiment, 8-week-old C57BL/6 mice were ordered from the Jackson Laboratory and 19- and 20-month-old C57BL/6 mice were ordered from the National Institutes on Aging. For the cohousing in germ-free mice experiment, we used 3-5-month-old germ-free female C57BL/6 mice bred at the Penn Gnotobiotic Mouse Facility.
Wild animals	No wild animals were used in this study.
Reporting on sex	The DRiDO study used only female mice to prevent aggressive competition for limited food resources. The longitudinal B6 experiment -- in which there was no dietary restriction -- used only male mice. The cohousing experiments used only female mice to avoid potential violence due to rehousing of animals.
Field-collected samples	No field animals were used in this study.
Ethics oversight	The DRiDO study was approved by the IACUC at the Jackson Laboratory (protocol #06005). The longitudinal B6 and cohousing experiments were approved by the IACUC at the Perelman School of Medicine at the University of Pennsylvania (protocol # 806361).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a