

Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades

Nicholas D. Youngblut^{§,*1}, Georg H. Reischer^{§,2,4}, William Walters¹, Nathalie Schuster², Chris Walzer³, Gabrielle Stalder³, Ruth E. Ley¹, Andreas H. Farnleitner^{2,4,5}

¹Department of Microbiome Science, Max Planck Institute for Developmental Biology, Tübingen, Germany

²Technische Universität Wien, Institute of Chemical, Environmental and Bioscience Engineering, Research Group of Environmental Microbiology and Molecular Diagnostics, Vienna, Austria

³Research Institute of Wildlife Ecology, University of Veterinary Medicine, Vienna, Austria.

⁴Interuniversity Cooperation Centre for Water & Health, Vienna, Austria

⁵Research Division Water Quality and Health, Karl Landsteiner University for Health Sciences, Krems, Austria

[§] These authors contributed equally to the publication

* Corresponding author: Nicholas Youngblut

Running title: *Diet and evolution differently explain microbiome diversity*

Key words: animal microbiome, diet, host evolution, gut, community assembly

Abstract

Multiple factors modulate microbial community assembly in the gut, but the magnitude of each can vary substantially across studies. This may be in part due to a heavy reliance on captive animals, which can have very different gut microbiomes versus their wild counterparts. In order to better resolve the influence of evolution and diet on gut microbiome diversity, we generated a large and highly diverse animal distal gut 16S rRNA microbiome dataset, which comprises 80 % wild animals and includes members of Mammalia, Aves, Reptilia, Amphibia, and Actinopterygii. We decoupled the effects of host evolutionary history and diet on gut microbiome diversity and show that each factor explains different aspects of diversity. Moreover, we resolved particular microbial taxa associated with host phylogeny or diet, and we show that Mammalia have a stronger signal of cophylogeny versus non-mammalian hosts. Additionally, our results from ecophylogenetics and co-occurrence analyses suggest that environmental filtering and microbe-microbe interactions differ among host clades. These findings provide a robust assessment of the processes driving microbial community assembly in the vertebrate intestine.

Introduction

Our understanding of the animal intestinal microbiome has now extended far beyond its importance for digestion and energy acquisition, with many recent studies showing that the microbiome contributes to detoxification, immune system development, behavior, postembryonic development, and a number of other factors influencing host physiology, ecology, and evolution^{1,2}. Clearly, the adaptive capacity of an animal species is not determined solely by the host genome, but must also include the vast genetic repertoire of the microbiome³. Concretely understanding how environmental perturbations, host-microbe co-evolution, and other factors dictate the microbial diversity in the animal intestine holds importance for the

conservation and management of animal populations along with determining their adaptive potential to environmental change⁴. However, we are still far from this understanding, especially regarding non-mammalian species and non-captive species in their natural environment.

A number of factors have been either correlated or experimentally shown to modulate microbiome diversity in the animal intestine^{5,6}. While biogeography, sex, reproductive status, and social structure have all been associated with animal gut microbiome diversity in certain animal clades, the consistently dominant drivers appear to be host evolutionary history and diet⁷⁻⁹. For instance, diet can rapidly and reproducibly alter the microbiome in humans and mice^{10,11}. Still, each individual seems to possess a unique microbiome, and studies on humans and animals have identified microbes whose abundances are determined by host genetics (ie., heritable microbes)^{12,13}. Among animal microbiome studies, the magnitude these two drivers can differ substantially among studies. For example, diet was the dominant predictor of microbiome diversity in recent studies of great apes¹⁴, mice¹⁵, and myrmecophagous mammals¹⁶. Other research points to a strong signal of host-microbiome co-evolution (ie, phyllosymbiosis) across many animal clades¹⁷⁻¹⁹, and yet other studies have found very little or no effect of host phylogeny^{15,20,21}.

A current challenge is determining whether these inter-study discrepancies are the result of technical artifacts inherent to differing experimental designs or whether the modulating effects of host diet and evolution on the gut microbiome do truly differ among host clades and/or microbial lineages. Resolving this question has been hampered by multiple factors. First, most studies have focused on narrow sections of the animal phylogeny (eg., primates), with a predominant focus on mammals⁹. In fact, the meta-analysis of Colston and Jackson revealed that <10 % of studies investigating the gut microbial communities of vertebrates were conducted on non-mammalian species⁶. Although meta-analyses can greatly expand the diversity of hosts analyzed, the heterogeneous sample collection and processing methods employed among individual studies can lead to large batch effects and obscure true biological effects^{9,22}. Second,

due to the challenge of sample collection and metadata gathering from wild animals, many studies have utilized captive animals. However, the gut microbiome of wild and captive animals can differ substantially^{6,23,24}, which has led to calls for more studies that assess the microbiomes of wild animals^{9,25}. Third, studies vary in how the effects of evolutionary history are assessed. Host phylogenies are inferred from differing molecular data or sometimes only host taxonomy used as a coarse proxy for evolutionary history^{6,19,21,26}. Finally, host intra-species variation is often removed (ie., just one randomly selected sample used per species), or alternatively it is retained but the potential biases and treatment group imbalances are ignored in hypothesis testing^{8,19}.

To address this challenge, we generated a very large and highly diverse vertebrate distal gut microbiome 16S rRNA dataset, comprising 80 % wild animals that include members of Mammalia, Aves, Reptilia, Amphibia, and Actinopterygii (which diverged from a last common ancestor ~435 MYA). Unlike meta-analyses, this dataset was generated with the same collection methods and molecular techniques performed in the same facility, which reduces batch effects that plague meta-analyses. We utilized a robust analytical framework to resolve the relative importance of host diet and evolutionary history (along with other host characteristics) on gut microbiome diversity. Moreover, we identified particular microbial operational taxonomic units (OTUs) that associate with diet or host phylogeny after controlling for the effect of the other factor. Finally, we utilized ecophylogenetic and co-occurrence analyses to investigate the effects of environmental filtering and microbe-microbe interactions on microbial community assembly in the vertebrate intestine.

Methods

Sample collection

Fecal sampling was conducted between February 2009 and March 2014. Only fresh samples with confirmed origin from a known host species were collected, most of them by wildlife biologists conducting long-term research on the respective species in its habitat. This also ensured that sampling guidelines and restrictions were adhered to, where these were applicable. Human DNA samples were taken from a previous study²⁷. Samples originate predominantly from Central Europe (Austria and neighboring countries). However, in order to cover as much of vertebrate diversity as possible, many samples were also taken from other countries around the world (19 countries on 6 continents; see Supplementary Fig. 1). Detailed metadata on the sampled animal species (eg., diet and habitat), the sampling location and conditions were collected alongside the fecal samples (Supplementary Table 1).

All fecal samples were collected in sterile sampling vials, transported to a laboratory and frozen within 8 hours. Samples were stored at -20°C and shipped on dry ice to TU Wien in Vienna, Austria within weeks after collection. In Vienna, DNA extraction was performed within two months after receiving the samples using the PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, USA) in combination with bead-beating (FastPrep-24, MP Biomedicals, Santa Ana, USA). DNA concentration in extracts was measured using a NanoDrop ND 1000 UV spectrophotometer and the Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher Scientific Inc., Vienna, Austria). DNA extracts were stored at -80°C until further analysis.

16S rRNA gene sequencing

PCR amplicons for the V4 region of the 16S rRNA gene were generated with primers 515F-806R²⁸ and were sequenced with the Illumina MiSeq 2x250 v2 kit at the Cornell University Institute for Biotechnology. DADA2²⁹ was used to call 100 % sequence identity OTUs (ie.,

sequence variants). Taxonomy was assigned to OTUs with the QIIME2 q2-feature-classifier³⁰ using the SILVA database (v119)³¹. The phyloseq³² R package was used to rarefy total OTU counts to 5000 per sample due to the multiple orders of magnitude difference in raw counts among samples. A phylogeny was inferred for all OTU sequences with fasttree³³ based on a multiple sequence alignment generated by mafft³⁴. All samples lacking metadata used in the study were filtered from the dataset. In cases where an individual host was sampled multiple times, we randomly selected one sample.

Host phylogeny

A dated host phylogeny was obtained from TimeTree.org³⁵. To create a phylogeny for all samples (Supplementary Fig. 2), sample-level tips were grafted onto the species-level tips with a negligible branch length.

Intra-species sensitivity analysis

The dataset contained a variable number of samples per host species, and species were asymmetrically represented among clades (Fig. 1). Moreover, the host phylogeny did not include within-species relatedness information, which would cause zero-inflation in our analyses of coevolution. Therefore, we used a sensitivity analysis approach (inspired by the sensiphy³⁶ R package) that assessed the sensitivity of all analyses in this study (unless noted otherwise) to intra-species heterogeneity in microbiome diversity and host metadata. This method consisted of generating 100 subsamples of the dataset, each with just one randomly selected sample per host species. For each hypothesis tested in the study, the test was applied to each dataset subset, and the overall hypothesis test was considered significant if ≥ 95 % of the subsets were each considered significant after correcting for multiple hypothesis testing with the Benjamini-Hochberg procedure.

Data analysis

General manipulation and basic analyses of the dataset were performed in R³⁷ with the phyloseq, dplyr, tidyr, and ggplot2 R packages³². High-throughput compute cluster job submission was performed with the batchtools³⁸ R package. Phylogenies were manipulated with the ape³⁹ and caper⁴⁰ R packages and visualized with iTOL⁴¹. Networks were manipulated and visualized with the tidygraph⁴² and ggraph⁴³ R packages, respectively.

Similarity of OTUs to cultured representatives in the SILVA All Species Living Tree database³¹ was conducted by BLASTn⁴⁴ of OTU representative sequences versus the 16S sequence database. We filtered out all BLASTn hits with an alignment length of <95 % the query sequence length. Similarity of OTUs to any representatives in SILVA was conducted in the same manner, but the BLAST database was SILVA release 132, de-replicated at 99 % sequence identity.

Multiple regression on matrices (MRM) was performed with the Ecodist⁴⁵ R package, with rank-based correlations. We converted all regression variables to distance matrices through various means. The host phylogeny was represented by the patristic distance (branch lengths). We calculated the Gower distance for the detailed diet data, detailed habitat data, and sample type data (wild/captive animal + gut/feces sample origin; see Fig. 1). Geographic distance was represented as great-circle distance based on latitude and longitude. Alpha-diversity was converted to a distance matrix by taking the Euclidean distance among all pairwise sample comparisons.

Procrustean Approach to Cophylogeny (PACo)⁴⁶ and Parafit³⁹ were performed on the host phylogeny and microbial 16S rRNA phylogeny, along with a matrix of OTU presence/absence among hosts. The “Cailliez” correction for negative eigenvalues was applied for both PACo and Parafit. For PACo, we used the quasiswap null model, which does not assume that the symbiont is tracking the evolution of the host or vice versa (a conservative

approach). For each method, 1000 permutations were used. Phylogenetic signal of OTUs was tested with the PhyloSignal⁴⁷ R package. Binomial regression on OTU presence/absence was used to regress out the effects of diet, and the residuals were used for tests of phylogenetic signal.

The Local Indicator of Phylogenetic Association (local Moran's I; "LIPA") was calculated with 9999 permutations. Phylogenetic general least squares models (PGLS) conducted with caper⁴⁰ R package. A Brownian motion model of evolution was used. For beta-diversity, the first 5 PCoA eigenvectors were used. Co-occurrence analyses were conducted with the cooccur⁴⁸ R package. The walktrap algorithm⁴⁹ for defining sub-networks in the co-occurrence network. For OTU-specific tests (LIPA, PGLS, and co-occurrence), only OTUs present in >5 % of samples were included.

Jupyter notebooks describing the entire data analysis process are available on GitHub at https://github.com/leylabmpi/animal_gut_16S-uni.

Data availability

The raw sequence data are available from the European Nucleotide Archive under the study accession number PRJEB29403. Metadata associated with each sample are provided in Supplementary Table 1.

Results

Concept of sampling

In order to have a comprehensive representation of vertebrate intestinal microbiota, we collected fresh fecal/gut samples of animals from five host classes: Mammalia, Aves, Reptilia, Amphibia, and Actinopterygii. Sampling was mostly restricted to animals living in the wild, with

193 some additional samples originating from domesticated livestock and pets (see Supplementary
 194 Table 1). We generally refrained from collecting samples from animals living in zoos (20 of the
 195 39 samples from captive animals) because artificial habitat, diet, and medication may have
 196 strong confounding effects on the natural intestinal communities. No samples were collected
 197 from aquariums. The majority of the samples were collected in Central Europe and
 198 supplemented with samples from other regions to cover phylogenetic groups lacking extant
 199 members in this region (eg., Afrotheria, Marsupialia, Primates or Cetacea). Samples were
 200 obtained by specialized wildlife biologists working with the host species in the field in order to be
 201 certain of the sample origin. In total, the dataset includes 213 samples from 128 species, each
 202 with detailed diet, habitat, and additional metadata (Fig. 1).

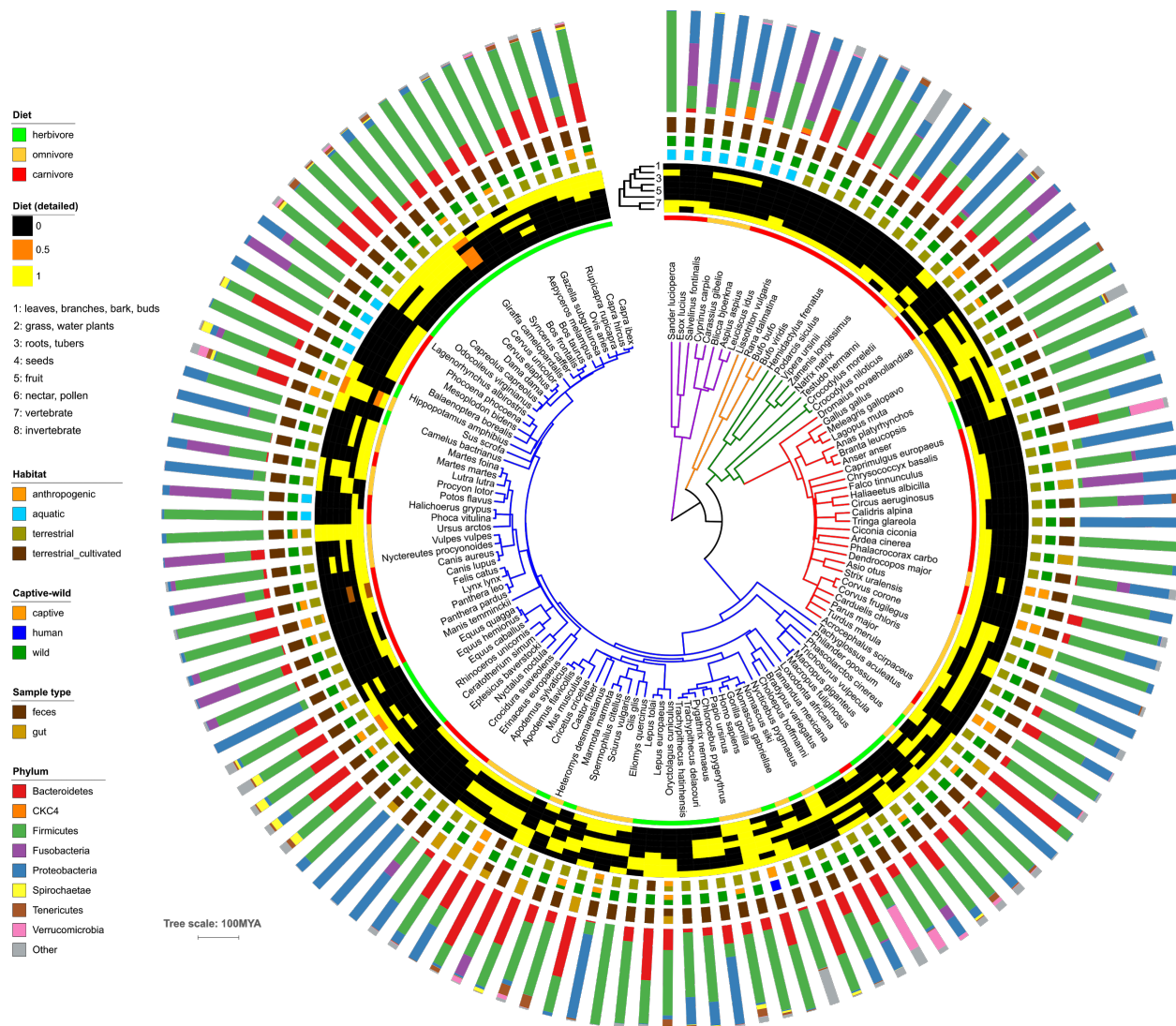


Fig. 1. Phylum-level grouping of microbiome diversity by host phylogeny and host metadata. The dated host phylogeny was obtained from TimeTree.org, with branches colored by host class (purple = Actinopterygii; orange = Amphibia; green = Reptilia; red = Aves; blue = Mammalia). The tree scale is MYA. From inner to outer, the data mapped onto the tree is host diet (general), host diet (detailed breakdown), host habitat, host captive/wild status, the microbiome sample type, and the relative abundances of microbial phyla in each host. Relative abundances are an estimated average generated via subsampling OTUs from all samples for each host species.

Low prevalence and limited representation of cultured isolates

We sequenced the 16S rRNA V4 region from feces or gut contents of all 213 samples and generated OTUs (resolved at 100 % sequence identity) with the DADA2²⁹ pipeline, which produced a total of 30,290 OTUs. Most OTUs were only detected in ≤ 5 % of samples (Supplementary Fig. 3), which may be due to the high taxonomic and ecological diversity of the hosts. Therefore, we utilized presence-absence for all subsequent OTU-based analyses unless noted otherwise (eg., for abundance-based beta-diversity metrics). At the phylum level, 2 clades were found in at least one individual per species: Firmicutes (mainly Clostridia) and Proteobacteria (mainly Beta- and Gammaproteobacteria). The next-most prevalent phyla were Actinobacteria and Bacteroidetes, which were found in ~ 87 and ~ 86 % of host species, respectively (Supplementary Fig. 3).

Mapping phylum-level relative abundances onto the host phylogeny revealed some clustering of microbiome composition by host clade and diet (Fig. 1). Notably, hosts from the same species generally showed similar phylum-level abundances (Supplementary Fig. 2). We quantified this clustering of microbiome composition on the host tree by calculating beta-dispersion (beta-diversity variance within a group) at each host taxonomic level (host class down to species), and indeed we found beta-diversity to be constrained (more clustered) at finer taxonomic resolutions regardless of the beta-diversity metric (Supplementary Fig. 4).

Many of the phylum-level distributions resembled observations from other studies. For instance, Actinopterygii (ie., ray-finned fishes) samples were mostly dominated by Proteobacteria (Fig. 1), which is consistent with a meta-analysis of fish gut microbiomes⁵⁰. Fusobacteria abundance ranged from 6-35 % among the *Crocodylus* species, which is reflective of high Fusobacteria abundance previously observed in alligators⁵¹. Spirochaetae showed high clade specificity for Perissodactyla, Artiodactyla, and Primates, which matches previous observations⁵²⁻⁵⁴. The CKC4 phylum, which lacks cultured representatives, was markedly

abundant in many Actinopterygii samples, reflecting its previous observation in marine species^{55,56}.

Given the potential for observing novel cultured and uncultured microbes among the phylogenetically diverse and mostly wild hosts, we assessed how many OTUs in the dataset were closely related to cultured and uncultured representatives in the SILVA database. We found that the vast majority (~67 %) lacked a BLASTn hit to a cultured representative at a 97 % sequence identity cutoff (Supplementary Fig. 5A). Even at a 90 % cutoff, ~27 % of OTUs lacked a representative. Most OTUs lacking a representative were Bacteroidetes or Firmicutes (46 and 12 %, respectively; Supplementary Fig. 5B). Mammalia hosts possessed the majority of OTUs lacking closely related cultured representatives, but still hundreds of OTUs, mainly belonging to Actinobacteria, Proteobacteria, and Verrucomicrobia phyla were associated with non-mammalian hosts (Supplementary Fig. 5C). In regards to completely novel diversity, ~22 % of the OTUs lacked any representative sequence in the entire SILVA r132 database at a 97 % sequence ID cutoff. These novel OTUs showed a similar taxonomic composition and distribution among host classes as those OTUs lacking cultured representatives (Supplementary Fig. 5C).

Host phylogeny and diet explain microbiome diversity

We utilized multiple regression on matrices (MRM) to test how well gut microbiome diversity could be explained by host phylogeny, diet, habitat, geographic location, and technical variation. We chose MRM because host phylogeny and geographic location can be directly represented as distance matrices (patristic distance and great-circle distance, respectively) and measuring host phylogenetic similarity as a continuous variable (patristic distance) versus a discrete variable (taxonomic groupings) alleviates imbalances in representation for specific host taxonomic groups (eg., Mammalia was highly represented). Host metadata that could not inherently be described as a distance matrix (eg., the diet components of each species) were converted to distance matrices by various means (see Methods). We had no data on the genetic

similarity of individuals within host species, and thus we conducted our analysis at the species level. To estimate the effects of intra-species variation in host microbiome and metadata on our MRM analysis, we performed the analysis on 100 subsampled datasets, each comprising one randomly selected sample per species. Unless noted otherwise, we used this sensitivity analysis approach for all hypothesis testing in this study (see Methods).

Each of our four MRM models (one per diversity metric) had a significant overall fit ($P < 0.005$ for all models). Host diet and phylogeny were the only significant explanatory variables (Fig. 2). Diet explained a substantial amount of alpha- and beta-diversity variation (~20-30 %) and was significant for all diversity metrics tested (ie., Shannon index, Faith's PD, unweighted Unifrac, and weighted Unifrac). However, host phylogeny was only significant for unweighted Unifrac and explained approximately 15 % of the variance, suggesting that host phylogeny mainly dictates community composition, but not OTU abundances. These findings were supported by principal component analysis (PCoA) ordinations of weighted and unweighted Unifrac values, which displayed clustering by host taxonomy and diet (Supplementary Fig. 6).

Neither host habitat nor geographic location were significant; however, we must note that the experimental design was not directly designed to test this hypothesis (Supplementary Fig. 1). Importantly, the "Technical" covariate, which comprised sample type (feces versus gut contents) and captivity status (wild versus captive) also lacked significance for all models, suggesting no substantial effect of technical variation in our dataset. Also, we did not detect any major outlier samples in our dataset that may be skewing our results (Supplementary Fig. 7). Lastly, we obtained similar results to our initial MRM analysis when we randomly selected one sample per family instead of per species (Supplementary Fig. 8), which reduced the Mammalia:non-Mammalia bias from 64 % of samples being mammalian to 42 %. However, phylogeny was not quite significant ($P = 0.12$), likely due to the reduced sample size.

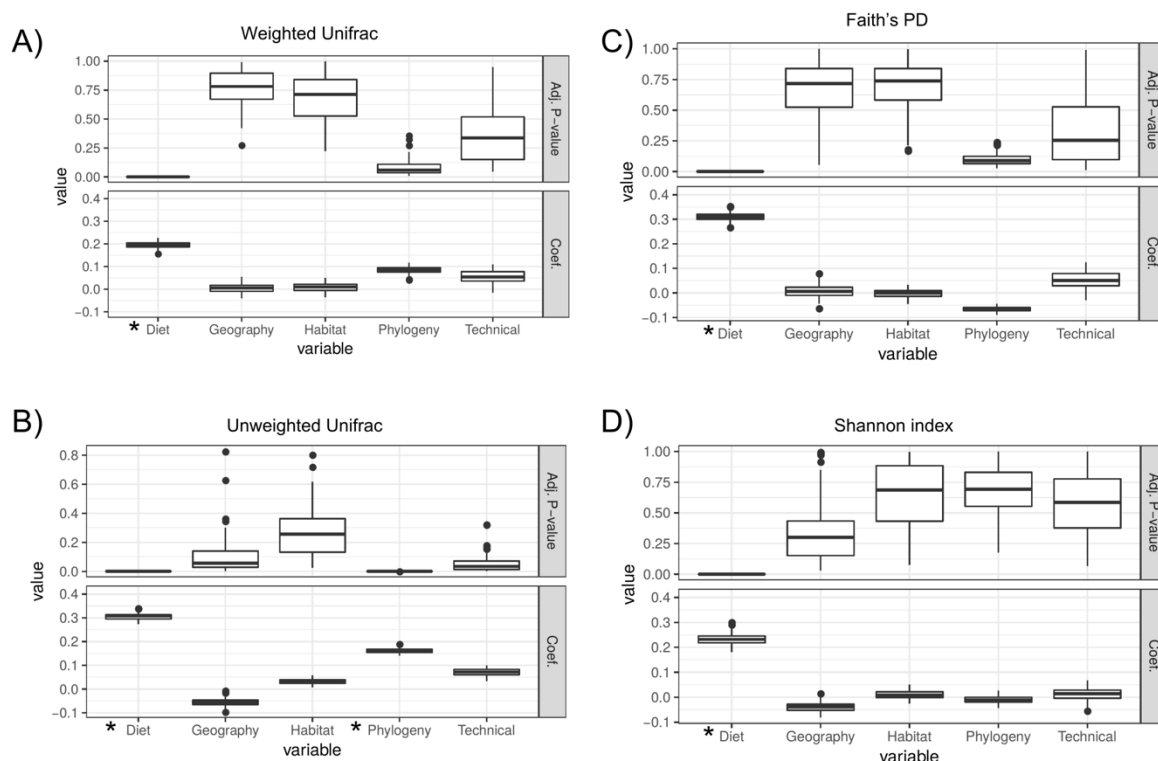


Fig. 2. Host phylogeny and diet significantly explain aspects of microbiome diversity. The plots show the BH-adjusted p-values (“Adj. P-value”) and partial regression coefficients (“Coef.”) for multiple regression on matrices (MRM) tests used to determine how much alpha- or beta-diversity variance was explained by host diet, geographic location, habitat, phylogeny, and technical parameters (see Methods). The boxplots show the distribution of values obtained when running MRM on each of 100 random dataset subsets, with each subsample comprising just one sample per species. The boxplots show the MRM rho coefficient and coefficient p-value for each subsample. See Methods for a description of how each distance matrix for the MRM models was generated. “*” signifies significance (Adj. p-value <0.05 for ≥95 % of dataset subsets; see Methods).

Further resolving the effects of host phylogeny and diet

Our MRM analyses suggest that host phylogeny and diet explain gut microbiome diversity, but this is only one line of evidence, and it does not finely resolve which particular aspects of diversity (eg., particular OTUs) correspond with host diet and phylogeny. Therefore, we employed complementary tests to our MRM analyses to support and further investigate our findings. While animal host phylogeny is somewhat correlated with diet, our dataset comprised a highly taxonomically diverse set of species with substantially varying diets, which often did not correspond to phylogenetic relatedness (Fig. 1). We exploited this lack of complete correspondence between host phylogeny and diet to decouple the effects of each variable on microbial community diversity.

We used phylogenetic generalized least squares (PGLS) to quantify the association of diet with microbial diversity while accounting for host phylogeny. In support of our MRM results, both alpha- and beta-diversity were significantly explained by host diet (Fig. 3A, B). We also conducted the analysis on individual OTUs, and found only 2 OTUs to be significant (Fig. 3C). These OTUs belonged to the Ruminococcaceae and Bacteroidaceae families, respectively. Mapping the distribution of these 2 OTUs onto the host phylogeny revealed that the Ruminococcaceae OTU was associated with many hosts in the herbivorous Artiodactyl clade and also in the southern white-cheeked gibbon (*Nomascus siki*), which is a herbivore in the distantly related primate clade (Supplementary Fig. 9). In contrast, the Bacteroidaceae OTU was predominantly present among multiple distantly related herbivorous clades. The ability of diet to explain overall community alpha- and beta-diversity but only two OTUs support a hypothesis where diet predominantly selects for functional guilds of microbes (eg., cellulolytic consortia) rather than specific OTUs.

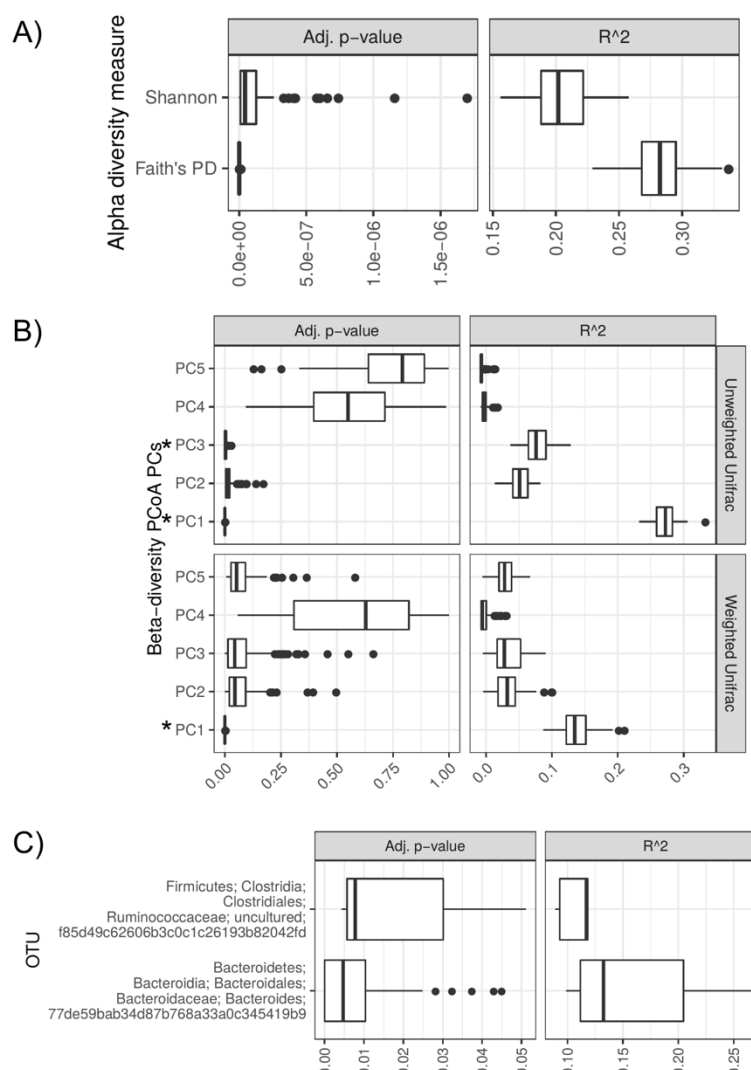


Fig. 3. After accounting for host phylogeny, diet significantly explained alpha- and beta-diversity components but could only explain the prevalence of two OTUs. The boxplots are distributions of PGLS R² and Adj. P-values for 100 random subsamples of the datasets (1 per species for each subsample). A) Both alpha-diversity measures were found to be significant. B) Some PCoA PCs were significantly explained by diet ("*" denotes Adj. P-value <0.05). The percent variance explained for each unweighted Unifrac PC is 18.1, 6.9, 4.2, 3.6 and 2.1 and each weighted Unifrac PC is 27.2, 10.6, 9.6, 6.4, 6.0, and 5.5. C) Only two OTUs were found to be significant.

To assess the effects of host phylogeny while controlling for diet, we utilized tests for phylogenetic signal after regressing out diet. More specifically, we utilized the Local Indicator of Phylogenetic Association (LIPA) to assess whether OTU prevalence (ie., % of samples where present) was similar among closely related hosts. In contrast to the PGLS analysis, we found very little phylogenetic signal of alpha-diversity (Supplementary Fig. 10). This finding is consistent with the MRM analysis results. Also in contrast to the PGLS analysis, we identified 121 OTUs with significant local phylogenetic signal in the host tree (Fig. 4A). These “LIPA-OTUs” differed greatly in which host clades they were associated with. More specifically, the number of LIPA-OTUs per host species ranged from 1 to 34, with only 21 hosts possessing at least 1 LIPA-OTU. OTU-specific phylogenetic signal was only associated with Mammalia species, suggesting weak or no effects of evolutionary history for non-mammalian hosts. Herbivorous species possessed the majority of LIPA-OTUs, but a minority of these OTUs were associated with some omnivorous and carnivorous species (Fig. 4A). LIPA-OTU composition varied among host clades, regardless of whether they shared the same diet (Fig. 4B), which indicates that the phylogenetic signal is indeed a result of host evolutionary history and not contemporary diet. LIPA-OTUs were most predominant among Artiodactyla species, with Primates and Perissodactyla ranked a distant second and third (Fig. 4B). This finding suggests that the effects of host evolutionary history within Mammalia are most pronounced for Artiodactyla. Interestingly, there was no OTU-specific phylogenetic signal for any macropods, even though they are foregut fermenters similar to the Artiodactyla. The same is true of Carnivora species, except for 2 members of the Felidae clade (*Felis catus* and *Panthera pardus*). Altogether, these findings support the hypothesis that mammalian evolutionary history dictates the prevalence of certain OTUs.

The LIPA-OTUs belonged to 7 bacterial phyla and 1 archaeal phylum (Fig. 4C; Supplementary Fig. 11). Firmicutes was dramatically more represented than other phyla, with *Bacteroides* the second-most common. Members of Bovidae consistently had the highest

numbers of these two phyla. This finding is supported by Sasson and colleagues¹³, who only identified Bacteroides and Firmicutes to be heritable in cattle. The majority of the Firmicutes OTUs were members of the Ruminococcaceae, and while most of Ruminococcaceae OTUs were associated with Artiodactyla hosts, some were also observed in certain members of the Primates, Rodentia, and Perissodactyla. Other OTU clades with significant phylogenetic signal included Christensenellaceae, *Blautia*, and *Methanobrevibacter*, which were all found to be consistently heritable among multiple human cohort studies^{12,57}. Interestingly, while humans are represented in this dataset, and a few OTUs were associated with some of the primate species, no OTUs showed a phylogenetic signal with humans (Fig. 4A). Among some very closely related OTUs, we observed that host clade specificity differed, suggesting that these taxa have diversified via adaptive specialization for particular hosts (Supplementary Table 2; Supplementary Fig. 11).

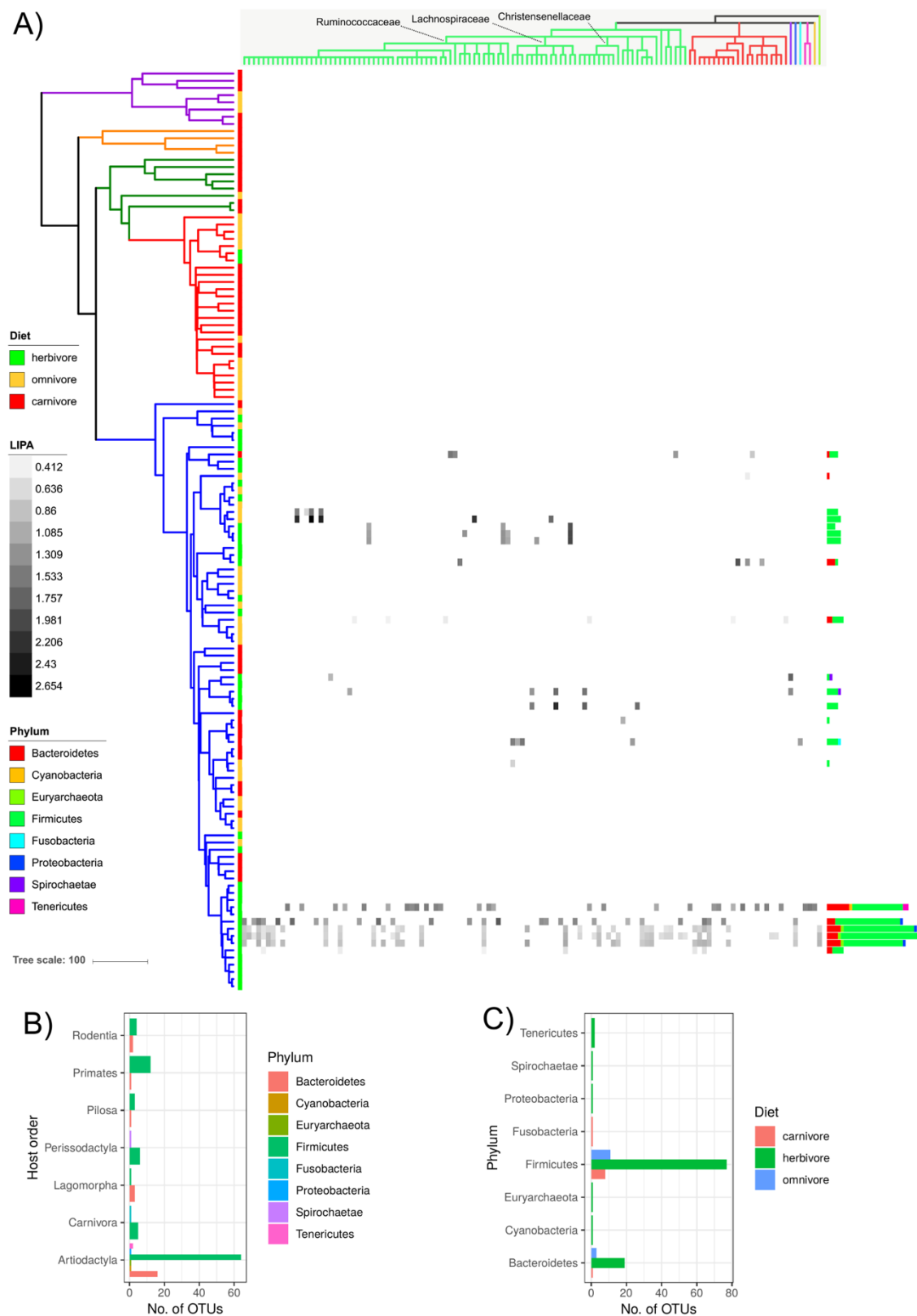


Fig. 4. Many OTUs display a local phylogenetic signal in specific host clades after accounting for diet. A)

The phylogeny is the same as shown in Fig. 1. The heatmap depicts Local Indicator of Phylogenetic Association (LIPA) values for each OTU-host association, with stronger values indicating a stronger phylogenetic signal of OTU presence (with diet regressed-out). White boxes in the heat map indicate non-significant LIPA indices. The dendrogram on the top of the heatmap is a cladogram based on the SILVA-derived taxonomy for each OTU (see Supplementary Fig. 11 for the full taxonomy). The dendrogram is colored by phylum. The bar plots in B) and C) show the number of OTUs with a significant LIPA index per host (OTUs are colored by Phylum; the number of OTUs per host ranges from 0 to 34). B) The bar plots summarize the number of significant OTUs per host order and diet. The bar plots in C) are the same as B) except the data is grouped by OTU phylum.

A stronger pattern of cophylogeny in Mammalia versus non-mammals

Our finding that only Mammalia possessed OTUs with local phylogenetic signal suggests that the effects of evolutionary history on intestinal microbiome diversity may be stronger for Mammalia versus non-mammalian species. We investigated this finding by performing cophylogeny analyses, which determines whether the phylogenies of the host and symbiont (microbe) correspond in their branching patterns. While a positive correlation can be the result of other processes besides co-cladogenesis⁵⁸, the pattern is consistent with model of host-symbiont coevolution. We first utilized Procrustean Approach to Cophylogeny (PACo), which performs Procrustes superimposition to infer the best fit between host and symbiont phylogenies based on symbiont occurrences in the hosts. This permutation-based approach does not rely on distribution assumptions. Moreover, the analysis generates residuals of the Procrustean fit, which describes the contribution of each individual host-symbiont association to the global fit (smaller residuals means a better fit).

The PACo analysis showed a significant global fit, regardless of intra-species heterogeneity ($P < 0.002$ for all dataset subsets). Host-microbiome residuals decreased in the order of Actinopterygii > Amphibia > Reptilia \geq Aves > Mammalia, with the most dramatic

decrease between Aves and Mammalia (Fig. 5), indicating that Mammalia show the strongest signal of cophylogeny. The residuals significantly differed by both host class and diet (ANOVA, $P = 1e-16$ for both), but the effect size was much larger for class versus diet (F-value of 972.3 vs 536.3). Thus, while diet may somewhat confound the signal of cophylogeny, it is likely not the main driver. Conducting PACo and just mammalian species still showed a significant global fit ($P < 0.002$), and we found that Artiodactyla have the smallest distribution of residuals (Supplementary Fig. 12A). Excluding all Artiodactyla samples did not substantially change the results (global fit: $P < 0.003$); neither did subsampling just one sample per family in order to decrease the imbalance of host species per clade (global fit: $P < 0.003$; Supplementary Fig. 12B, C).

We additionally evaluated patterns of cophylogeny with the Parafit analysis, which is also a permutation-based method but assesses similarity of principal coordinates derived from the host and symbiont phylogenies. As with PACo, the global Parafit test was significant ($P < 0.001$), and Mammalia showed the strongest signal of cophylogeny (Fig. 5). Altogether, these data support a model of host-microbe coevolution, with Mammalia displaying the strongest cophylogeny signal.

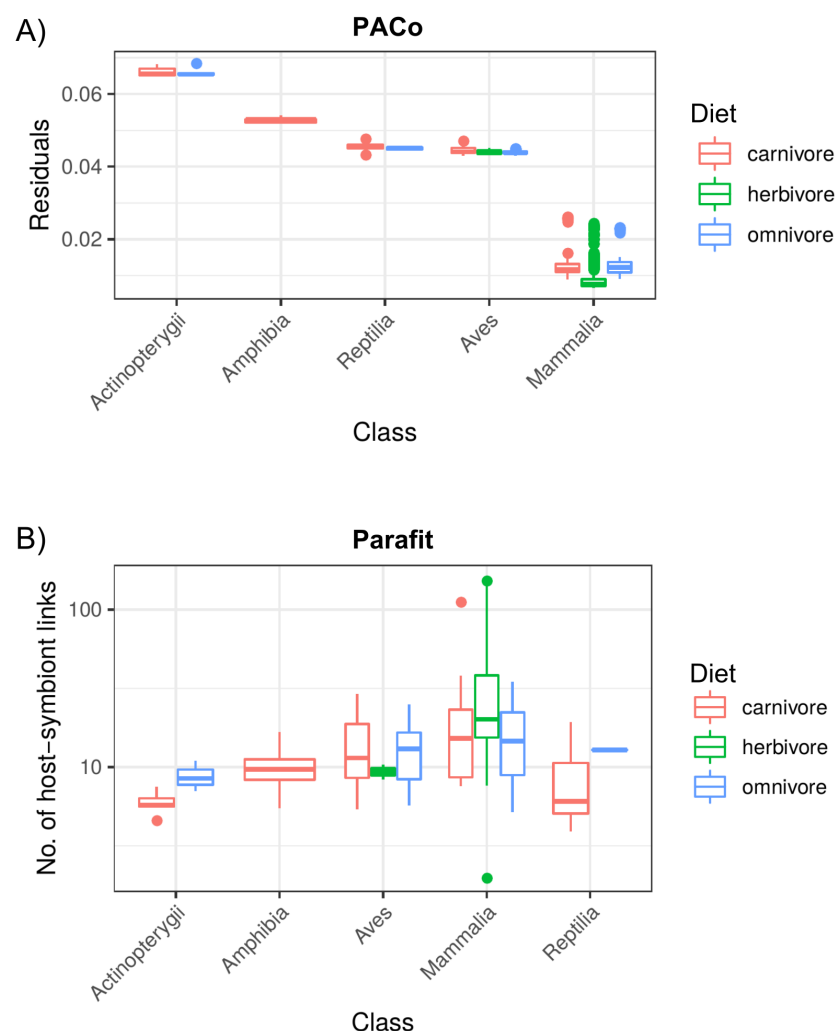


Fig. 5. PACo and Parafit show a stronger cophylogeny signal for Mammalia versus non-mammals. A) Boxplots of PACo residuals between hosts and OTUs (smaller residuals means a stronger cophylogeny signal), with residuals grouped by host class and diet. B) Boxplots of significant host-symbiont links as determined by Parafit analysis, with links group by host class and diet. For both PACo and Parafit, 1000 permutations were performed on each of the 100 dataset subsets.

The roles of environment filtering and microbe-microbe interactions in community assembly

Our findings that diet and host evolutionary history significantly explain microbiome diversity indicate that environmental filtering plays a substantial role in microbial community assembly. In order to further test this notion and to assess how environmental filtering may differ among host clades, we utilized two ecophylogenetics analyses: Mean Phylogenetic Distance (MPD) and Mean Nearest Taxon Distance (MNTD). These tests assess the degree of phylogenetic clustering within each sample (host) relative to a permuted null model. Assuming phylogenetic niche conservatism (ie., closely related taxa overlap along niche axes), then host diet or gut physiology may select for phylogenetically clustered taxa with overlapping niches. While in the absence of such strong selection, competition via niche conservatism would lead to phylogenetic overdispersion⁵⁹. Phylogenetic overdispersion may also result from facilitation (ie., beneficial microbe-microbe interactions), such as when distantly related taxa form consortia to break down complex plant polymers⁵⁹. MPD is more sensitive to overall patterns of phylogenetic clustering and evenness, while MNTD is more sensitive to patterns at the tree tips⁶⁰.

We found that the majority of host species showed significant clustering for MNTD, with close to half for MPD (Fig. 6). Very few species showed phylogenetic evenness. Of those that did, all belonged to the Artiodactyla, except for the long-eared owl (*Asio otus*; Fig. 6). In support of these findings, Gaulke and colleagues also found lower signals of phylogenetic clustering in the Artiodactyla relative to other mammalian clades⁶¹. These findings suggest that community assembly differs between Artiodactyla and non-Artiodactyla mammals, with microbe-microbe competition and/or facilitation surpassing gut environmental filtering among Artiodactyla species.

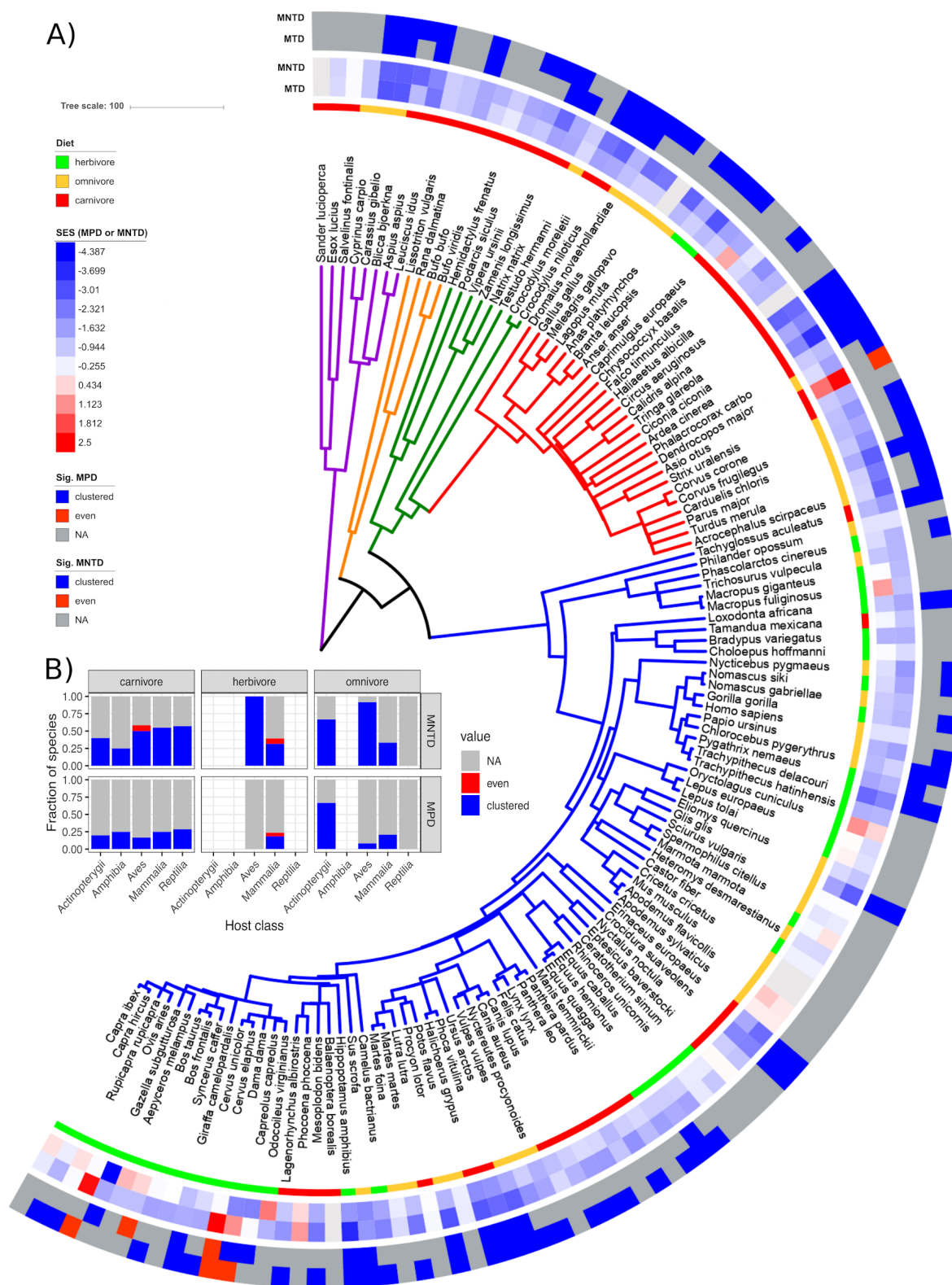


Fig. 6. Microbial communities are generally phylogenetically clustered versus evenly distributed. A) The

phylogeny is the same as shown in Fig. 1. From inner to outer, the data mapped onto the tree is host diet, mean standardized effect sizes for MPD and MNTD, and samples with significant phylogenetic clustering or evenness based on MPD or MNTD. The animal species possessing microbial communities that were phylogenetically evenly distributed were the long-eared owl (*Asio otus*), fallow deer (*Dama dama*), red deer (*Cervus elaphus*), cattle (*Bos taurus*), and sheep (*Ovis aries*). B) The bar charts depict the fraction of host species for each host class/diet where microbial taxa are more phylogenetically clustered (“clustered”) or evenly distributed (“even”) than expected from the null model, or those that did not deviate from the null model (“NA”).

We next tested how microbes co-occur among hosts, which can be influenced by selective pressures or microbe-microbe interactions. Specifically, we conducted a co-occurrence analysis to determine which OTUs significantly positively or negatively co-occurred relative to a permuted null model. Our analysis revealed that almost all significant co-occurrences were positive (Fig. 7A; Supplementary Fig. 13A). The co-occurrence network consisted of 4 sub-networks, each with differing taxonomic compositions and existence of “hub” OTUs (Fig. 7A, D). Sub-networks 1 and 2 were dominated by Ruminococcaceae and Peptostreptococcaceae, with Ruminococcaceae OTUs acting as central hubs in both (Supplementary Fig. 14). Sub-network 3 contained an Enterobacteriaceae (Proteobacteria) OTU hub and also possessed more members of Clostridiaceae, Lachnospiraceae and Enterobacteriaceae. Sub-network 4 did not have a strong hub OTU and contained the most taxonomic diversity (Fig. 7A, D). Interestingly, *Methanobrevibacter* OTUs were only found in Sub-network 1 and significantly co-occurred with Christensenellaceae OTUs as previously seen in a large human cohort study⁵⁷. The presence of OTUs from each sub-network differed substantially among host clades (Fig. 7B). Sub-networks 3 and 4 were generally most prevalent in many host orders, with only 1 of the 2 networks being highly prevalent. Sub-network 1 was only prevalent in the Artiodactyla, suggesting strong host specificity of this microbial consortium.

In support of this finding, the network contained a substantially higher proportion of OTUs with local phylogenetic signal among hosts relative to the other sub-networks (Fig. 7D). Sub-network 2 was only prevalent in 4 mammalian orders: Artiodactyla, Diprotodontia, Pilosa, and Primates. The sub-networks showed distributional shifts among diets, with sub-networks 1 and 2 being most prevalent among herbivores, Sub-network 4 dominating in omnivores, and sub-networks 3 and 4 showing equal prevalence among carnivores (Fig. 7C).

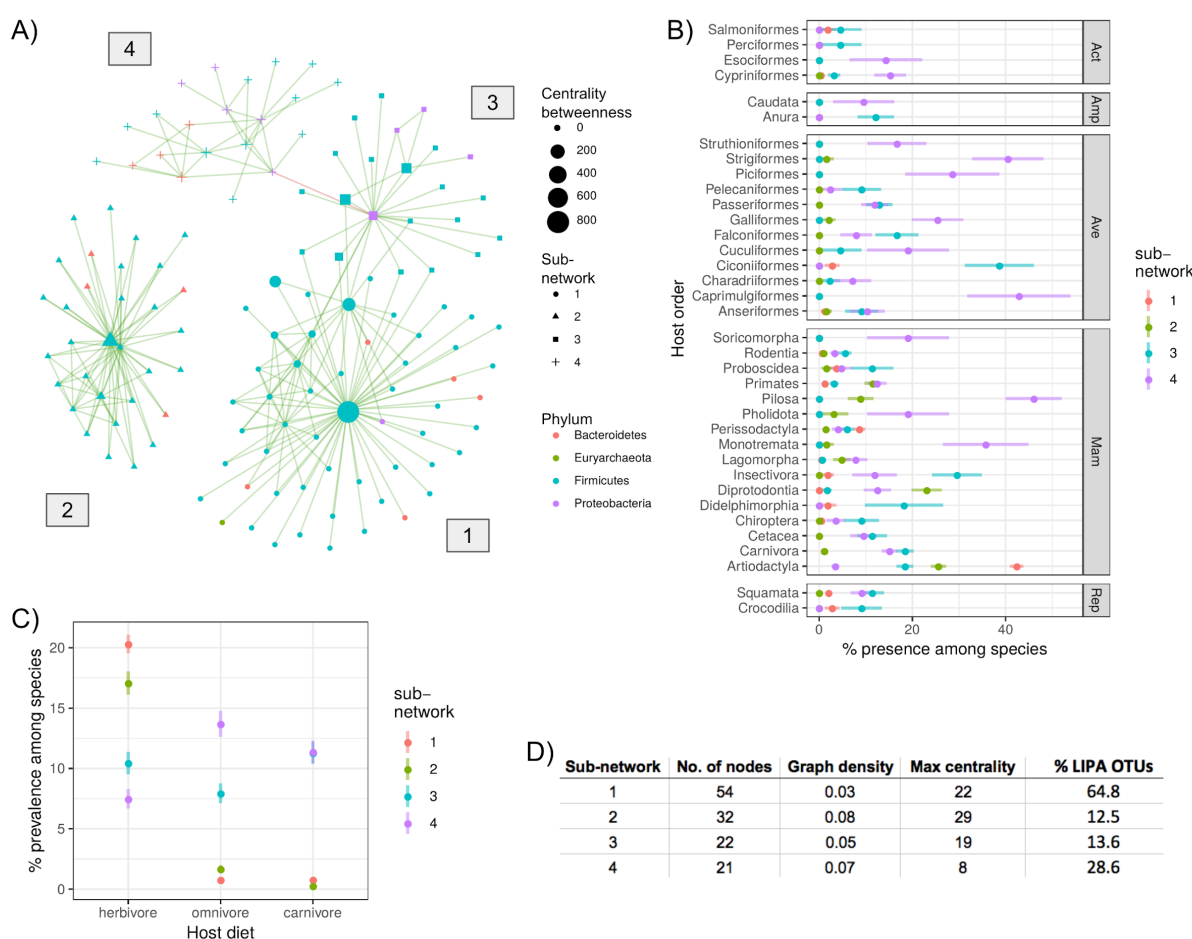


Fig. 7 Co-occurrence analysis revealed 4 sub-networks of co-occurring OTUs that differ in taxonomic composition, network structure, and distribution across host taxonomic orders. A) The network is a presence-absence co-occurrence network, with only significant edges shown (see Methods). Sub-networks are represented by differing node shapes and labels next to each sub-network. “Centrality

betweenness” is a measure of how often the shortest path between two nodes transverses through the focal node. B) The distribution of community presence among samples from each species (% of samples per species) shown for each host taxonomic order, with points representing the mean and line ranges representing +/- the standard error of the mean. The plots are faceted by host taxonomic class (‘Act’ = Actinopterygii; ‘Amp’ = Amphibia; ‘Ave’ = Aves; ‘Mam’ = Mammalia; ‘Rep’ = Reptilia). C) Similar to B) but grouped by host diet. D) A table of sub-network statistics, with “Graph density” defined as “number of edges / total possible edges”, “Max centrality” defined as the max number of shortest paths between any two nodes that cross the focal node, and “% LIPA OTUs” defined as the percent of OTUs with significant local phylogenetic signal (see Fig. 4).

Discussion

While various studies have shown that host diet and phylogeny modulate the animal intestinal microbiome^{5,6}, we have expanded on this previous work by performing a robust assessment of each factor’s effect on a homogeneously generated dataset of highly diverse and predominantly wild animals. Because our dataset consisted of animals from diverse lineages that consume a range of dietary components, we were able to decouple of the effects of host phylogeny and diet on both aggregate diversity metrics and at the individual OTU level. We employed multiple analytical methods to support our findings, and we also directly assessed the sensitivity of our analyses to intra-species microbiome and metadata heterogeneity, which has been found to be non-trivial^{7,14,62,63}. We must acknowledge that we did not have inter-individual replicates for some host species in our dataset, which limited our ability to determine the impact of this factor for certain host clades. Still, our findings suggest that host diet and evolution are strong modulators despite the intra-species variability that we measured. We did not find that habitat or geographic distance explained microbiome diversity, which is consistent with some animal microbiome studies^{6,26}, but clashes with others^{6,22,64}. Possibly, these factors may only

modulate the microbiome of certain host clades, or our dataset is underpowered in regards to testing these potential modulators.

Sparsely distributed and sparsely cultured microbial taxa

Only a couple of very coarsely-resolved taxonomic groups were present in (nearly) all host species (Supplementary Fig. 3). This finding suggests that most microbial clades, especially finely resolved clades, are somewhat constrained to certain host clades. Indeed, we did find beta-diversity to be more constrained at finer host taxonomic levels (Supplementary Fig. 4). The largest exception to this trend was the Clostridiales order, which was found in ~98 % of host species (Supplementary Fig. 3). Many members of Clostridiales generate resistant spores, which may allow for high inter-species or environment-host migration. This process could generate source-sink dynamics, where Clostridiales only transiently pass through specific gut environments, but high migration rates from source hosts, soil, water, etc. continually replenish these ephemeral sink populations. In contrast, our data support true specialization of certain Clostridiales for specific host clades. Specifically, we found that the majority of OTUs displaying a local phylogenetic signal belonged to Clostridiales (Fig. 4). Importantly, these Clostridiales OTUs showed specificity for differing host clades, which have different exposures to potential source communities, and thus the signal of host specificity is unlikely to have resulted from transient populations maintained by high migrations rates. While only two OTUs were significantly modulated by host diet after controlling for phylogeny, one belonged to Clostridiales (Fig. 3), suggesting that specialization to specific host clades (and in some instances, diet) contributed to adaptive speciation in this lineage.

New culturomics techniques are greatly reducing the number of uncultured microbes in the human gut⁶⁵; however, our analysis suggests that microbes from other animals are far less represented (Supplementary Fig. 5). This even applies to Mammalia, which have received the lion's share of focus for gut microbiome studies⁶. Our limited knowledge of gut-inhabiting

microbes of many animals is typified by the CKC4 phylum, which we found to be a relatively abundant phylum in a number of samples (Fig. 1), but the clade has no cultured representatives and is thus poorly characterized⁶⁶. So as with other calls for more studies of wild animal microbiomes^{9,25}, our findings also advocate for more research utilizing both culture-dependent and independent methods to characterize the physiology, ecology, and evolution of vertebrate gut-inhabiting microbes.

Host diet and phylogeny modulate different aspects of gut microbial diversity

While we found both host diet and evolutionary history to significantly explain microbiome diversity, each factor explained differing aspects of that diversity. Diet was a relatively strong predictor of both alpha- and beta-diversity, but the association was strongest with alpha-diversity (Fig. 2; Fig. 3; Supplementary Fig. 10). However, at the single OTU level, the distribution of only 2 OTUs was significantly explained by diet (Fig. 3). In contrast, host phylogeny was only a significant predictor of differences in microbiome composition (Fig. 2B; unweighted Unifrac). While at the OTU level, 121 OTUs displayed a significant phylogenetic signal after first accounting for diet (Fig. 4).

Taken together, these results support a scenario in which diet mediates community assembly through environmental filtering predominantly at the level of functional guilds (eg., cellulolytic consortia), while host evolutionary history mainly dictates the prevalence of specific OTUs (ie., heritable microbial taxa). By modulating the distribution of functional guilds, host diet would expand or contract alpha-diversity depending on the diversity of guilds selected for. If these guilds are somewhat labile in their taxonomic composition due to functional redundancy, then the diversity of the functional guild would be dictated by diet, but taxonomic composition could vary among hosts that have the same specific diets. To illustrate, consider that a consortium degrading cellulose or other recalcitrant plant polymers in a herbivorous diet would likely require a larger assemblage of primary and secondary degraders versus a less

recalcitrant meat-based diet. While microbial function can only be indirectly inferred by 16S rRNA sequencing, metagenomics studies support this concept that diet is strongly selective of microbial function, at least in the mammalian gut^{26,67}.

Interestingly, the recent meta-analysis on mammal gut microbiomes by Nishida and Ochman showed that phylogenetic signal is strongest at finer taxonomic levels, which coincides with our observations that host phylogeny mainly dictates that distribution of specific OTUs²². Our findings also correspond with studies of microbial heritability in humans, in which the abundances of only certain specific taxonomic groups have been consistently found to be dictated by host genetics across multiple independent studies¹². Moreover, we observed significant phylogenetic signal for OTUs belonging to all three clades identified by Goodrich and colleagues to be consistently heritable in humans: *Methanobrevibacter*, Christensenellaceae, and *Blautia*. No OTUs in our study showed significant phylogenetic signal for humans, and only a few OTUs were associated with any of the 10 primate species in our study. These finding indicate that the effects of host evolutionary history are stronger outside of this clade. This finding could help to explain why relatively large cohorts are necessary to identify heritable microbial taxa in humans¹². Alternatively, intra-species diversity is greater in large human cohort studies compared to what we measured in this work, and this higher intra-species variance may obscure signals of coevolution.

Both tests of phylogenetic signal at the OTU-level and tests of co-speciation support the hypothesis that host evolutionary history more strongly determines microbial diversity among mammals versus non-mammals (Fig. 4; Fig. 5; Supplementary Fig. 12). Multiple non-exclusive mechanisms could explain these findings. First, the gut microbiomes of non-mammal species may contain more transient microbes from environmental sources. This may be especially true of the Actinopterygii, given that the surrounding environment is thought to be one of the primary mechanisms of microbiota acquisition for fish⁶⁸. Second, when considering the evolution of digestive physiology, mammals have developed highly complex digestive systems in relation to

most non-mammalian species in our study⁶⁹. This is especially true for ruminants, which have developed complex multi-chambered forestomachs and a system of regurgitation and mastication in order to efficiently degrade complex plant polymers via enhanced microbial fermentation. We observed the strongest cophylogeny signal for ruminants, especially among cattle (Bovidae), which have arguably the most complicated digestive physiology⁷⁰. Interestingly, Nishida and Ochman found that rates of microbiome divergence have accelerated in Cetartiodactyla²², which may be the result of evolving the complex forestomach and other digestive traits specific to this clade. Indeed, an increased microbial yield and fiber digestion are thought to represent important selective advantages in foregut fermenters⁷⁰. Third, vertical transmission for microbial taxa from parent to offspring may also differ between mammals and non-mammals. Mammalian microbiome acquisition occurs during the birthing process and is further developed through nursing, maternal contact, and social group affiliation⁷¹. Much less is known about how non-mammals acquire their gut microbiomes, but at least for some species, coprophagy, eating soil in the nest, and eating regurgitated food are important modes of vertical transmission⁶. Still, mixed-mode transmission (vertical transmission and transmission from unrelated hosts or the environment) is considered to be more prevalent among non-mammals⁷².

The role of microbe-microbe interactions in community assembly

Our ecophylogenetic and co-occurrence tests further resolved differences in microbial community assembly among host species. The majority of microbial communities showed significant phylogenetic clustering (Fig. 6), which supports our hypothesis that diet and host phylogeny impose environmental filtering on specific functional guilds and/or certain taxa. Interestingly, members of Artiodactyla showed little signal of phylogenetic clustering, and in some cases, we observed significant phylogenetic evenness (Fig. 6). This is consistent with a hypothesis that the effects of environmental filtering are limited among Artiodactyla compared to processes selecting for unrelated taxa. Similar observations were recently reported by Gaulke

and colleagues, who found less signal of phylogenetic clustering among Artiodactyla relative to other mammalian clades⁶¹. The high water content of ruminant feces may help to explain this lack of phylogenetic clustering⁷³, given that high water content in soil has been shown to reduce phylogenetic clustering relative to dry soils^{74,75}. Another non-exclusive explanatory factor may be that the refractory composition of the ruminant diet requires functional guilds composed of distantly related taxa, resulting in phylogenetic evenness. In support of this hypothesis, Sub-network 1 in our co-occurrence analysis showed high specificity to Artiodactyla relative to the other sub-networks (Fig. 7), and it is the only one to contain OTUs from all 4 phyla present among the sub-networks (Supplementary Fig. 14).

The “hub” OTUs present in 3 of the 4 sub-networks suggests that keystone species (OTUs) contribute to community assembly (Fig. 7). Interestingly, the maximum betweenness score in each sub-network directly corresponded with the prevalence of the sub-networks in herbivores, while the sub-network with the lowest centrality scores (Sub-network 4) was the most prevalent among omnivores and carnivores (Fig. 7). Therefore, it appears that the herbivorous diet selects for co-occurring consortia containing keystone species. These keystone species may form the foundation in which functional guilds are based. The other members of each sub-network would thus represent the taxonomically stable portion of the functional guild, while functionally redundant taxa in the guild would not show a stable co-occurrence pattern. In support of this concept, the hub OTUs of sub-networks 1 and 2 both belong to the Ruminococcaceae (Supplementary Fig. 14), and this clade contains members that can play a major role in plant cell wall breakdown into substrates utilized by other members of the consortium⁷⁶. Indeed, Ruminococcaceae taxa have previously been identified as keystone species in human and ruminant gut communities⁷⁶. The gain or loss of these putative keystone species in hosts may cause relatively large, diet-dependent health and fitness effects on the host.

Conclusions

Our findings help to resolve the major modulators of intestinal microbiome diversity, which have not been well-studied in wild animals, especially non-mammalian species. We posit that diet primarily selects for functional guilds while host evolutionary history mainly determines the prevalence of specific microbial OTUs. A metagenomics-based analysis on this dataset will help to resolve how diet and host phylogeny modulate microbial function versus taxonomy. The modulating effect of host evolutionary history was most pronounced in mammals, especially for Artiodactyla. In general, our findings suggest that microbial community assembly in the Artiodactyla clade differs substantially from other mammalian clades, which may be the result of the complex digestive physiology that has evolved in ruminants. The putative keystone species identified in our co-occurrence analysis may be of special interest for future work determining how dietary changes can modulate the animal gut microbiome, such as in the context of captivity or climate change.

Author contributions

GR and AF created the study concept. GR, CW, and GS performed the sample collection and metadata compilation. GR and NS performed the laboratory work. NY and WW performed the data analysis. NY, GR, RL, and AF wrote the manuscript.

Acknowledgements

This study was supported by the Austrian Science Fund (FWF) research projects P23900 granted to Andreas H. Farnleitner and P22032 granted to Georg H. Reischer. Further support came from the Science Call 2015 "Ressource und Lebensgrundlage Wasser" Project SC15-016 funded by the Niederösterreichische Forschungs- und Bildungsgesellschaft (NFB).

This work was supported by a David and Lucile Packard Foundation Fellowship (to REL) and the Max Planck Society.

We would like to thank the following collaborators for their huge efforts in sample and data collection: Mario Baldi, School of Veterinary Medicine, Universidad Nacional de Costa Rica; Wolfgang Vogl and Frank Radon, Konrad Lorenz Institute of Ethology and Biological Station Illmitz; Endre Sós and Viktor Molnár, Budapest Zoo; Ulrike Streicher, Conservation and Wildlife Management Consultant, Vietnam; Katharina Mahr, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine Vienna and Flinders University Adelaide, South Australia; Peggy Rismiller, Pelican Lagoon Research Centre, Australia; Rob Deaville, Institute of Zoology, Zoological Society of London; Alex Lécu, Muséum National d'Histoire Naturelle and Paris Zoo; Danny Govender and Emily Lane, South African National Parks, Sanparks; Fritz Reimoser, Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna; Anna Kübber-Heiss and Team, Pathology, Research Institute of Wildlife Ecology, University of Veterinary Medicine Vienna; Nikolaus Eisank, Nationalpark Hohe Tauern, Kärnten; Attila Hettyey and Yoshan Moodley, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine Vienna; Mansour El-Matbouli and Oskar Schachner, Clinical Unit of Fish Medicine, University of Veterinary Medicine; Barbara Richter, Institute of Pathology and Forensic Veterinary Medicine, University of Veterinary Medicine Vienna; Hanna Vielgrader and Zoovet Team, Schönbrunn Zoo; Reinhard Pichler, Herberstein Zoo. We explicitly thank the Freek Venter of South African National Parks and the National Zoological Gardens of South Africa for granting access to their Parks for sample collection. We also thank Carolin Kolmeder for her helpful discussions on this project.

Competing interests

No competing interests declared.

680

681 **References**

- 682 1. Suzuki, T. A. Links between Natural Variation in the Microbiome and Host Fitness in Wild
683 Mammals. *Integr. Comp. Biol.* **57**, 756–769 (2017).
- 684 2. McFall-Ngai, M. *et al.* Animals in a bacterial world, a new imperative for the life sciences.
685 *Proc. Natl. Acad. Sci. U. S. A.* **110**, 3229–3236 (2013).
- 686 3. Rosenberg, E. & Zilber-Rosenberg, I. The hologenome concept of evolution after 10 years.
687 *Microbiome* **6**, 78 (2018).
- 688 4. Bahrndorff, S., Alemu, T., Alemneh, T. & Lund Nielsen, J. The Microbiome of Animals:
689 Implications for Conservation Biology. *Int. J. Genomics Proteomics* **2016**, 5304028 (2016).
- 690 5. Ley, R. E. *et al.* Evolution of Mammals and Their Gut Microbes. *Science* **320**, 1647–1651
691 (2008).
- 692 6. Colston, T. J. & Jackson, C. R. Microbiome evolution along divergent branches of the
693 vertebrate tree of life: what is known and unknown. *Mol. Ecol.* **25**, 3776–3800 (2016).
- 694 7. Maurice, C. F. *et al.* Marked seasonal variation in the wild mouse gut microbiota. *ISME J.* **9**,
695 2423–2434 (2015).
- 696 8. Groussin, M. *et al.* Unraveling the processes shaping mammalian gut microbiomes over
697 evolutionary time. *Nat. Commun.* **8**, 14319 (2017).
- 698 9. Pascoe, E. L., Hauffe, H. C., Marchesi, J. R. & Perkins, S. E. Network analysis of gut
699 microbiota literature: an overview of the research landscape in non-human animal studies.
700 *ISME J.* **11**, 2644–2651 (2017).
- 701 10. Faith, J. J., McNulty, N. P., Rey, F. E. & Gordon, J. I. Predicting a human gut microbiota's
702 response to diet in gnotobiotic mice. *Science* **333**, 101–104 (2011).
- 703 11. David, L. A. *et al.* Diet rapidly and reproducibly alters the human gut microbiome. *Nature*
704 **505**, 559–563 (2014).

12. Goodrich, J. K., Davenport, E. R., Clark, A. G. & Ley, R. E. The Relationship Between the Human Genome and Microbiome Comes into View. *Annu. Rev. Genet.* (2017).
doi:10.1146/annurev-genet-110711-155532
13. Sasson, G. *et al.* Heritable Bovine Rumen Bacteria Are Phylogenetically Related and Correlated with the Cow's Capacity To Harvest Energy from Its Feed. *MBio* **8**, (2017).
14. Hicks, A. L. *et al.* Gut microbiomes of wild great apes fluctuate seasonally in response to diet. *Nat. Commun.* **9**, 1786 (2018).
15. Carmody, R. N. *et al.* Diet dominates host genotype in shaping the murine gut microbiota. *Cell Host Microbe* **17**, 72–84 (2015).
16. Delsuc, F. *et al.* Convergence of gut microbiomes in myrmecophagous mammals. *Mol. Ecol.* **23**, 1301–1317 (2014).
17. Ochman, H. *et al.* Evolutionary Relationships of Wild Hominids Recapitulated by Gut Microbial Communities. *PLoS Biol.* **8**, e1000546 (2010).
18. Brooks, A. W., Kohl, K. D., Brucker, R. M., van Opstal, E. J. & Bordenstein, S. R. Phylosymbiosis: Relationships and Functional Effects of Microbial Communities across Host Evolutionary History. *PLoS Biol.* **14**, e2000225 (2016).
19. Ross, A. A., Müller, K. M., Weese, J. S. & Neufeld, J. D. Comprehensive skin microbiome analysis reveals the uniqueness of human skin and evidence for phylosymbiosis within the class Mammalia. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E5786–E5795 (2018).
20. Degnan, P. H. *et al.* Factors associated with the diversification of the gut microbial communities within chimpanzees from Gombe National Park. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 13034–13039 (2012).
21. Martinson, V. G., Douglas, A. E. & Jaenike, J. Community structure of the gut microbiota in sympatric species of wild *Drosophila*. *Ecol. Lett.* **20**, 629–639 (2017).
22. Nishida, A. H. & Ochman, H. Rates of gut microbiome divergence in mammals. *Mol. Ecol.* **27**, 1884–1897 (2018).

23. Clayton, J. B. *et al.* Captivity humanizes the primate microbiome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10376–10381 (2016).
24. McKenzie, V. J. *et al.* The Effects of Captivity on the Mammalian Gut Microbiome. *Integr. Comp. Biol.* **57**, 690–704 (2017).
25. Hird, S. M. Evolutionary Biology Needs Wild Microbiomes. *Front. Microbiol.* **8**, 689 (2017).
26. Henderson, G. *et al.* Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* **5**, 14567 (2015).
27. Reischer, G. H. *et al.* Performance characteristics of qPCR assays targeting human- and ruminant-associated bacteroidetes for microbial source tracking across sixteen countries on six continents. *Environ. Sci. Technol.* **47**, 8548–8556 (2013).
28. Caporaso, J. G. *et al.* Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 1**, 4516–4522 (2011).
29. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* **13**, 581–583 (2016).
30. 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).
31. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* **41**, D590–6 (2013).
32. McMurdie, P. J. & Holmes, S. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013).
33. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490 (2010).
34. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
35. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).

36. Paterno, G. B., Penone, C. & Werner, G. D. A. sensiPhy : An r -package for sensitivity analysis in phylogenetic comparative methods. *Methods Ecol. Evol.* **9**, 1461–1467 (2018).
37. R Core Team. R: A Language and Environment for Statistical Computing. (2016).
38. Lang, M., Bischl, B. & Surmann, D. batchtools: Tools for R to work on batch systems. *JOSS* **2**, 135 (2017).
39. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
40. Orme, D. & Others. The caper package: comparative analysis of phylogenetics and evolution in R. *R package version* **5**, 1–36 (2013).
41. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–5 (2016).
42. Pedersen, T. L. tidygraph: A Tidy API for Graph Manipulation. (2018).
43. Pedersen, T. L. ggraph: An Implementation of Grammar of Graphics for Graphs and Networks. (2018).
44. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
45. Goslee, S. C. & Urban, D. L. The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *Jstatssoft* **22**, 1–19 (2007).
46. Hutchinson, M. C., Cagua, E. F., Balbuena, J. A., Stouffer, D. B. & Poisot, T. paco: implementing Procrustean Approach to Cophylogeny in R. *Methods Ecol. Evol.* **8**, 932–940 (2017).
47. Keck, F., Rimet, F., Bouchez, A. & Franc, A. phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol. Evol.* **6**, 2774–2780 (2016).
48. Griffith, D., Veech, J. & Marsh, C. cooccur: Probabilistic Species Co-Occurrence Analysis in R. *Journal of Statistical Software, Code Snippets* **69**, 1–17 (2016).
49. Pons, P. & Latapy, M. Computing Communities in Large Networks Using Random Walks. in

Computer and Information Sciences - ISCIS 2005 284–293 (Springer Berlin Heidelberg, 2005).

50. Sullam, K. E. *et al.* Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Mol. Ecol.* **21**, 3363–3378 (2012).
51. Keenan, S. W., Engel, A. S. & Elsey, R. M. The alligator gut microbiome and implications for archosaur symbioses. *Sci. Rep.* **3**, 2877 (2013).
52. Costa, M. C. & Weese, J. S. The equine intestinal microbiome. *Anim. Health Res. Rev.* **13**, 121–128 (2012).
53. Petri, R. M. *et al.* Characterization of the core rumen microbiome in cattle during transition from forage to concentrate as well as during and after an acidotic challenge. *PLoS One* **8**, e83424 (2013).
54. Clayton, J. B. *et al.* The gut microbiome of nonhuman primates: Lessons in ecology and evolution. *Am. J. Primatol.* e22867 (2018).
55. Liu, Y. *et al.* Influence of Endogenous and Exogenous Estrogenic Endocrine on Intestinal Microbiota in Zebrafish. *PLoS One* **11**, e0163895 (2016).
56. Gainza, O., Ramírez, C., Ramos, A. S. & Romero, J. Intestinal Microbiota of White Shrimp *Penaeus vannamei* Under Intensive Cultivation Conditions in Ecuador. *Microb. Ecol.* **75**, 562–568 (2018).
57. Goodrich, J. K. *et al.* Human genetics shape the gut microbiome. *Cell* **159**, 789–799 (2014).
58. Moran, N. A. & Sloan, D. B. The Hologenome Concept: Helpful or Hollow? *PLoS Biol.* **13**, e1002311 (2015).
59. Horner-Devine, M. C. & Bohannan, B. J. M. Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**, S100–8 (2006).
60. Kembel, S. W. *et al.* Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
61. Gaulke, C. A. *et al.* Ecophylogenetics Clarifies the Evolutionary Association between

Mammals and Their Gut Microbiota. *MBio* **9**, (2018).

62. Davenport, E. R. *et al.* Seasonal variation in human gut microbiome composition. *PLoS One* **9**, e90731 (2014).

63. Trosvik, P. *et al.* Multilevel social structure and diet shape the gut microbiota of the gelada monkey, the only grazing primate. *Microbiome* **6**, 84 (2018).

64. Moeller, A. H. *et al.* Dispersal limitation promotes the diversification of the mammalian gut microbiota. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 13768–13773 (2017).

65. Lagier, J.-C. *et al.* Culturing the human microbiota and culturomics. *Nat. Rev. Microbiol.* **16**, 540–550 (2018).

66. Youssef, N. H., Couger, M. B., McCully, A. L., Criado, A. E. G. & Elshahed, M. S. Assessing the global phylum level diversity within the bacterial domain: A review. *J. Advert. Res.* **6**, 269–282 (2015).

67. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–974 (2011).

68. Nayak, S. K. Role of gastrointestinal microbiota in fish. *Aquac. Res.* **41**, 1553–1573 (2010).

69. Edward Stevens, C. & Hume, I. D. *Comparative Physiology of the Vertebrate Digestive System*. (Cambridge University Press, 2004).

70. Clauss, M., Hume, I. D. & Hummel, J. Evolutionary adaptations of ruminants and their potential relevance for modern production systems. *Animal* **4**, 979–992 (2010).

71. Gilbert, S. F. A holobiont birth narrative: the epigenetic transmission of the human microbiome. *Front. Genet.* **5**, 282 (2014).

72. Ebert, D. The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission. *Annu. Rev. Ecol. Evol. Syst.* **44**, 623–643 (2013).

73. Clauss, M., Nunn, C., Fritz, J. & Hummel, J. Evidence for a tradeoff between retention time and chewing efficiency in large mammalian herbivores. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **154**, 376–382 (2009).

74. Placella, S. A., Brodie, E. L. & Firestone, M. K. Rainfall-induced carbon dioxide pulses result from sequential resuscitation of phylogenetically clustered microbial groups. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 10931–10936 (2012).
75. Barnard, R. L., Osborne, C. A. & Firestone, M. K. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *ISME J.* **7**, 2229–2241 (2013).
76. Ze, X., Le Mougen, F., Duncan, S. H., Louis, P. & Flint, H. J. Some are more equal than others: the role of ‘keystone’ species in the degradation of recalcitrant substrates. *Gut Microbes* **4**, 236–240 (2013).