Supplementary Methods, Results and Figures to:

Experimental evolution of a mammalian holobiont: stability of the distinct microbiome composition in bank voles selected for herbivorous capability

Małgorzata M. Lipowska, Edyta T. Sadowska, Kevin D. Kohl, Paweł Koteja

Supplementary Methods

Animal model and the ongoing selection experiment

This work was performed on bank voles (*Myodes = Clethrionomys glareolus* Schreber 1780) from generation 27 of an ongoing artificial selection experiment maintained at the Jagiellonian University (Poland). The rationale, history and protocols of the ongoing experiment were presented in our earlier papers (Sadowska et al. 2008, 2015; Lipowska et al. 2020). Briefly, the colony was established with about 320 wild voles captured in 2000 and 2001. After 5-6 generations of random breeding, the selection experiment has been started, with "Herbivorous" (H) lines selected for the ability to maintain body mass during a 4-day trial, during which the young, growing animals are fed a low-quality diet, "diluted" with dried grass powder. Four replicate H lines, and four unselected Control (C) lines are maintained to allow valid tests of the effects of selection (Henderson 1997), with 15–20 reproducing families in each of the 8 lines (to avoid excessive inbreeding). As average litter size in the voles is only about 4.5, up to three subsequent litters from each family are reared to provide enough animals for an effective selection. The animals are weaned at the age of 17 days (at day 18 a next litter can be born) and kept initially in family groups. At the age of 32-36 days the 4-day selection trial is performed on most animals from the selected lines (except individuals assigned to separate experiments) and a sample of individuals from the Control lines. The selection criterion is body mass change during the trial adjusted for body mass at weaning and body mass gain between the weening and the start of the trial (i.e., a residual of regression on the two covariates). The adjustment is made to avoid selecting for just a high or low values of body mass or growth rate. The selection is performed mostly within-families, i.e., from each full-sib family 1-2 males and 1-2 females with the highest scores are chosen for reproduction. However, when more than 17 families are available, the families in which all individuals have below-average scores (residuals lower than zero) are excluded from reproduction. If the best animals from the family fail to reproduce, next ones (if available) are selected.

Over the course of the selection experiment, the composition of the low-quality diet has been modified a few times in attempts to ensure that it poses a challenge, but not an overly severe one (Sadowska et al. 2015). This, together with the fact that the composition of the grass powder changed across time and seasonal changes (despite controlled thermal and light conditions), leads to considerable variation in the selected trait values across generations (Fig. S1). Nevertheless, already since generation 3, the H-line animals have been consistently able to maintain a more positive body mass balance during the trial when compared to the C-line ones, and despite the large among-generation variation, the difference in body mass balance between the selected and control lines was about 1.5-2.0g, corresponding to 1-2 units of phenotypic standard deviation. In generation 25, the last in which the selection was performed, voles from the H lines have gained during the test 1.55±0.97g (mean±SD from pooled observations from the four replicate lines, 7.4% of the initial

body mass), whereas those from C line gained only 0.10±0.89g (0.5% of the initial body mass). In generations 16-17, and in two generations (25-26) preceding the experiment reported here the selection was relaxed, and the regular tests with low-quality diet were not performed.

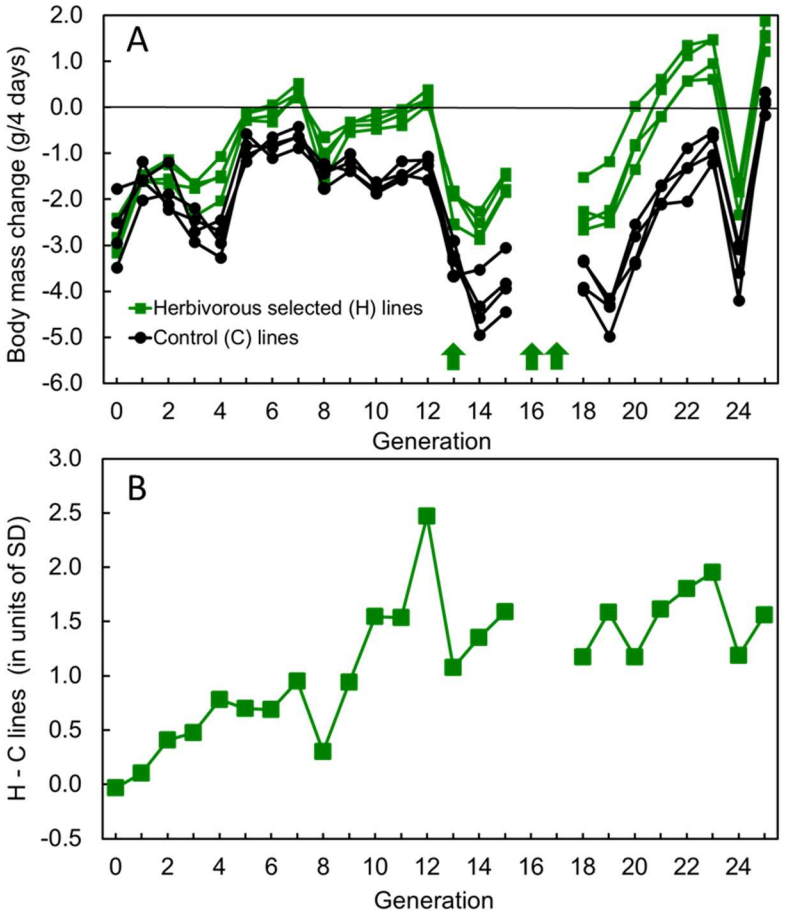

Fig. S1 Direct effects of selection towards an increased ability to maintain body mass in a 4-day test with low-quality herbivorous diet. **A)** Mean values of body mass change (g/4days) in the four replicate H-selected and four replicate Control lines; **B)** the difference between means of the selected and control lines in the units of phenotypic standard deviation; arrows indicate generations in which selection was relaxed.

Reproductive pairs were kept in standard polypropylene mouse cages (model 1290D, Tecniplast, Bugugiatte, Italy) fitted with sawdust bedding, a clay pot "shelter" and nest-building material (paper towel and a cardboard tube), with *ad libitum* access to water and food (a standard rodent chow: 23.9% protein, 4.5% fat, 5.3% fiber, 14.3 kJ/g metabolizable energy in dry mass; Labofeed H, Kcynia, Poland), at constant temperature (20±1°C) and photoperiod (16:8 light:dark; light phase starting at 02:00 hours).

All the breeding, selection and experimental procedures were approved by the Local Ethical Committees in Krakow, Poland (decision no. 170/2014 – 1st Local Ethical Committee for Animal Experiments, Faculty of Pharmacy, Jagiellonian University Medical College in Kraków; 257/2017 – 2nd Local Institutional Animal Care and Use Committee, Institute of Pharmacology Polish Academy of Sciences in Kraków), and in accordance with the EU directive 2010/63/EU. This study is reported in accordance with ARRIVE guidelines.

The cohabitation procedure

We conducted a cohabitation experiment to allow for microbial transfer between alternate lines to test for microbial dispersal and effects on host phenotypes. The experiment was performed on animals from 69 C-line families and 80 H-line families from generation 27 (15-20 families within each replicate line, up to 7 siblings within family) (Fig. 1 in the main text). Parents of these animals were not subjected to the selection test. Because the first litter of a breeding couple usually differs from the further litters, the animals chosen for the experiment were sampled from 2nd and 3rd litters of each family. At the age of 17 days, the pups were weaned, weighed in clean cups, marked temporarily by fur clipping and moved, in groups of up to 5 siblings per cage, to standard housing cages (model 1264C, Tecniplast, Bugugiatte, Italy). The pups were provided sawdust bedding and a nesting material (a paper towel), and *ad libitum* access to food and water.

At the age of 21 or 22 days, experimental animals were separated from their siblings and moved to individually-ventilated cages (AERO Mouse IVC Green Line: Tecniplast, Italy), which prevented microbiome exchange with animals other than the cohabitant, fitted with sawdust bedding *and ad libitum* access to food and water. There, the animals were housed with another, unrelated experimental individual of same sex, forming a cohabitation pair. The pairs were formed with two individuals derived from either different line types (CH, HC – where the first letter denotes the linetype of the focal individual, and the second letter denotes the cohabitant linetype), or two animals from the same line type (CC, HH), but not the same replicate line (Fig. 1). At least one animal within a pair was 22 days old, animals a day younger were used if age-matched cohabitants were not available. If necessary, marking by additional fur clipping was applied to ensure distinction between cohabitants. Whenever possible, same-sex siblings were assigned to different replicate-line combinations. Within each combination of replicate lines, 13-19 pairs were formed (414 pairs = 828 individuals total). The cohabitant pairs were housed together for 10 days, after which the animals were separated into individual cages. In cases where one animal was younger than the other, it was maintained in the cohabitation cage for an additional day.

One pair was incidentally killed during cohabitation. In two more pairs one of the cohabitants had died. In one of these cases bite marks were found on the corpse, indicating that the surviving animal had altered its diet and hence it was excluded from the experiment. In the other case the carcass was found fresh and intact at the day of separation, so the surviving animal was maintained in the experiment. Hence, 823 animals successfully completed cohabitation (see Table S1 for a summary of the number of animals that completed subsequent stages of the experiment).

The feeding trial

Next, we conducted a factorial experiment where animals from each cohabitation combination were fed two different diets. The animals were separated from their cohabitants at the age of 32 days and moved to individual cages where they were subjected to a feeding trial. The animals were assigned to four combinations of two factors: two categories of diet and two categories of cage type. The assignment was randomized, with a restriction that cohabitants were assigned to different cage type categories, and same-sex siblings were distributed possibly evenly across cage and diet type combinations. The two diet groups received during the feeding trial either the standard diet (SD), the same as used in the regular maintenance and breeding (see above), or low-quality diet (LQD), similar to that used in the H-line selection tests, but containing less plant material (pellets made of the

mixture of 60% Labofeed H and 40% powdered dried grass: 20.4% protein, 4.4% fat, 16.1% fiber, 11.4 kJ/g metabolizable energy in dry mass).

The “standard” cage type (SC) was the same as applied during the H-line selection test: standard open-top cages (model 1264C, Tecniplast, Bugugiatte, Italy), fitted with sawdust bedding. The “metabolic” cage type (IVC) were individually-ventilated cages (AERO Mouse IVC Green Line: Tecniplast, Italy), the same as used in the post-weaning period, but fitted with perforated plastic bottoms suspended above the cage floor instead of bedding, which allowed to collect all uneaten food and feces (blotting paper was placed at the floor to absorb urine). The two types of cages were applied because reliable estimates of food consumption and digestibility require using metabolic cages, but on the other hand the lack of bedding in such cages is stressful, and therefore both the pattern of body mass changes and the microbiome composition could differ from those in voles maintained in standard cages with bedding.

The animals were habituated to the cages for four days. During this period, they were provided *ad libitum* the standard food in the feeder, but also offered a small pellet of the experimental diet (either SD or LQD, depending on the diet group assignation) on the cage bottom. The LQD was introduced in the habituation phase to minimize the effect of novelty at the onset of the proper trial.

After the 4-day habituation, at the age of 36 days, the 5-day feeding trial was started (day 0; Fig. 1). For technical reasons, all the procedures on a given day were performed first on animals kept in the metabolic IVC cages (ca. 7:30 - 11:30 hours) and later on those in the standard cages (ca 9:30 - 13:30 hours). To minimize the effect of the measurement timing on the estimates of body mass changes and food consumption, in the subsequent stages of the feeding trial the procedures were performed on animals ordered in the same way. At day 0, the animals were weighed in clean cups, moved to fresh cages (of the same type), and were given either SD or LQD. In the standard cages, the food was provided in excess to the overhead feeder, as in the routine breeding or standard selection tests. In the metabolic cages, a pre-weighed portion of ca 12g food (weighted to the nearest 0.001g) was served on the cage bottom (to obtain reliable estimates of the food consumption the amount of food provided must be small, and with the small amount of food pellets some animals had problems with eating it from the feeders – 28 animals which originally received food to the feeders were excluded from further analyses). At the same time weighted samples of the food were taken for measuring dry mass content. At days 1 and 3 the animals were weighed in clean cups and either returned to the same cage (standard cages group) or moved to fresh cages with a pre-weighed, 23-g portion of food (metabolic cages). Although the food portions were designed to be more than sufficient for 2-day periods (days 1-3 and 3-5), some animals tended to grind it and the orts fell below the perforated floors. Therefore, the amount of food available was inspected every day, and was restocked if needed. Nevertheless, some animals had ground all available food to orts and temporarily lost access to food as it fell through the floor, and those 22 animals were excluded from further analyses.

Uneaten food and feces collected from the metabolic cages were pre-dried, sorted, dried (two days at +60°C in vacuum drier) and weighed (to the nearest 0.001g), alongside with the samples of food taken for dry-mass content estimation. The rate of food consumption (FC, g/day) was calculated for days 1-5 as the difference between the dry mass of food provided and dry mass remaining in the cage, averaged over the 4 days. For the same four days, the rate of food digestion (FD g/day) was

calculated as a difference between the food consumption and feces production, and apparent digestive efficiency (ADE, %) was calculated as the FD/FC ratio. The first day of the trial (days 0-1) was not included in these calculations because the feces excreted for at least several initial hours were remains of the pre-trial food, and therefore the estimations of digestibility would be meaningless.

At day 5 the animals were moved to a separate room, weighed, euthanized with isoflurane (Aerrane, Baxter, USA) and dissected using flame-sterilized tools. The caecum was extracted, cut just before the entrance to ileum, and its contents were transferred to a clean Eppendorf tube. The tubes were immediately put on dry ice and stored in -80˚C within 2.5 hours.

During the feeding trial, 20 animals died, 5 showed signs of poor health (sudden mass changes, unhealthy look, cage wetting and smell distinct to early symptoms of diabetes) and one got accidently exposed to external microbial sources (non-sterilized laboratory equipment). Additionally, as was mentioned earlier, 50 animals had difficulties in accessing food in the IVC cages: 28 animals had their food provided in a feeder, where it was less available than on the cage floor, and 22 animals tended to fragment all available food into orts which fell through the perforated floor, and might have experience brief fasting before the daily cage checking. Overall, samples were collected from 747 individuals, and microbial DNA was successfully extracted from 745 samples.

Microbial DNA analyses

Microbial DNA was extracted with DNeasy Power Soil Pro kit (Qiagen, Germany), according to the producer recommendation. The tubes containing caecal contents were moved onto dry ice, and partially thawed on wet ice just before the DNA extraction. The contents of the tube were mixed with a flame-sterilized spatula, and a subsample of approximately 150mg was taken for the extraction. The extracts were further processed with a procedure targeting the V4 region of the 16S ribosomal RNA gene, compatible with the Earth Microbiome Project and utilizing a two-step PCR library preparation protocol (Method for in: (Glenn et al. 2019; Marquina et al. 2021)).

In the first PCR, the target region was amplified using custom 515F and 806R primers with variable-length inserts and Illumina adapter tails. The products were purified on home-made SPRI (sold phase reversible immobilization) magnetic beads and indexed in a second PCR reaction using a custom set of 192 forward and 192 reverse indexing primers (Iwaszkiewicz-Eggebrecht et al. 2023). Each sample was indexed with a unique combination of two primers, and each of the primers was used in no more than 8 combinations. The indexed amplicons were pooled and sequenced by Novogene (UK) using the Illumina Novaseq PE250 technology. Approximately 50,000 raw read pairs per sample were obtained.

The sequences were processed using the Qiime2 bioinformatic package (Bolyen et al. 2019; Marizzoni et al. 2020; Prodan et al. 2020). The primers were trimmed off with the *cutadapt* tool, which also filtered out the sequences in which the error rate within the primer region exceeded 10% (2 bases per primer), or which were shorter than 200b after trimming. The sequence pairs were assembled using the *PEAR* tool (Zhang et al. 2014), with a minimum overlap of 15b, minimum quality threshold of 30, and maximum assembly length of 300b. The assembled reads were clustered into sequence variants with the *deblur denoise-16S* tool, based on reads with length limited to 252b. The amplicon sequence variants (ASVs) were aligned and used to construct phylogenetic trees using the

phylogeny align-to-tree-mafft-fasttree function. The taxonomic information of the sequence variants was obtained with the feature-classifier *clarify-consensus-vsearch* tool and the *SILVA 138* database (Quast et al. 2013). The sequences derived from mitochondria, chloroplasts and archaea were excluded from the list with the *feature-table filter-features* function; the same function was also used to remove sequences found in only one sample. The feature table was rarefied to 13,874 sequences per sample with the *feature-table rarefy* function, to match the number of reads in the second-to-last sample in terms of sequence richness. Twenty of such rarefied tables were generated for further bootstrap analyses.

Statistical analyses

The alpha- and beta-diversity measures of the microbiome composition were obtained with tools available within the Qiime2 package. The *diversity alpha* tool was used on each of the rarefied tables to obtain three alpha-diversity metrics in 745 individuals: number of observed ASVs (N_{ASV}), Shannon diversity index and Pielou evenness index. The values were averaged across the twenty repetitions to obtain the bootstrapped alpha-diversity values for each sample. Similarly, weighted and unweighted UniFrac distance matrices were obtained for each of the rarefied tables with the *diversity beta-phylogenetic* tool, and the matrices were then averaged to form a single bootstrapped matrix for each metric. A PCoA analysis was performed on the matrices with the *diversity pcoa* function.

Based on these initial results, we noticed that a subset of 55 animals (7.4%) were characterized by strikingly low microbiome diversity, and formed a separate cluster both in the heatmap and the beta-diversity (PCoA axes) plots (supplementary Results, Fig. S2, below). The subset could be nearly perfectly distinguished by a single criterion: the presence of bacteria from an undescribed genus from *Clostridium innocuum* group (*Ci*), which did not appear in any other individuals. The *Ci*-present animals were distributed nearly evenly across all the experimental groups. As a consequence, all the quantitative traits describing microbiome were plagued by an extreme non-normality of the within-groups distribution, which precluded any meaningful statistical tests concerning the experimental factors. In addition, the Ci-present voles had also a lower body mass and lower food digestibility (supplementary Results). Therefore, because those 55 outlying individuals would distort the analyses of both the microbial and the physiological performance traits, we removed them from further investigation, leaving 690 individuals for the proper statistical analyses.

The statistical analyses included three main parts: a) univariate analyses of the physiological traits measured in the feeding trials and the bacterial alpha diversity indices, b) multivariate and univariate analyses of the bacterial community composition and structure and abundances of particular phyla and genera, and c) multivariate and univariate analyses of correlations between the physiological and microbial traits.

The effects of diet and the origin of the individual (the genetic component of the effect of selection) and of the cohabitant (environmental effect associated with the distinct selection lines) on body mass, performance traits in the feeding trial, and alpha-diversity characteristics of the caecal microbiome at the end of the trial, were performed with cross-nested mixed ANCOVA models, using Mixed procedure of SAS (v. 9.4, (SAS Institute Inc. 2011)), with REML method of estimation and variance components restricted to positive values. All the models included the selection direction (linetype) of the individual and its cohabitant (H *vs* C lines), diet (SD vs LQD) and sex as the main

fixed factors, interactions between these main factors, and respective random effects of replicate line of both the individual and the cohabitant (nested within respective selection groups), and random interaction of the lines with diet and sex. This basic model structure was further expanded to accommodate additional factors and covariates adequate for specific analyses.

The response variables were body mass change during the feeding trial (MD_{FT}; g/5 days), food consumption (FC; g/day) and digestion rate (FD; g/day), apparent digestive efficiency (ADE; %), and three alpha diversity metrics: number of ASVs (N_{ASV}), Shannon diversity index, and Pielou evenness index. Except of FC, FD and ADE, which were measured only in IVC cages, analyses of the other traits were performed both separately for each of the cage types, and for all individuals in one model, which included the cage type (SC *vs* IVC) as cofactor. All the analyses were performed in two versions: for all individuals with *Ci* presence as an additional cofactor, and separately for the main, *Ci*-free group.

Each of the above models included initially all first-order interactions among all the main fixed categorical factors, and, if applicable, also the second order interaction between the effects of individual's and cohabitant's origin and diet, and corresponding random interaction terms (interactions between replicate lines and the respective fixed factors). Then the models were step-wise reduced by removing non-significant interactions. However, interactions between the three focal factors, the origin of individual and it's cohabitant and diet, were always retained in the final models.

The set of random effects included in the above models reflected the actual structure of the experimental design (with two levels of random nested effects and numerous interactions at the level of replicate lines), and corresponded in a minimalistic way to the set of fixed effects in the model. We realize that, despite the large sample size, the number of these random effects was too large to be effectively estimated. However, as it was not possible to determine *a priori* which subset of random effects would be estimable for a particular dependent variable (and the sets turned out to be different for different variables), we decided to keep the excessive set and let the SAS Mixed procedure find the best solution. In all the models the majority of random effects were fixed to zero and only a few positive variance components were estimated. Therefore, the models effectively provided the same solution for the fixed effects as would be obtained in models not including the excessive random effects. Because we used the Satterthwaite's approximation of degrees of freedom (df), the excessive, fixed-to-zero random effects did not affect results of ANOVA F and t tests, either. Note, that with Satterthwaite's approximation the effective dfs are computed from a combination of the dfs of respective random grouping effects and residual term, weighted by variance contribution of the terms (SAS Institute Inc. 2011), and therefore the dfs can take non-integer values.

Several analyses revealed outlying individuals (absolute value of studentized residual ≥ 4.0). These individuals were excluded from analyses of one or more traits, but were retained in analyses of other traits, in which their residuals did not stand out. There was one such individual for body mass at the start of the trial, three for MD_{FT}, five for ADE (which were also excluded from analyses of FC and FD) and six for the alpha-diversity metrics. The exclusion of these individuals from respective analyses improved the normality of residual distribution and the model's goodness of fit (judged by the models' AIC values).

Complete tables with group composition, descriptive statistics, results of the linear mixed models (significance of all the effects and adjusted least squares means with confidence intervals, LSM±95%CI) are provided in Supplementary Tables. The adjusted least squares means were computed based on the final models for mean values of the covariates (the same for all analyses: litter size = 5.5; body mass at the onset of the feeding trial = 21.96g).

To analyze the effects of the focal factors (the origin of the biological and foster mothers and diet) on the multivariate beta-diversity characteristic of the microbial community we used permutational multivariate analysis of variance (PERMANOVA, with 9999 permutations) implemented in R (v4.3.0) *vegan* package (v2.6-4; (Anderson 2017; Oksanen et al. 2022)). The analyses were performed for both the unweighted UniFrac distance matrix (describing the community membership) and the weighted UniFrac distance matrix (describing the community structure). The models included also sex and cage type as additional main effects, as well as the covariates present in the univariate ANCOVA models described above. Initial models included all first-order interactions and the second-order interaction between the effects of individual's and cohabitant's origin and diet type. Then the models were step-wise reduced in the same way as the univariate models presented above. As the analyses showed significant interactions between the three focal factors, in the next steps the analyses were performed separately for the diet and individual's or cohabitant's origin subgroups. Although *adonis2* PERMANOVA can handle random effects (Anderson 2017; Oksanen et al. 2022), it cannot cope with unbalanced nested designs. Therefore, in these analyses the random effects of replicate lines were not included (c.f. (McNamara et al. 2021; Hanhimäki et al. 2022)).

To get an insight in what taxonomic groups contributed to the differences in the microbiome beta diversity between the experimental groups, we used the *adonis2* PERMANOVA also to perform univariate analyses of the relative abundances of 11 phyla (we omitted Fusobacteriota, which were present practically only in the *Ci*-present group) and 115 genera that were preset in at least 10% individuals. We used this approach because the distributions of the abundances were non-normal (and for many taxa severely zero-inflated), and hence the regular linear model could not be used. In these analyses the dependent variable was the abundance of a particular taxon, and the structure of the predictor variables was such as used in the multivariate model presented above. The analysis was performed for the Euclidean distance matrix, and therefore the analysis was equivalent to PERMANOVA on Bray-Curtis dissimilarity for relative abundances of the focal taxon and "all other taxa combined" (summing up to 1), or PERMANOVA on the weighted UniFrac distance for such two operational "taxa" with the sum of the length of phylogenetic branches set to 1. On the other hand, such an analysis is also equivalent to the classical univariate linear model for the abundance of the focal taxon, except that the reported ANOVA F test of significance uses Monte Carlo F distributions (generated with permutations) instead of the theoretical F distribution based on normality assumption. P-values obtained in these analyses were corrected using False Discovery Rate correction for multiple comparisons (Benjamini and Hochberg 1995) "BH" option in R function *p.adjust*).

The above analyses based on relative abundance of particular taxa are conceptually compatible with the multivariate analyses based on the distances (computed from the matrix of relative abundances of ASVs), but the drawback of the approach is that the tests for particular taxa are not independent (an increased abundance of a taxon implies decreased abundance of others). Therefore, we have applied also the ANCOMBC (*Analysis of Compositions of Microbiome with Bias Correction*; *ancombc2*

function in R package ANCOMBC, v. 2.4.0; (Lin and Peddada 2020a, 2020b), b), to compare the bias corrected “absolute” abundances. The method corrects for the bias resulting from differences in sampling fractions among individuals, and fits log-linear models to the corrected abundances. Thus, the comparisons of the abundances across groups concerns log-fold differences. Compared with several other methods, ANCOMBC shown in simulations the best performance, both in terms of controlling the bias and the False Discovery Rate, and maintaining a high power of detecting differences in bacterial composition (Lin and Peddada 2020a, 2020b). The analyses were performed for the same set of phyla and genera, and with the same factors included in the model, as in the analyses for relative abundances. As well, the BH correction for False Discovery Rate was applied (but within the *ancombc2* function, rather than externally).

The last part of the analyses was aimed at testing correlations between the traits characterizing performance in the feeding trials (MD_{FT}, FC, FD, ADE) and microbial characteristics at the level of individual variation, within the groups of the main factors (i.e., partial correlations). To assess the association of the performance traits with the overall microbial community membership (unweighted UniFrac distances) and community structure (unweighted UniFrac distances), we applied the same *adionis2* PERMANOVA models as described above, but with the performance traits and their interaction with diet as additional predictors. Each of the performance traits was analyzed in a separate model. In the same way we used *ancombc2* to analyze the association of the performance traits with the bias-corrected “absolute” abundances of particular taxa (phyla and genera), again, by adding the performance traits as additional predictors to the same models as used for comparing the abundances. The correlations of the performance traits with relative abundances of the particular phyla and genera were tested in more intuitive way, by fitting linear models (R *lm* function) with the performance traits as the dependent variable, and the microbiome traits as predictors (and the same set of the fixed predictors as used in analyses aimed at testing the effects of experimental factors on the performance traits). In both of the analyses of correlations with abundances of particular taxa, P-values were corrected using False Discovery Rate correction (“BH” option in R function *p.adjust*).

Supplementary Results

The Ci-present microbiome

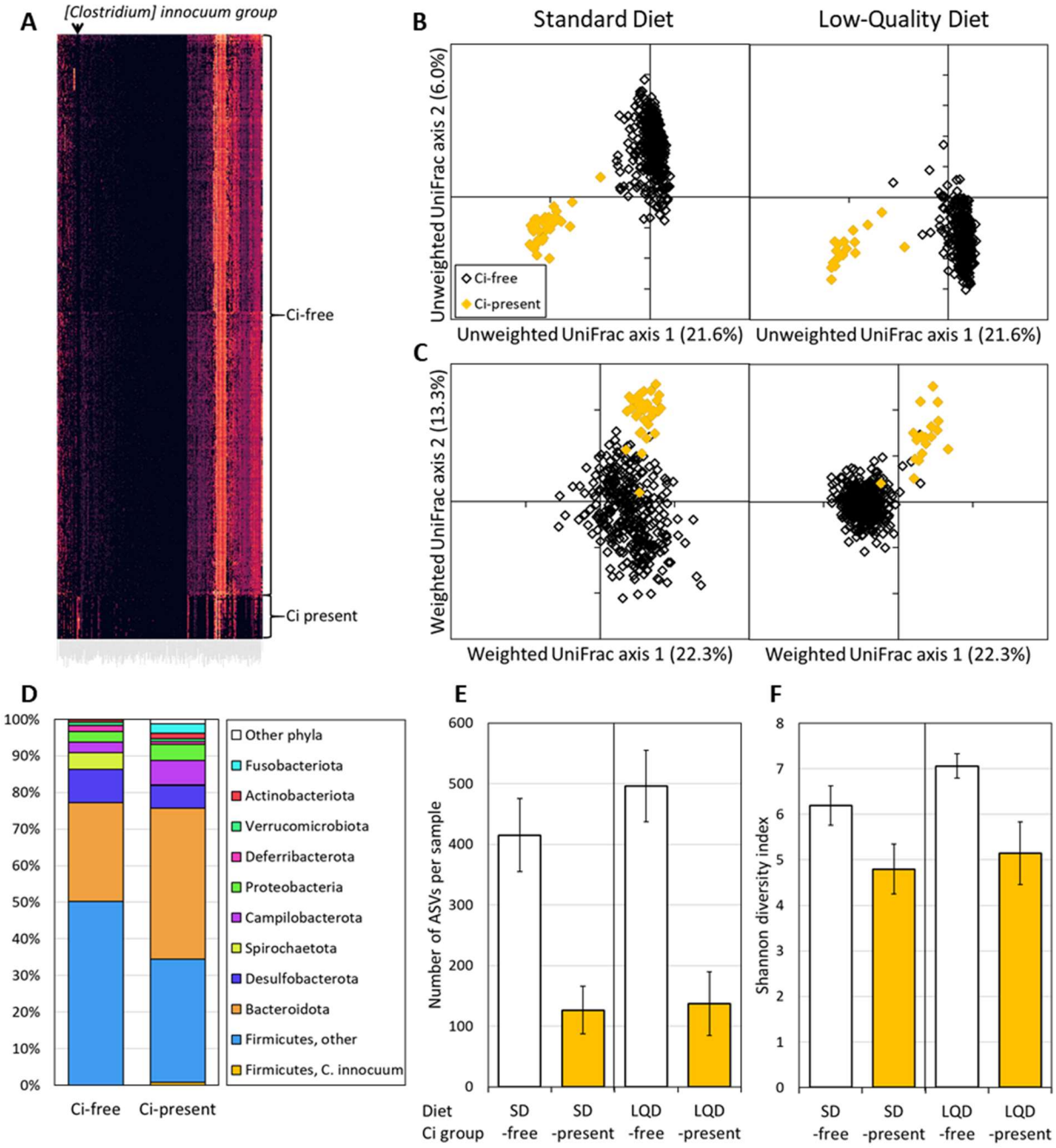

Fig. S2 Caecal microbiome characteristics in bank voles distinguished by presence or absence of bacteria from the [*Clostridium*] *innocuum* group (Ci). **A)** heatmap of abundances of bacterial genera (bacterial genera on horizontal axis, individuals on vertical axis; **B,C)** scores of microbiomes of individual voles on the first two Principal Coordinates Analyses (PCoA) axes based on unweighted **(B)** and weighted **(C)** UniFrac distances (for a better clarity, each displayed on two panels split by diet, but PCoA was performed for all individuals together); **D)** relative abundance of main bacterial phyla ([*Clostridium*] *innocuum* group highlighted within the Firmicutes phylum); and Mean ± SD of **E)** the number of amplicon sequence variants (ASVs) detected in a sample, or **F)** Shannon diversity index of the sample **(F)**.

Preliminary analyses revealed a group of 55 voles (7.4%) with a strikingly distinct bacterial community membership and structure (Fig. S2A,B), which could be nearly perfectly separated by a single criterion, the presence of bacteria from *[Clostridium] innocuum* group. The microbiome of individuals from this group was less diverse, as shown by a significantly lower number of ASVs (N_{ASV}) and Shannon index (Fig. 2E,F). In the *Ci*-free voles (690 individuals), the majority of the bacterial community was formed by two phyla: Firmicutes (50.2%) and Bacteroidota (27.1%), whereas in the 55 *Ci*-present voles, the rank of these phyla was inverted (Bacteroidota 41.3%, Firmicutes 34.4%; Table S3, Fig S2D). The 55 *Ci*-present voles were distributed nearly equally across cage types and sexes (chi-square test of independence performed separately for each of the factors, $p \geq 0.27$), but *Ci-present* phenotype was more common in the group fed SD than LQD (10% vs. 5%; chi-square p=0.026). However, they were present in only 21 out of 148 families, and in 8 of these families all individuals belonged to the *Ci*-present category. The association of the *Ci* presence with family was distinctly non-random (chi-square test with p values based on Monte Carlo randomization, p<1E-6). With such a strong family effect and the overall low proportion of *Ci-presence*, individual-level analyses would be ineffective in answering the question whether the Ci-presence is associated with linetype. Instead, an analysis of frequencies at the level of the full-sib families, in which either at least one individual belonged to the *Ci*-present group *vs* those in which all individuals were *Ci*-free, could be applied. The *Ci*-present families appeared less frequently in the H lines (9/70 families, 13%) than in the C lines (12/57 families, 21%), but the difference was not statistically significant (p=0.35).

The *Ci*-present voles had also a significantly lower body mass at the onset of the feeding trial 20.8±1.22g, vs. 22.0.9±1.05; p=0.007), and a 3.1% points lower apparent food digestibility (p<0.001).

Although we operationally used presence of the bacteria from the [*Clostridium*} *innocuum* group to distinguish the *Ci*-present and *Ci*-free categories, we do not claim that the presence of this particular bacteria was the causal factor behind the distinct microbiomes. Although it is tempting to hypothesize that an infection with this specific bacterium has led to the extinction or decreasing abundance of many bacterial taxa, and has created conditions in which only a few other taxa found favorable conditions (e.g., *Fusobacterium*, which was nearly absent in *Ci*-free voles, or an unnamed genus representing Muribaculaceae, whose abundance grossly increased in the *Ci*-present group; Table S3), the direction of the causal effect could be reversed. Moreover, the development of the distinct microbiome may have been initiated by specific physiological conditions in the vole's caecum, whether determined genetically or environmentally, rather than by the invasion of a particular bacterial species. We have observed a similarly distinct microbiome in a comparable proportion of voles in other studies based on our selection experiment (Lipowska et al., *in* review, Hämäläinen et al., unpublished). Thus, the specific microbiome did not result from an incidental infection during this experiment.

Table R1. Summary of the effects of selection on the relative and absolute abundances of bacterial phyla and genera, and correlations between the these microbial traits and body mass change (MD_{FT}) during the feeding trial and apparent digestive efficiency (ADE). Prev. % - prevalecce, Log_{10} RA% - log_{10}(relative abundance, %); SC - starndard cage, IVC - individually ventilated metabolic cage, SD - standard diet; LQD - low-quality diet. The effects are shown as: plus - positive effect (H lines > C lines, or positive correlaton), minus - negative effect (H < C , or negative correlation); ++ or ––: p<0.01; + or –: p<0.05; (+) or (–): significant overall effect for combined diets, but the differenses or correlations for a given diet considered separatelly not significant. Complete results are presented in Supplementary Tables S5, S7-12.

					offect of				correlations											
			Prev.	Log_{10}	selection				MD_{FT}, SC				MD_{FT}, IVC				ADE			
ITax	Taxon		%	RA %	SD		LQD		SD		LQD		SD		LQD		SD		LQD	
					relative (R) or absolute (A) abundace															
					R	A	R	A	R	A	R	A	R	A	R	A	R	A	R	A
	Phylum																			
1	**Fir**micutes		100	1.7																
2	**Bac**teroidota		100	1.4	–	–	–	–			––						––			
3	**Des**ulfobacterota		100	1.0	++		+				+									
4	**Spi**rochaetota		97	0.7							+									
5	**Cam**pilobacterota		86	0.5	–	–	–	–												
6	**Pro**teobacteria		100	0.5	++	++		+			+									
7	**Def**erribacterota		94	0.2	++	++	+	+												
8	**Ver**rucomicrobiota		94	0.0		–	––	–			++	++								
9	**Act**inobacteriota		100	-0.3	++						+									
10	**Cya**nobacteria		92	-0.7	–								–	–	–	–				
11	**Pat**escibacteria		81	-1.4																
	(Phylum abbrev.) **Genus**																			
159	Fir.	Lachnospiraceae_NK4A136_group	100	0.9																
160	Fir.	uncultured f. Lachnospiraceae	100	0.6	–	–	–	–									––			
161	Fir.	unclassified f. Lachnospiraceae	100	0.4													–			
162	Fir.	Roseburia	100	0.4	–		–				+									
163	Fir.	Lachnospiraceae_UCG-001	99	0.2													++			
164	Fir.	Blautia	99	-0.2	–		–													
165	Fir.	Lachnoclostridium	100	-0.3													––			
166	Fir.	A2	93	-0.3																
167	Fir.	[Eubacterium]_xylanophilum_group	97	-0.3													–		–	
168	Fir.	Lachnospiraceae_UCG-006	99	-0.4	–		–													
169	Fir.	Tyzzerella	97	-0.4							++	+								
170	Fir.	ASF356	87	-0.5		–		–												
171	Fir.	Coprococcus	22	-0.5										––		––				
172	Fir.	Acetatifactor	80	-0.6	+	+	+	+									––			
173	Fir.	GCA-900066575	96	-0.7							––									
174	Fir.	Lachnospiraceae_NK4B4_group	83	-0.8																
175	Fir.	Lachnospira	62	-1.0						+		++								
176	Fir.	Tuzzerella	92	-1.1		–		–			––	–								
177	Fir.	Lachnospiraceae_FCS020_group	89	-1.1													––		–	
178	Fir.	Acetitomaculum	73	-1.3	+															
179	Fir.	Oribacterium	65	-1.4	+		+													
180	Fir.	Lachnospiraceae_AC2044_group	70	-1.7																
181	Fir.	[Ruminococcus]_gauvreauii_group	22	-1.9	–		–					––		–						
182	Fir.	[Ruminococcus]_torques_group	32	-2.0	–	–	–	–		++										
184	Fir.	Eisenbergiella	49	-2.0	+		+							––		–				

					offect of				correlations											
			Prev.	Log_{10}	selection				MD_{FT}, SC				MD_{FT}, IVC				ADE			
ITax	Taxon		%	RA %	SD		LQD		SD		LQD		SD		LQD		SD		LQD	
					relative (R) or absolute (A) abundace															
					R	A	R	A	R	A	R	A	R	A	R	A	R	A	R	A
186	Fir.	Lachnospiraceae_UCG-010	21	-2.6	–		–			(+)		(–)								
187	Fir.	[Eubacterium]_fissicatena_group	22	-2.6	+		+			++				––						
196	Fir.	uncultured f. Oscillospiraceae	100	0.6	–	–	–	–			––									
197	Fir.	Colidextribacter	100	-0.2							––						––			
198	Fir.	Oscillibacter	100	-0.2	+		+													
199	Fir.	UCG-005	97	-0.5							+	+								
200	Fir.	UCG-003	100	-0.7	––	–	––	–			–									
201	Fir.	NK4A214_group	98	-0.7																
202	Fir.	unclassified f. Oscillospiraceae	78	-1.5	+		+										––		–	
203	Fir.	UCG-002	34	-1.5	–	–	–	–						(+)		––				
204	Fir.	UCG-007	17	-2.6																
205	Fir.	Flavonifractor	21	-2.6										++						
207	Fir.	uncultured f. Ruminococcaceae	100	0.4	+		+													
208	Fir.	Ruminococcus	77	-0.1	+		+							–		–	+		+	
209	Fir.	Incertae_Sedis	100	-0.2													––		(–)	
210	Fir.	[Eubacterium]_siraeum_group	91	-0.2	+		+													
211	Fir.	UBA1819	99	-0.4																
212	Fir.	Fournierella	81	-0.5	––	––	–	(–)												
213	Fir.	Ruminococcaceae	71	-0.9																
214	Fir.	unclassified f. Ruminococcaceae	86	-1.4	(+)	(+)	++	++												
215	Fir.	Paludicola	45	-2.2																
216	Fir.	Anaerofilum	29	-2.3	–		–			++				––						
217	Fir.	Anaerotruncus	34	-2.4																
219	Fir.	Butyricicoccus	94	-0.9	–															
220	Fir.	UCG-009	96	-0.9																
221	Fir.	[Eubacterium]_coprostanoligenes_group	69	-1.2	––	–	–	–									+		+	
222	Fir.	UCG-010	66	-1.5										–			++		+	
224	Fir.	unclassified o. Oscillospirales	19	-2.7	+		+			––										
226	Fir.	uncultured f. Christensenellaceae	83	0.4														+		++
227	Fir.	unclassified f. Christensenellaceae	85	0.4													++	+	+	++
228	Fir.	Christensenella	36	-2.4	+		++							–						
230	Fir.	Clostridia_vadinBB60_group	100	0.4							+									
231	Fir.	uncultured f. Peptococcaceae	94	-0.3																
232	Fir.	Peptococcus	95	-1.0	–						––	–								
233	Fir.	Family_XIII_UCG-001	87	-1.1																
234	Fir.	[Eubacterium]_brachy_group	96	-1.2	+															
235	Fir.	Family_XIII_AD3011_group	78	-1.4	(+)		++				–									
236	Fir.	unclassified f. Anaerovoracaceae	24	-2.3	+		+													
237	Fir.	Anaerovorax	23	-2.6													––			
240	Fir.	Monoglobus	33	-0.7			+										–			
241	Fir.	unclassified c. Clostridia	87	-1.3																
242	Fir.	Clostridia_UCG-014	82	-1.3																
243	Fir.	Candidatus_Arthromitus	15	-1.3												––				++
245	Fir.	Anaerofustis	12	-3.0																
247	Fir.	Lactobacillus	95	0.7	++	+	++	+									––			

					offect of				correlations											
			Prev.	Log_{10}	selection				MD_{FT}, SC				MD_{FT}, IVC				ADE			
ITax	Taxon		%	RA %	SD		LQD		SD		LQD		SD		LQD		SD		LQD	
					relative (R) or absolute (A) abundace															
					R	A	R	A	R	A	R	A	R	A	R	A	R	A	R	A
248	Fir.	Streptococcus	56	-1.9						–	––	–					–		–	
254	Fir.	Ileibacterium	18	-0.2	++	++				––							++	++		
255	Fir.	unclassified f. Erysipelotrichaceae	10	-0.2	––	––								––				––		
256	Fir.	uncultured f. Erysipelotrichaceae	27	-0.8	––	––	(–)													++
263	Fir.	Anaeroplasma	50	-0.3	++	++	++	(+)		––				++						
264	Fir.	RF39	86	-1.1																
265	Fir.	Ureaplasma	13	-2.0		(–)		–						+		++				
274	Fir.	Veillonellaceae_UCG-001	18	-0.9	––	++	––	––		+						––				
276	Fir.	Syntrophomonas	31	-2.1	+		+							+						
278	Bac.	Muribaculaceae	100	1.2	––						––							––		
279	Bac.	Alistipes	100	0.5	+		(–)				+									
280	Bac.	Rikenella	69	-0.6		–	–	–						––						
281	Bac.	Rs-E47_termite_group	76	0.5	++													++		
282	Bac.	Odoribacter	97	0.2	++							+								
283	Bac.	unclassified o. Bacteroidales	67	0.2																
286	Des.	Desulfovibrio	100	0.9	+		+				+									
287	Des.	Bilophila	98	-0.5			–													
288	Des.	unclassified f. Desulfovibrionaceae	28	-1.0		++		+		––				++		+				––
289	Des.	uncultured f. Desulfovibrionaceae	77	-1.5	––		–				––	–								
290	Des.	unclassified o. Desulfovibrionales	92	-0.3	++	+	+	+				+					+		+	+
292	Spi.	Termite_Treponema_cluster	69	0.6						+		+		–						
293	Spi.	Treponema	44	-0.8										––			++	++		
295	Spi.	Brachyspira	82	-1.6																
296	Cam.	Helicobacter	86	0.5		–		–												
297	Pro.	uncultured o. Rickettsiales	100	0.3	++	+					++	+								
298	Pro.	uncultured f. Paracaedibacteraceae	90	-0.2	++	++						+								
299	Pro.	Bauldia	17	-1.6						––		(–)		++		––				––
300	Pro.	uncultured o. Rhodospirillales	36	-1.9						–		––								
301	Pro.	unclassified f. Cellvibrionaceae	41	-0.7						––		(–)								++
302	Pro.	Oxalobacter	91	-1.1			–													
303	Pro.	Nitrosomonas	84	-1.6										––						
307	Pro.	unclassified f. Pasteurellaceae	23	-2.4						––	–	(–)								
315	Def.	Mucispirillum	94	0.2	++	++	+	+												
316	Ver.	unclassified f. Puniceicoccaceae	92	0.0	–	–	––	–			++	++								
317	Ver.	uncultured f. Puniceicoccaceae	21	-0.9	++	++	++													
319	Act.	Enterorhabdus	100	-0.5													++			
320	Act.	unclassified f. Eggerthellaceae	60	-1.6																
321	Act.	Adlercreutzia	68	-1.7			–													
323	Act.	uncultured f. Coriobacteriales_Incertae_Sed	32	-1.5																
324	Act.	Bifidobacterium	12	-1.0	++	++				––				++						
325	Act.	Corynebacterium	59	-1.8	++		+													
326	Act.	Brachybacterium	18	-2.8						++				++						
327	Act.	Rothia	11	-3.0												––				
330	Cya.	Gastranaerophilales	92	-0.7	–									–		–				
331	Pat.	Candidatus_Saccharimonas	81	-1.4										–						

The relation between feeding-trial traits and body mass

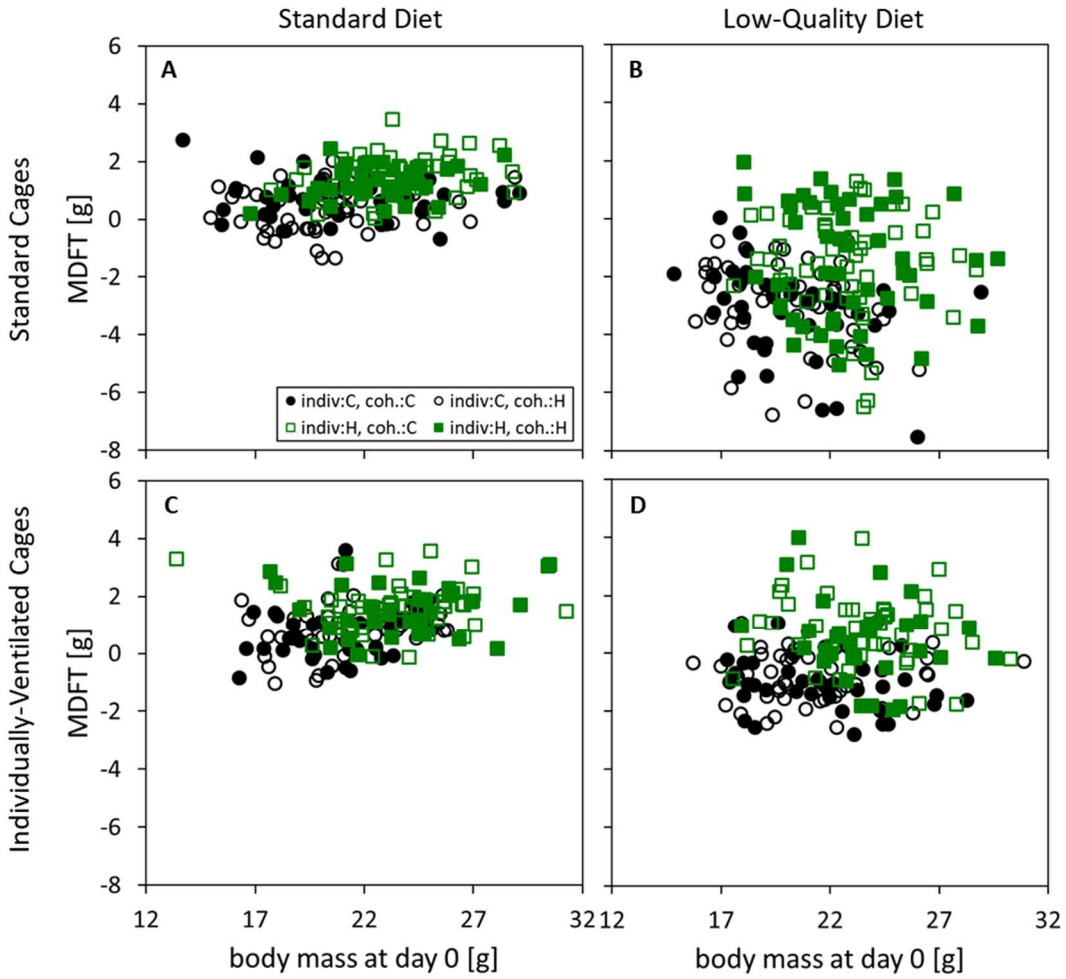

Fig. S3 The relationship between body mass change during the feeding trial (MD_{FT}) and initial body mass, in animals tested in standard cages (top row) or individually-ventilated cages (bottom row) and fed either standard diet (left column) or low-quality diet (right column).

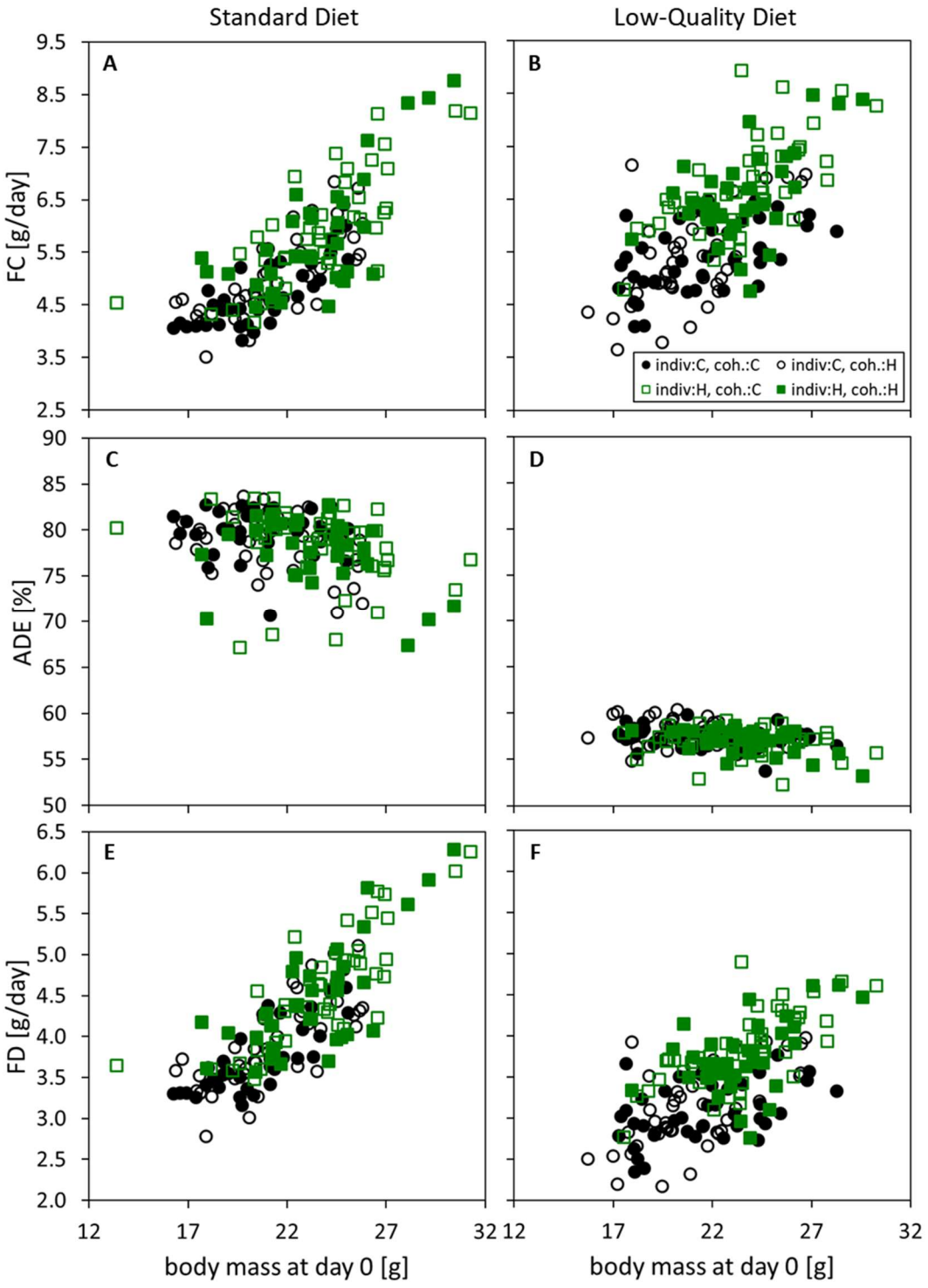

Fig. S4 The relationship between feeding trial performance traits and initial body mass in animals tested in the individually-ventilated cages and fed either standard diet (left column) or low-quality diet (right column). FC - rate of food consumption, ADE – apparent digestive efficiency, FD – rate of food digestion (ADE = 100 × FD/FC).

References

Anderson M.J. 2017. Permutational Multivariate Analysis of Variance (PERMANOVA) . Wiley StatsRef Stat Ref Online 1–15.

Benjamini Y. and Y. Hochberg. 1995. Controlling The False Discovery Rate - A Practical And Powerful Approach To Multiple Testing. J R Stat Soc 57:289–300.

Bolyen E., J.R. Rideout, M.R. Dillon, N.A. Bokulich, C.C. Abnet, G.A. Al-Ghalith, H. Alexander, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857.

Glenn T.C., T.W. Pierson, N.J. Bayona-Vásquez, T.J. Kieran, S.L. Hoffberg, J.C. Thomas, D.E. Lefever, et al. 2019. Adapterama II: Universal amplicon sequencing on Illumina platforms (TaggiMatrix). PeerJ 2019.

Hanhimäki E., P.C. Watts, E. Koskela, P. Koteja, T. Mappes, and A.M. Hämäläinen. 2022. Evolved high aerobic capacity has context-specific effects on gut microbiota. Front Ecol Evol 10.

Henderson N.D. 1997. Spurious associations in unreplicated selected lines. Behav Genet 27:145–154.

Iwaszkiewicz-Eggebrecht E., P. Łukasik, M. Buczek, J. Deng, E.A. Hartop, H. Havnås, M. Prus-Frankowska, et al. 2023. FAVIS: Fast and versatile protocol for nondestructive metabarcoding of bulk insect samples. PLoS One 18:1–13.

Lin H. and S. Das Peddada. 2020a. Analysis of compositions of microbiomes with bias correction. Nat Commun 11:1–11.

_____. 2020b. Analysis of microbial compositions: a review of normalization and differential abundance analysis. npj Biofilms Microbiomes 6.

Lipowska M.M., E.T. Sadowska, U. Bauchinger, W. Goymann, B. Bober-Sowa, and P. Koteja. 2020. Does selection for behavioral and physiological performance traits alter glucocorticoid responsiveness in bank voles? J Exp Biol 223:1–14.

Marizzoni M., T. Gurry, S. Provasi, G. Greub, N. Lopizzo, F. Ribaldi, C. Festari, et al. 2020. Comparison of Bioinformatics Pipelines and Operating Systems for the Analyses of 16S rRNA Gene Amplicon Sequences in Human Fecal Samples. Front Microbiol 11.

Marquina D., M. Buczek, F. Ronquist, and P. Lukasik. 2021. The effect of ethanol concentration on the morphological and molecular preservation of insects for biodiversity studies. PeerJ 9:1–22.

McNamara M.P., J.M. Singleton, M.D. Cadney, P.M. Ruegger, J. Borneman, and T. Garland. 2021. Early-life effects of juvenile Western diet and exercise on adult gut microbiome composition in mice. J Exp Biol 224.

Oksanen J., G.L. Simpson, F.G. Blanchet, R. Kindt, P. Legendre, P.R. Minchin, R.B. O'Hara, et al. 2022. Package "Vegan" Title Community Ecology Package. Cran 1–297.

Prodan A., V. Tremaroli, H. Brolin, A.H. Zwinderman, M. Nieuwdorp, and E. Levin. 2020. Comparing bioinformatic pipelines for microbial 16S rRNA amplicon sequencing. PLoS One 15:1–19.

Quast C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, et al. 2013. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res 41:590–596.

Sadowska E.T., K. Baliga-klimczyk, K.M. Chrząścik, and P. Koteja. 2008. Laboratory model of adaptive radiation: a selection experiment in the bank vole. Physiol Biochem Zool 81:627–640.

Sadowska E.T., C. Stawski, A. Rudolf, G. Dheyongera, K. Baliga-Klimczyk, P. Koteja, and K.M. Chrząścik. 2015. Evolution of basal metabolic rate in bank voles from a multidirectional selection experiment. Proc R Soc B 282:20150025.

SAS Institute Inc. 2011. Base SAS® 9.3 Procedures Guide. Cary, NC: SAS Institute Inc.

Zhang J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620.