AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

Check for updates

Allometry and Ecology of the Bilaterian Gut Microbiome

Scott Sherrill-Mix,[a] Kevin McCormick,[a] Abigail Lauder,[a] Aubrey Bailey,[a] Laurie Zimmerman,[a] Yingying Li,[b] Jean-Bosco N. Django,[c] Paco Bertolani,[d] Christelle Colin,[e] John A. Hart,[f] Terese B. Hart,[f] Alexander V. Georgiev,[g,h] Crickette M. Sanz,[i] David B. Morgan,[j] Rebeca Atencia,[k] Debby Cox,[k] Martin N. Muller,[l] Volker Sommer,[m] Alexander K. Piel,[n] Fiona A. Stewart,[n] Sheri Speede,[o] Joe Roman,[p] Gary Wu,[q] Josh Taylor,[r] Rudolf Bohm,[r] Heather M. Rose,[s] John Carlson,[t] Deus Mjungu,[u] Paul Schmidt,[v] Celeste Gaughan,[w] Joyslin I. Bushman,[w] Ella Schmidt,[w] Kyle Bittinger,[w] Ronald G. Collman,[b] Beatrice H. Hahn,[b] Frederic D. Bushman[a]

[a]Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[b]Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[c]Department of Ecology and Management of Plant and Animal Resources, Faculty of Sciences, University of Kisangani, Kisangani, Democratic Republic of the Congo

[d]Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, United Kingdom

[e]Projet Primates France, Centre de Conservation pour Chimpanzés, Faranah, Republic of Guinea

[f]Lukuru Wildlife Research Foundation, Tshuapa-Lomami-Lualaba Project, Kinshasa, Democratic Republic of the Congo

[g]Department of Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

[h]School of Biological Sciences, Bangor University, Bangor, United Kingdom

[i]Department of Anthropology, Washington University in St. Louis, St. Louis, Missouri, USA

[j]Lester E. Fisher Center for the Study and Conservation of Apes, Lincoln Park Zoo, Chicago, Illinois, USA

[k]Tchimpounga Chimpanzee Rehabilitation Center, Jane Goodall Institute, Pointe Noire, Republic of Congo

[l]Department of Anthropology, University of New Mexico, Albuquerque, New Mexico, USA

[m]Department of Anthropology, University College London, London, United Kingdom

[n]Department of Natural Sciences, Liverpool John Moores University, Liverpool, United Kingdom

[o]Sanaga-Yong Chimpanzee Rescue Center, IDA-Africa, Portland, Oregon, USA

[p]Gund Institute for Environment, Rubenstein School for Environment and Natural Resources, University of Vermont, Burlington, Vermont, USA

[q]Division of Gastroenterology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[r]Tulane National Primate Research Center, Tulane University Health Science Center, Covington, Louisiana, USA

[s]Thomas Jefferson University, Philadelphia, Pennsylvania, USA

[t]Southeast Fisheries Science Center, National Oceanic and Atmospheric Administration Fisheries Service, Panama City, Florida, USA

[u]Gombe Stream Research Centre, The Jane Goodall Institute, Kigoma, Tanzania

[v]Department of Biology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[w]Division of Gastroenterology, Hepatology, and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA

Address correspondence to Frederic D. Bushman, bushman@pennmedicine.upenn.edu.

ABSTRACT Classical ecology provides principles for construction and function of biological communities, but to what extent these apply to the animal-associated microbiota is just beginning to be assessed. Here, we investigated the influence of several well-known ecological principles on animal-associated microbiota by characterizing gut microbial specimens from bilaterally symmetrical animals (*Bilateria*) ranging from flies to whales. A rigorously vetted sample set containing 265 specimens from 64 species was assembled. Bacterial lineages were characterized by 16S rRNA gene sequencing. Previously published samples were also compared, allowing analysis of over 1,098 samples in total. A restricted number of bacterial phyla was found to account for the great majority of gut colonists. Gut microbial composition was associated with host phylogeny and diet. We identified numerous gut bacterial 16S rRNA gene sequences that diverged deeply from previously studied taxa,

identifying opportunities to discover new bacterial types. The number of bacterial lineages per gut sample was positively associated with animal mass, paralleling known species-area relationships from island biogeography and implicating body size as a determinant of community stability and niche complexity. Samples from larger animals harbored greater numbers of anaerobic communities, specifying a mechanism for generating more-complex microbial environments. Predictions for species/abundance relationships from models of neutral colonization did not match the data set, pointing to alternative mechanisms such as selection of specific colonists by environmental niche. Taken together, the data suggest that niche complexity increases with gut size and that niche selection forces dominate gut community construction.

IMPORTANCE   The intestinal microbiome of animals is essential for health, contributing to digestion of foods, proper immune development, inhibition of pathogen colonization, and catabolism of xenobiotic compounds. How these communities assemble and persist is just beginning to be investigated. Here we interrogated a set of gut samples from a wide range of animals to investigate the roles of selection and random processes in microbial community construction. We show that the numbers of bacterial species increased with the weight of host organisms, paralleling findings from studies of island biogeography. Communities in larger organisms tended to be more anaerobic, suggesting one mechanism for niche diversification. Nonselective processes enable specific predictions for community structure, but our samples did not match the predictions of the neutral model. Thus, these findings highlight the importance of niche selection in community construction and suggest mechanisms of niche diversification.

Classical ecology provides extensive theories for investigating community dynamics that are beginning to be applied to sequence-based studies of microbes. For example, species-area relationships (1) have been called "one of community ecology's few genuine laws" (2, 3). According to this idea, for islands located at the same distance from a mainland source, the larger islands accumulate the greater number of species, because (i) the larger population sizes that are possible only on larger islands reduce the chances of stochastic fluctuations leading to extinction and (ii) larger islands have a greater number of niches, supporting more different types of residents. Species-area relationships have been investigated previously for the gut microbiota but with inconsistent results. One study using published values from 16S rRNA gene sequencing studies of rRNA gene segments led to the conclusion that the mammalian gut microbiota did not show greater richness in intestines of larger organisms (4). Another study, which used a DNA fingerprinting method to characterize bacterial gut communities, concluded that richness indeed increased for animals of larger sizes (5).

Another ecological framework, Hubbell's neutral model, makes specific predictions for neutral colonization processes that can then be compared to data to interrogate assembly mechanisms. Hubbell's neutral theory models construction of a population via migration from a source population using information on birth and death rates, migration rates, and abundances while assuming all organisms are equally fit colonists (6). Neutral theory provides quantitative predictions for the relative abundances of community members, which can then be compared to empirical data to assess whether neutral theory provides a compelling explanation of observations (7). In the absence of a compelling fit, other processes, typically involving niche selection, can be hypothesized to be more likely. In a recent study, human microbiome specimens were tested and usually found to diverge from the predictions of neutral theory (8), but this has not been investigated for a larger number of animals.

We thus sought to investigate microbial community structure as a window into assembly processes and so studied a wide set of members of the *Bilateria*—the bilaterally symmetrical clade of animals that includes protostomes and deuterostomes.

We focused on intestinal bacterial communities to access a consistent sample type and to interrogate an organ system relevant to human and veterinary medicine. In all, we sampled gut contents from 10 classes of animals, representing 64 species. We purified DNA from feces or dissected gut tissue from species that ranged in size from bedbug (0.009 g) to right whale (72,000,000 g), spanning almost 10^9 g in body size. In this analysis, we assumed that overall organismic size is a viable surrogate for gut size, though it is clearly an approximation (9). To analyze bacterial communities, we amplified and sequenced 16S rRNA gene segments, allowing quantification of the richness and diversity of bacterial communities. Samples were rigorously vetted to control for artifacts due to low biomass in starting samples (10, 11). We also compared data from several published studies (12–19), yielding 1,098 total samples.

We found that gut microbial community structure was associated with host phylogeny. Species-area analysis showed a positive correlation between organismic size and gut microbial species richness. Further analysis showed that community structure did not fit predictions of neutral assembly models, with the total data implying a prominent role for niche selection in assembly of gut microbial communities.

RESULTS

Experimental strategy. To compare gut microbial populations among bilaterians of different sizes, we collected 444 specimens of feces or dissected intestines (see Table S1 in the supplemental material). Because some samples contained host tissue, we used 16S rRNA gene primers for amplification and sequencing, so that our sequencing effort queried bacterial colonists and not sequences from the host. Amplification was carried out with primers targeting the V1V2 region of the 16S rRNA gene, which has been used previously for phylogenetic comparisons of gut microbiota (17, 20).

Low amounts of starting template are problematic because reagent contamination can become a significant proportion of the total (10, 11). We thus carried out 16S quantitative PCR (qPCR) analysis of all samples studied and analyzed only samples where at least 1,000 16S rRNA gene copies could be introduced into PCRs for library preparation. Samples with fewer than 1,000 reads after quality filtering were also excluded. All DNA isolation and sequencing steps were carried out side by side with negative controls, allowing us to query contamination from reagents, dust, and other sources (see Fig. S1 in the supplemental material). Samples with extensive representation of lineages found in negative controls typically contained low starting copy numbers of bacterial DNA and were among those excluded by our qPCR criteria. This yielded 265 samples from 10 classes, representing 64 species of *Bilateria* (Table S1) (Fig. S1A).

We expected that bacteria in little-studied bilaterian species would be disproportionately difficult to assign to genera and species, so bacterial lineages were characterized as operational taxonomic units (OTUs) based on sequence and not as phylogenetic assignments based on database references. Sequence reads were condensed into clusters of 97% identity using UCLUST (Table S2). Phylogenetic placement was assessed where possible by comparison to the Greengenes database (Fig. 1A). On average, each specimen was characterized by 104,998 sequence reads (Fig. S1B and C), yielding on average 3,551 OTUs. Numbers ranged from 21 OTUs in a sample (boxelder bug) to 26,991 OTUs (Jersey cow).

We also collected additional data from previously published studies on the microbiome of insects (12), birds (13), primates (14), myrmecophagous mammals (15), fish (16), and other mammals (17–19). Combining newly generated data with published data yielded 1,098 specimens for analysis.

Bacterial communities in the *Bilateria*. Figure 1 summarizes the samples newly studied here and their bacterial communities (similar summaries for the published data sets are presented in Fig. S2A to J). Overall, 64 species were queried, which included broad sampling of both the protostomes and deuterostomes. The predominant bacterial phyla are summarized in the heat maps in Fig. 1A, organized by the host phylogeny (a detailed summary of sample phylogeny is presented in Fig. S1A). Among
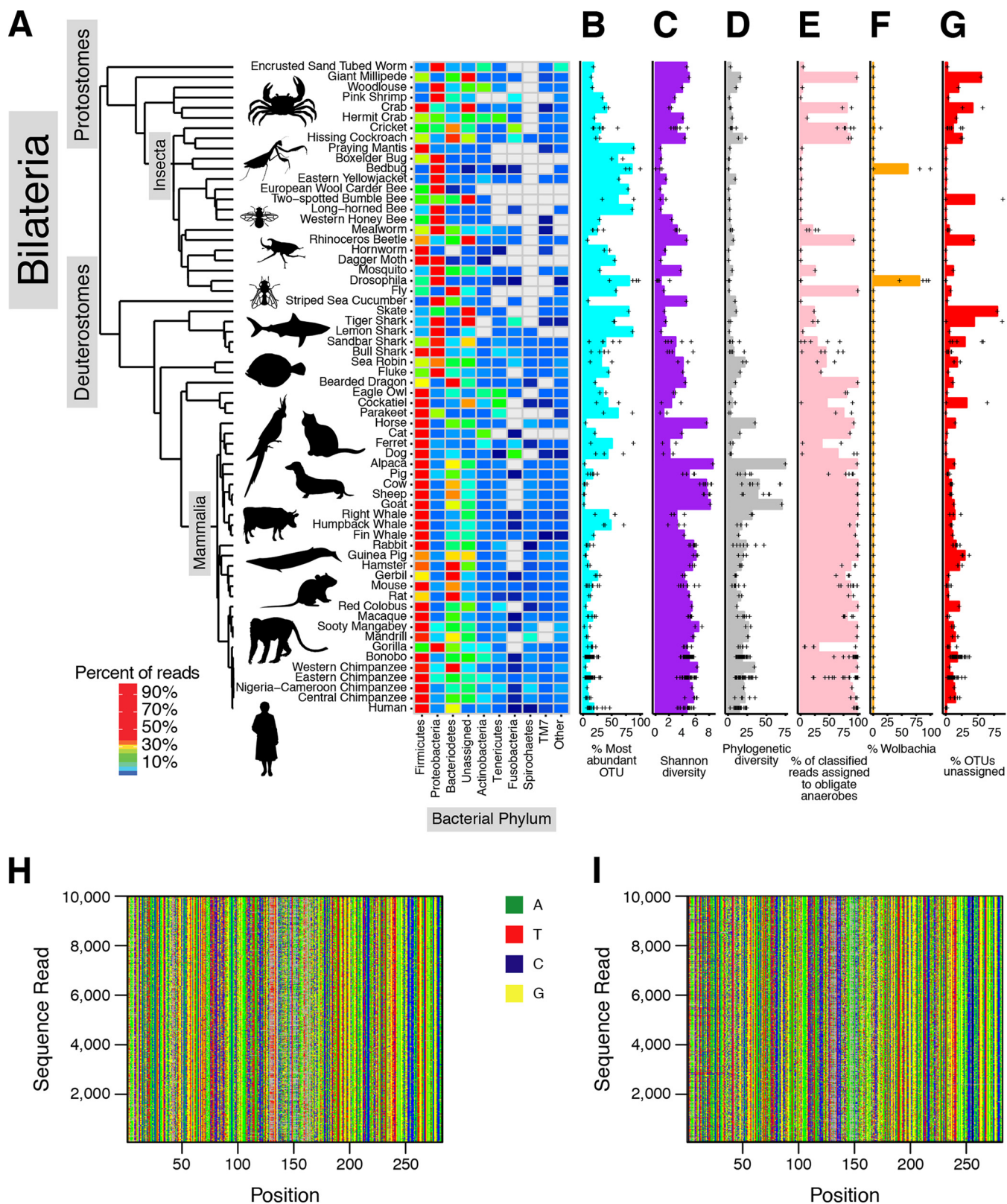
Sherrill-Mix et al.

mBio®

FIG 1 Gut microbiota of the *Bilateria*. The figure summarizes several forms of analysis of the 16S rRNA gene tag sequence data. Host species are indicated by their common names—formal genus and species designations are in Table S1. (A) The proportional abundance of the most abundant bacterial phyla averaged by species. (B) Bar graphs summarizing the percentages of reads assigned to the proportionally most abundant OTU in each species. The values were averaged for each species (bar), and values for individual samples are indicated by plus (+) symbols (the graph format repeats in panels C to G). (C) Bar graph summarizing Shannon diversity for the microbial communities of the various species. (D) Bar graph summarizing phylogenetic diversity for the microbial communities of the various species. (E) Bar graph summarizing the proportion of annotated obligate anaerobic bacteria for the various species. (F) Bar graph summarizing the

(Continued on next page)

the vertebrates, *Mammalia* were dominated by *Firmicutes* and *Bacteroides*, as has been reported in many studies. *Proteobacteria* and *Actinobacteria* were present to lesser extents. Birds showed high levels of colonization with *Firmicutes* and *Tenericutes*. Fish and sharks, in contrast, consistently showed high levels of colonization with levels of *Proteobacteria*. Among invertebrates, colonization by *Proteobacteria* typically predominated.

In our data, nine phyla accounted for 87.7% of the data—in order of abundance, these were *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Actinobacteria*, *Tenericutes*, *Spirochaetes*, *Fusobacteria*, *Cyanobacteria*, and TM7 (Fig. S1C). Twelve percent of sequences remained unassigned by comparison to the Greengenes database, emphasizing the frequency of unstudied microbial lineages in the bilaterian microbiome. These reads were most commonly found in little-studied host species, such as crab (an average of 42% of reads were not assigned), giant millipede (55%), and skate (80%). In contrast, for the well-studied case of the human microbiome (21, 22), an average of only 0.7% of reads were not classified.

Shared and unique features of bilaterian gut communities. We began to assess community structure by asking what proportion of the full sample was comprised of the most predominant OTU (Fig. 1B). As is discussed below, assessing the rank/abundance structure of communities provides potential insight into the process by which they were generated. A broad range of values were seen, with highly diverse communities in primates and ungulates having maximum OTU abundance ranging from only 1.6% (goat) to 18.1% (human), while many sharks and rays (3 of 5 species) and insects (10 of 17 species) showed 50% or more of the communities comprised of the most predominant OTU. The Shannon diversity index showed reciprocal behavior, as expected (Fig. 1C). Samples were further compared using phylogenetic diversity (Fig. 1D), which measures biodiversity as branch length on a phylogenetic tree. Here a strong separation was evident between protostomes, which showed low phylogenetic diversity, and deuterostomes, for which phylogenetic diversity was substantially higher (Wilcoxon rank sum $P < 10^{-6}$). Related patterns could be seen in the published data sets analyzed in Fig. S2A to J, with deuterostomes typically showing higher levels of phylogenetic diversity.

Given these survey data, we could also compare levels of oxygen utilization by gut bacterial communities. Literature data were used to annotate the oxygen utilization and sensitivity of each bacterial species; values were first pooled for bacteria in a community and then for individuals within a host species, to generate an overall value for each host species (Fig. 1E; see also Fig. S3A and B). A trend was observed in which the mammals tended to have a larger fraction of obligate anaerobes in their guts whereas the smaller insects typically showed a smaller fraction of anaerobes. We return to this point below.

Cells of many invertebrates are known to be potentially colonized by *Wolbachia*, a Gram-negative *Alphaproteobacteria* species of the *Rickettsiaceae* family. A previous sequence survey identified *Wolbachia* in multiple insect species (12) (Fig. S2C). Our sample set offered the opportunity to assess colonization broadly among the *Bilateria* (Fig. 1F). We found colonization in *Drosophila*, as has been well described, and also colonization in bed bugs and crickets. We did not detect *Wolbachia* outside the class *Insecta*.

Of particular interest is the issue of how frequently unstudied lineages were encountered. For example, recent studies of lineages from mouths of dolphins revealed unusual deep-branching lineages in host-associated communities (23). To begin to probe this issue, we asked what proportion of the lineages detected could not be assigned even to the phylum level (Fig. 1G). We identified specific host organisms that

FIG Legend (Continued)
percentages of reads annotated as *Wolbachia* for the various species. (G) Bar graphs summarizing the percentages of OTUs that were unable to be assigned to references in the Greengenes database. (H and I) Alignments of a 10,000-read random subset of assigned (H) and unassigned (I) sequence reads showing positions with greater than 10% non-gap sequences, emphasizing that the unassigned reads resemble the assigned reads in sequence.

had notably high fractions of unassigned bacterial lineages (as judged by comparison to the Greengenes database), including tiger shark, skate, giant millipede, two-spotted bumble bee, and rhinoceros beetle. Analysis of these sequences indicated that despite their divergence, they were recognizable as 16S rRNA gene sequences (Fig. 1H and I). Further analysis using BLAST allowed some of these sequences to be tentatively assigned, revealing, for example, possible members of the uncommon phyla GN02 and SR1 and candidate phylum SBR1093 (Fig. S4). Analysis of published data sets showed that unassigned lineages were enriched in invertebrates and fish as well and that more fully characterized bacterial lineages were more common in mammals (Fig. S2A to J). These findings provide promising starting points for future studies targeting discovery of novel gut bacteria.

Community structure, phylogeny, and lifestyle. We next compared bilaterian gut community structures on a phylogenetic scaffold using UniFrac. In this method, two samples are compared by arranging their OTU sequences on a common tree and quantifying how much of the branch length is unique to each sample. Figure 2 summarizes data for unweighted UniFrac, which generates distances based on presence/absence information. All pairwise distances between samples were calculated, and the resulting matrix was reduced to a two-dimensional summary using t-distributed stochastic neighbor embedding (t-SNE). Sample clustering associated with host phylogeny was apparent. Such clustering was also seen when species centroids were used rather than individual samples (Fig. S5A to D). Mammals formed a broad cluster, with notable subclusters for specific clades, e.g., *Cetacea*, *Artiodactyla*, *Langomorpha*, and *Rodentia*. Primates, extensively represented in this data set, formed a broad cluster, with specific subclusters classified by species. Members of *Insecta* formed a particularly tight group, with subclusters for *Orthoptera*, *Diptera*, *Hymenoptera*, and *Coleoptera*.

We used several approaches to measure this clustering with host taxonomy. Permutational multivariate analysis of variance (PERMANOVA) testing showed that gut microbial communities clustered with host taxonomy for all host taxonomic levels queried (PERMANOVA $P < 10^{-6}$ for phylum, class, order, family, genus, and species). In another analytic approach, evolutionary times of divergence among all host species were obtained from TimeTree (24), and the resulting host evolutionary distance matrix was compared to the unweighted UniFrac matrix using the Mantel test. The level of overlay of the two matrices was found to be significantly higher than would be expected by chance (Mantel $P < 10^{-6}$), indicating a strong association.

Samples also partitioned by diet, although diet was commonly correlated with host phylogeny and difficult to disentangle in our data. Samples clustered by diet type (carnivore, omnivore, or herbivore) at a high level of significance (PERMANOVA $P < 10^{-6}$). Even after accounting for phylum identity, the clustering was still highly significant (PERMANOVA $P < 10^{-6}$). Samples from *Chondrichthyes*, carnivorous sharks and rays, clustered (PERMANOVA $P < 10^{-6}$), as did samples from *Carnivora* (PERMANOVA $P = 0.0001$; $P = 0.02$ for clustering within mammals). Among the previously published whale samples (18), those from the filter-feeding whales were all dominated by *Firmicutes* (Fig. 1; see also Fig. S2I and J) and clustered seperately (PERMANOVA $P = 0.0002$) from the carnivorous toothed whales which were dominated by other phyla (Fig. S2I and J). Thus, our data are consistent with a role for diet in gut microbiota diversification, as has been suggested previously for humans (22, 25) and, broadly, for mammals (17, 26), but diet was convoluted with taxonomy in our data. Host organism weight was also significantly associated with community structure, even after accounting for host order (PERMANOVA $P < 10^{-6}$).

Species-area relationships in the bilaterian microbiome. We next investigated gut microbial colonization from the perspective of species-area analysis. In the ecological literature, a strong pattern has been observed whereby larger islands are more species rich (assuming that each of the islands compared is the same distance from a continental source of migrants) (1–3). Here we carried out a species-area analysis of over 1,098 gut microbiome samples, including our rigorously vetted sample series.
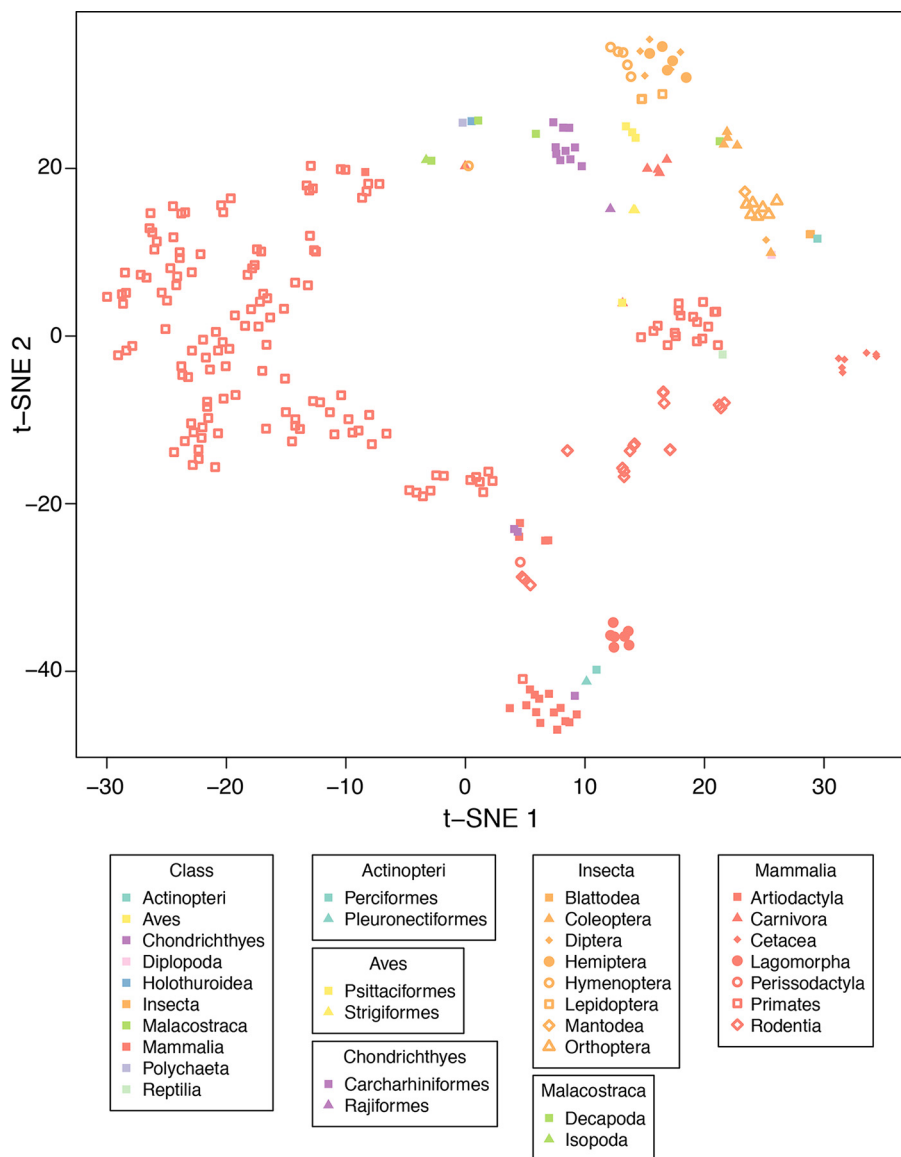
FIG 2 t-SNE plot of UniFrac distances. A two-dimensional representation of unweighted UniFrac distances was generated using t-distributed stochastic neighbor embedding (t-SNE). Samples are colored to indicate phylogenetic class, and orders are further broken out by point shape. Weighted and unweighted UniFrac t-SNE plots of species centroids (including species names) are shown in Fig. S5.

Organism weight was used to approximate gut size, allowing us to compare organisms ranging from bedbug (0.009 g) to right whale (72,000,000 g).

Modeling log bacterial richness (expected number of rarefied OTUs) as a linear function of log weight yielded a significant positive slope of 0.054 (95% credible interval of the posterior mean, 0.036 to 0.073), indicating greater richness in the larger guts (Fig. 3). This supports models in which (i) larger guts contain more niches and/or (ii) the larger microbial population sizes possible in larger guts oppose extinction due to stochastic fluctuations in numbers.

Assessment of oxygen utilization by bacterial communities showed a correlation with host body weight. Each bacterial lineage was analyzed where possible for its status as an anaerobe or aerobe, and the results were pooled for the bacterial lineages within a host individual and then for host individuals within a species (Fig. S3A). The ratio of aerobes to obligate anaerobes was significantly correlated with host weight (Fig. S3B; Pearson correlation = -0.73, $P < 10^{-6}$), suggesting that the larger guts may be more anaerobic in the luminal space and that this may result in new niche creation.
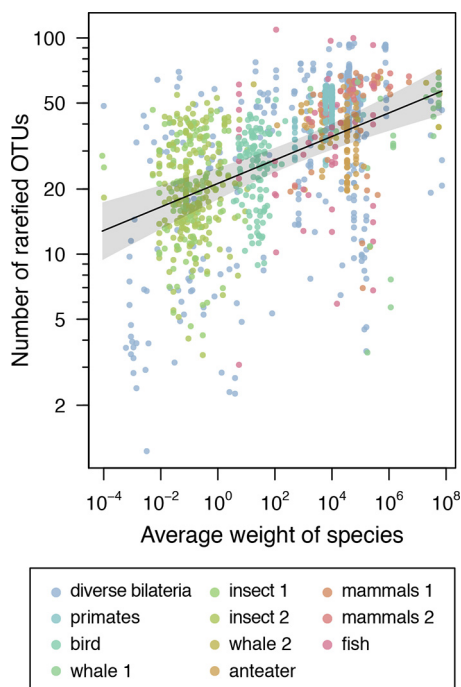
mBio

FIG 3 Species-area analysis. The relationship between the weight of the host organism and the number of gut bacterial OTUs found in fecal samples was estimated using a Bayesian regression model. All samples, both those newly determined here and those from previously published data sets, are included. The gray-shaded region shows the 95% credible interval for the slope. The OTU counts were normalized across sample sets as described in Materials and Methods. Origins of sample sets are as follows: diverse *Bilateria*, this work; primates, reference 14; birds, reference 13; whale 1 and 2, reference 18; insects 1 and 2, reference 12; anteaters, reference 15; fish, reference 16; mammals 1, reference 17; mammals 2, reference 19. Samples were rarefied to 100 reads each.

Community structure and models for community assembly. Ecological theory describing colonization of islands in the absence of deterministic factors—the neutral model—predicts the shape of rank abundance curves, allowing our data to be inter-rogated to assess agreement with this prediction. According to this model, colonization is a function of immigration of individuals from a source population, together with their birth and death rates, and all potential colonists are equal. Lack of agreement between predicted and observed rank-abundance curves allows rejection of the neutral model, pointing instead to alternative theories such as selection of specific colonists by environmental niches.

Figure 4 shows an assessment of fits to 10 model types, including the neutral model. Comparisons based on Akaike information content showed that the neutral model fits poorly relative to other simple models of species abundance (Fig. 4A). An elaboration of the neutral model where microbiota are acquired vertically from parents as well as from the environment yields rank-abundance curves that fit a log-normal distribution (27), which also does not match our results. Rather, the fit determined on the basis of a power law or modified power-bend curve provided a closer match of the data. Almost all gut samples had a long tail of rare species with many OTUs and few read counts (Fig. 4B to E), which has previously been reported for marine microbial communities (28) and is characteristic of bilaterian gut communities as well. Such long tails of rare species are inconsistent with the predictions of neutral theory.

Influence of data processing on the conclusions drawn. Several approaches were taken to assess the influence of sequencing error on our conclusions. Chimera filtering was not applied to the data initially due to the difficulty of identifying chimeras involving pairs of previously unstudied 16S rRNA gene sequences. To test for possible confounding by chimeras, we applied chimera filtering using ChimeraSlayer and USE-ARCH to our gut microbiome samples (29, 30). ChimeraSlayer found 0.16% potential
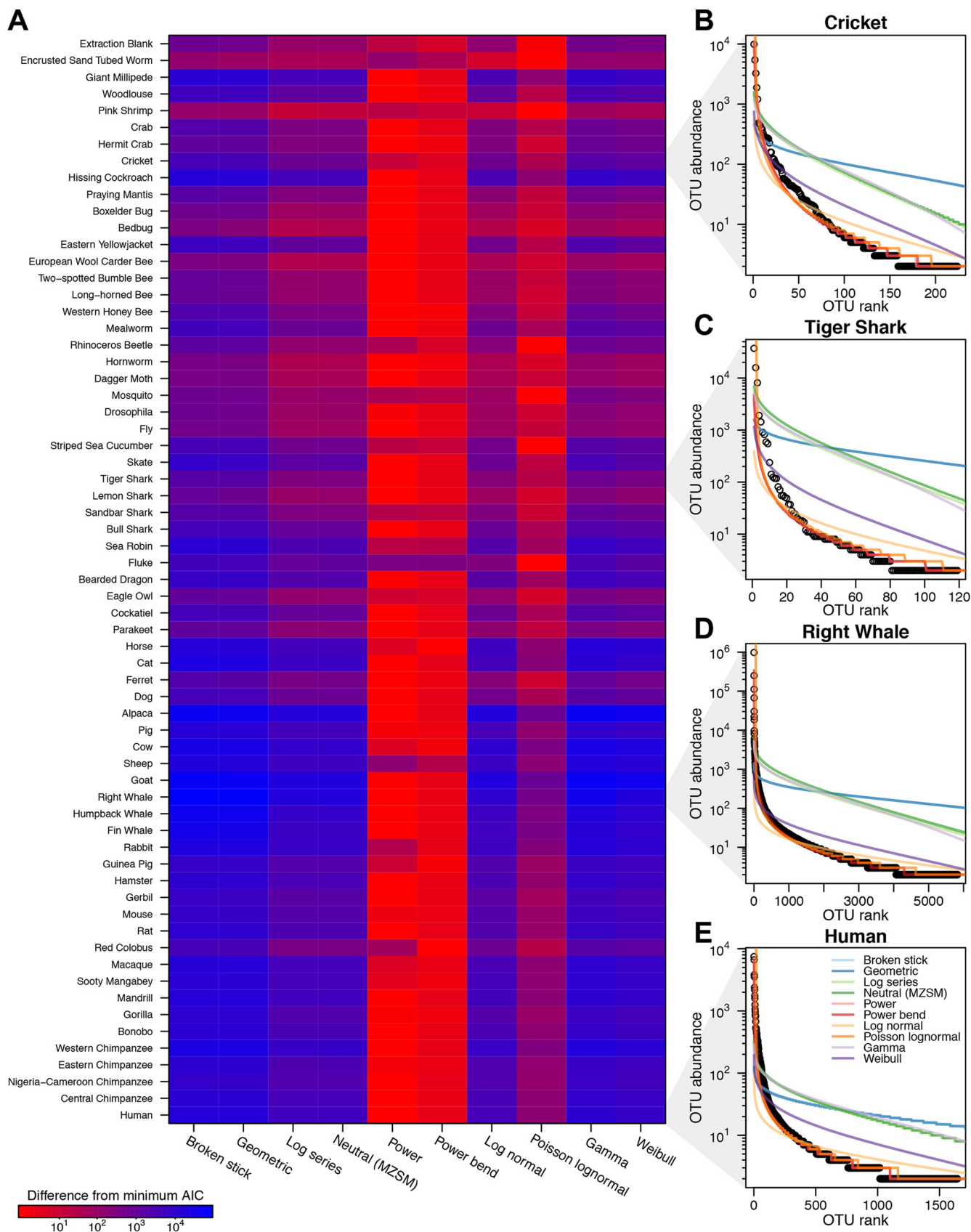
FIG 4 The bilaterian gut microbiota does not fit predictions of neutral assembly models. (A) The abundance of OTUs from each sample was assessed for their fits to community models (columns) using Akaike information content (AIC) and the AIC averaged within each species (rows). Within each row,

(Continued on next page)

chimeras and USEARCH found 3.4% potential chimeras, suggesting that chimera formation did not strongly influence our conclusions.

To test for possible effects of sequencing error, we denoised our data set using the filtering program DADA2 (31). DADA2 assumes that all sequence variants represented by single reads are products of error and removes them and clusters sequences into biological sequence variants based on the estimated probability of a read being a sequencing error. DADA2 yielded 84,259 sequence variants compared to the 521,510 OTUs yielded by the processing strategy described above. The shapes of species rank abundance curves were altered by DADA2 filtering such that the Poisson lognormal model was now favored over the power law models (Fig. S6A); however, data both before and after filtering did not match predictions of the neutral model. The species-area relationship was still observed after DADA2 filtering (Fig. S6B). Thus, our main conclusions were robust with respect to the data filtering strategy chosen.

DISCUSSION

Here we sought to test the hypothesis that the bilaterian gut microbiota is assembled by selective forces, resulting in communities diverging from the predictions of neutral colonization. For this, we assembled a broad sample set, ranging from small insects to baleen whales. Although myriad factors affect microbial community structure in each organism, our survey establishes several generalizations. The structure of gut microbial communities commonly partitioned with host taxonomy, indicating that gut microbiome structure can be preserved over long evolutionary time scales. Species-area analysis showed an increase in species richness with gut size, supporting the ideas that bacterial populations in larger guts are less likely to become extinct due to chance fluctuations and that larger guts may have more types of environmental niches. Consistent with these findings, predictions from quantitative models based on ecological neutral colonization did not match most samples, supporting an alternative model in which selection drives gut microbial community assembly. We also report numerous detections of highly divergent bacterial taxa, which can help target future studies to characterize global bacterial diversity.

Our finding of a positive association between bacterial species richness and host size suggests that mechanisms driving classic species-area relationships in island biogeography (1–3, 32) are also at play in gut microbial communities. Previous efforts to detect species-area relationships in the gut microbiota were mixed (4, 5)—the authors of the study that did not report a significant slope used data processed in multiple different ways from previous studies and themselves pointed out that this could have affected the outcome (4). For insects, estimates of microbial community size per individual can range from 10^9 bacterial cells to undetectable numbers (33); thus, fluctuations leading to extinction may be possible for the sparser communities. For larger organisms such as vertebrates, numerous studies have emphasized the heterogeneous anatomy of the gut at many spatial scales such that greater heterogeneity in microbial niches would be expected (9). Consistent with this, we found that the proportion of anaerobes was significantly higher in guts of larger animals. Previous data indicated that the lumen of the gut of mice is relatively anaerobic, whereas gut tissue is more oxygenated, and that this results in enrichment of numbers of anaerobic bacteria in the lumen and of aerobic bacteria near the gut wall (34). For bilaterians of smaller sizes, oxygen may permeate more readily throughout the gut, resulting in predominance of aerobes and reduced overall bacterial species richness. For larger bilaterians, the combination of aerobic and anaerobic environments may result in formation of bacterial communities that are richer in species.

Our values for the slope of the species-area relationship correspond to slopes that

FIG Legend (Continued)

the best-fitting models (i.e., those with the lowest AIC levels) appear red, with the color code showing the difference for each model from the minimum AIC. (B to E) Empirical rank-abundance curves and comparisons to model best fits are shown for single samples from human (B), right whale (C), tiger shark (D), and cricket (E).

are shallower than those typical of classical studies of island populations, which typically range from 0.12 to 0.35 (32). This shallower slope may indicate that animal guts are not as insular as islands with more-frequent (potentially low-abundance) immigrations from the environment/conspecifics or with a slower accumulation of habitat heterogeneity in large guts than on large islands. A potential technical issue which might have contributed to our findings is that smaller animals were often studied as dissected guts, potentially resulting in representation of a greater number of different niches than were present in fecal samples from larger animals.

Several aspects of our data argue against neutral (nonselective) processes in colonization of the bilaterian gut. The profiles of species-abundance curves display long tails (Fig. 4)—representative of the presence of many species at low abundance—which are not predicted by neutral theory. For marine organisms, abundant seawater lineages were scarce in gut samples, highlighting likely selection by the gut environment (see Table S2 in the supplemental material). For certain insects, some of the gut microbiotas have been reported to resemble local environmental taxa and to differ among individuals of the same species (33), suggesting possible neutral colonization from the environment. However, for the great majority of insects studied here, gut community rank-abundance curves did not show the profile expected for neutral colonization. For mammals, attempts to introduce new bacterial strains into the gut generally work poorly in a colonized host (35-38), consistent with gut bacterial communities adapting to their environments and so resisting invasion by new strains.

Numerous further theories have attempted to explain the shapes of species-abundance curves (39). Unfortunately, while interesting, these models rarely make unique predictions and so do not specify unique assembly mechanisms that can be applied here. Several lines of evidence have suggested that communities more open to immigration have a higher proportion of rare species (39), which provides a candidate mechanism contributing to the long tails typically observed for gut microbial species-abundance curves.

Despite the geographic dispersal of bilaterian individuals, we found that members of the same species usually resembled each other with respect to the characteristics under consideration more than they resembled members of other species. This supports the inference that dispersal of microbes between individuals is usually not limiting. Several mechanisms are known to assist transfer of microbiota between generations within a species. Vertical inheritance is well known for *Wolbachia* in invertebrates (40). Some insect mothers smear fecal material on newly deposited eggs (40). Cohabitation is associated with sharing bacterial lineages among social insects (33), laboratory mice (11), and humans and their dogs (41). Transmission via breastfeeding has been proposed to mediate human mother-to-infant transmission (42). Thus, the recurrent patterns of microbial community membership within host species suggest that dispersal of microbes is usually not a barrier to gut community construction.

This study had several limitations. The use of 16S rRNA gene tags provides a tractable window on community structure for a sample set that includes tissue specimens, but additional studies using shotgun metagenomic sequencing could yield richer (though quite complex) data. Although a large number of specimens and species were studied, our sampling scheme was opportunistic and the coverage of *Bilateria* uneven. Going forward, it would be valuable to target specific animal groups with focused questions, for example, to investigate the influences of diet and lifestyle (laboratory housed, domestic, or wild) and associations with humans.

In summary, these data suggest that neutral assembly models cannot explain the structure of the bilaterian gut microbiome and that niche selection is a likely driver. Larger animals harbor richer bacterial communities, potentially a consequence of the presence of more-diverse niches in gut, including an increase in anaerobic habitat, and greater resistance to stochastic extinction due to larger sizes of microbial populations. This work provides baseline data for understanding the structure and dynamics of the global bilaterian microbiome and provides a point of departure for additional studies.

mBio®

Going forward, the data and analytic methods presented here may be useful in assessing colonization mechanisms in human disease states (43, 44) and in evaluating the invasion of human-associated bacteria into global ecosystems.

MATERIALS AND METHODS

Sources of samples. Sources of samples are listed in Table S1 in the supplemental material, and additional sample information is available in the supplemental material.

DNA preparation. DNA from all samples was isolated in a sterile class II laminar flow hood. Due to differences in starting material, DNA was purified using three methods. First, DNA was isolated from fecal samples using a Mo Bio PowerSoil HTP 96 DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA). Sample inputs weighed between 0.0085 g (gerbil pellet) and 1.6408 g (gut contents of the large intestine of a sandbar shark). Samples were incubated for 10 min at 70°C and homogenized for 20 min with a Qiagen TissueLyser II instrument, and DNA was purified per the manufacturer's protocol. Second, to reduce background, DNA from low-biomass samples (primarily the dissected gut tissue from smaller organisms) was purified using a Qiagen DNeasy UltraClean microbial kit (Qiagen, Hilden, Germany), in single tubes, per the protocol of the manufacturer, including the recommended bead-beating step. Finally, DNA from whale feces was isolated with a PSP Stool DNA Plus kit (Stratec Biomedical, Berlin-Buch, Germany). Samples were homogenized in PSP stool DNA stabilization buffer using the TissueLyser II instrument and incubated at 95°C, and DNA was purified per the manufacturer's protocol. All DNA was stored at −20°C.

Quantification of 16S rRNA gene copies using qPCR. 16S rRNA gene copies were quantified by the amplification of the V1V2 region of the 16S rRNA gene by quantitative PCR (qPCR). Dilutions (1:4) of each DNA sample were used in reaction mixtures (total volume, 25 μl) and measured in triplicate. Primer and probe sequences and amplification conditions are described in Table S3 and reference 45.

PCR amplification of the V1V2 region of the bacterial 16S rRNA gene for Illumina sequence analysis. For each sample, the V1V2 region of the bacterial 16S rRNA gene was amplified with Golay bar-coded universal primers 27F and 338R (Table S3) (41, 46, 47). PCRs were performed in triplicate with 5 μl DNA, 7.21 μl PCR-grade water, 2.5 μl 10× buffer II, 0.19 μl *Taq*, 5 μl of the forward primer (2 μM), and 5 μl of the reverse primer (2 μM). PCRs were prepared in a PCR clean room using an EpMotion 5075 liquid handling workstation (Eppendorf, Hamburg, Germany) and run in quadruplicate on an Applied Biosystems GeneAmp PCR 9700 system (Thermo Fisher Scientific Inc., Waltham, MA) under the following cycling conditions: initial denaturation at 95°C for 5 min; 30 cycles of denaturation at 95°C for 30 s, annealing at 56°C for 30 s, and extension at 72°C for 90 s; and a final extension at 72°C for 8 min. Reaction replicates were pooled into libraries and purified using Agencourt AMPure XP beads (Beckman Coulter, Inc., Indianapolis, IN) per the manufacturer's protocol. The purified PCR products were then pooled and sequenced using an Illumina MiSeq platform. qPCR quantification of 16S rRNA gene copies was used to ensure that a minimum of 1,000 16S rRNA gene copies were input into each reaction. To pass our quality filter, samples were also required to contain at least 1,000 Illumina sequence reads.

Analytical methods. Sequence data was processed using QIIME 1.9.1 with default parameters (48). OTUs were selected by clustering reads at 97% sequence similarity. Taxonomic assignments were generated by comparison to the Greengenes reference database (49). The consensus taxonomy assignment implemented used the most detailed lineage description shared by two of the top three best-matching reference sequences in the Greengenes database. We mostly used data without rarefaction in order to maximize the amount of sequence information available. In cases where the amount of sequence analyzed could have affected the outcome, as in the species-area analysis, we did use rarefied data (for each sample, we calculated the number of OTUs expected to be observed for 100 reads). Additional analytical methods can be found in the Appendices.

Data availability. Sequences from this study are available at the NCBI SRA (https://www.ncbi.nlm.nih.gov/Traces/study/?acc=SRP115877).

APPENDIX

Additional sample information. Most of the information on samples studied is provided in Table S1. Some additional data is as follows.

Feces samples from healthy human individuals were obtained from a previous study approved by the Perelman School of Medicine at the University of Pennsylvania Institutional Review Board (50). Six rhesus macaques (*Macaca mulatta*) with no history of gastrointestinal disease were sampled at the Tulane National Primate Research Center. Fecal samples were collected from stainless steel collection pans below the animals' cages. Fecal samples (*n* = 111) from western (*P. t. verus*, *n* = 1), Nigeria-Cameroonian (*P. t. ellioti*, *n* = 1), central (*P. t. troglodytes*, *n* = 5) and eastern (*P. t. schweinfurthii*, *n* = 33) chimpanzees and from western (*G. g. gorilla*, *n* = 3) and eastern (*G. b. graueri*, *n* = 2) lowland gorillas, as well as from bonobos (*Pan paniscus*, *n* = 66), were obtained from existing specimen banks (51–56). Except for four specimens (CCptv01, SYptt43, TCptt98, and CFpts01) that were obtained from sanctuary chimpanzees, all other fecal samples were collected from wild-living ape populations at

remote forest sites (Table S1) (Fig. S1D). All samples were obtained noninvasively, preserved (1:1 [vol/vol]) in RNAlater, transported at ambient temperatures, and stored at −80°C. Fecal DNA was extracted using a QIAamp Stool DNA minikit (Qiagen). All specimens were subjected to host mitochondrial DNA analysis to confirm their species and subspecies origin (51–58). In the analysis, each of the four chimpanzee subspecies was treated as a separate species. Samples were also obtained from two mandrills from the Tchimpounga sanctuary and two habituated sooty mangabeys from the Tai Forest. Sampling sites are shown in Fig. S1D. All samples were obtained with approval from the respective governments as previously described (51–56) and were shipped in compliance with regulations from the Convention on International Trade in Endangered Species of Wild Fauna and Flora and country-specific import and export permits. Fecal pellets from healthy laboratory mice were collected at the Perelman School of Medicine at the University of Pennsylvania (protocol no. 803408). Right whale samples were collected under Fisheries and Oceans Canada license no. 325842 (59), the fin whale sample was collected under NOAA permit no. 16325, and the humpback samples were provided by Jan Straley, collected under NOAA permits 474-1700-02 and 14122. Shark specimens were collected as part of commercial fishing activities regulated by the National Marine Fisheries Service Federal Management Plan for Atlantic Tunas, Swordfish, and Sharks.

In instances where there was no available expelled fecal material, such as for some insects, samples were obtained by gut dissection of the frozen individuals. Each individual was weighed, washed with 70% ethanol, and rinsed with phosphate-buffered saline (PBS). Individuals were then placed on a sterile petri dish, and the gut was removed with sterile needle nose tweezers.

Additional analytic methods. The aerobic or anaerobic status of bacterial taxa was copied by hand from reports in Bergey's Manual of Systematics of Archaea and Bacteria (60). A detailed description of all annotations used is provided in Table S4. The anaerobic or aerobic status was scored for each bacterial lineage, starting with the lowest-ranking taxon and moving on to higher-ranking taxa. Aerobic or anaerobic status was assigned to taxa above the rank of genus if listed in the description of the taxon or if >99.9% of reads assigned to lower-ranking taxa observed in this study had a consistent status (further details are provided in Table S4).

Phylogenetic trees were inferred from the OTU data using FastTree (61). Similarity between samples was assessed by weighted and unweighted UniFrac distance (62, 63). Data files from QIIME were analyzed in the R environment for statistical computing. Phylogenetic distances were retrieved from TimeTree (24), using a closely related proxy species if necessary. We used the Adonis function from the vegan package to perform PERMANOVA.

Maximum likelihood estimates for various species abundance distributions were fitted using the sads R package (64) with functions modified slightly for additional robustness (e.g., multiple starting values for optimization and calculations on log values to avoid numeric overflow). Fitted models included the broken stick model (65); log series (66, 67); neutral metacommunity zero-sum multinomial (6, 66, 68); zeta, power, and power bend (69); Poisson lognormal (70); lognormal (71); Weibull; and negative binomial, gamma, and geometric distributions. Models were compared using the Akaike information criteria to control for overfitting in models with additional parameters.

We obtained additional data from previously published studies on the microbiome of insects (12), birds (13), primates (14), myrmecophagous mammals (15), fish (16), diverse mammals sequenced using the 454 method (17), whales and other mammals (18), and diverse mammals sequenced using the Illumina method (19). The whale data included both 454 sequencing and Illumina sequencing data, which were treated separately. The insect data contained unpaired reads from the 5′ and 3′ ends of the amplicon, which were treated separately. Thus, we had data from 8 additional studies broken into 10 data sets.

To estimate the species-area relationship in animal guts, we fit a Bayesian regression model of the rarefied OTU counts observed in each animal gut in relation to the mean weight of the hosts using Stan (72). This model included samples from our own data set and the 10 additional data sets. As extraction and amplification methods, amplicon targets, and sequencing technologies differed between data sets, we allowed each data set its own offset. To compare the data from the studies whose results are shown in Fig. 3, we subtracted the mean of the estimated posterior distribution for these offsets. The model assumed that the log rarefied OTU count, count_i in sample i were normally distributed as follows:

$$\text{count}_i = N(\mu_i, \sigma_{\text{dataset}_i})$$

where dataset_i is the data set that sample i came from, σ_x is the estimated variability in data set x, and μ_i is given by following equation:

$$\mu_i = \text{offset}_{\text{dataset}_i} + \beta\text{weight}_{\text{species}_i} + \theta_{\text{species}_i}$$

Where species_i is the host species for sample i, offset_x is the offset for data set x, weight_x is the log weight for species x, β measures the relationship between weight and OTU count, and θ_x is the mean offset for species x from a linear weight-count relationship. Species θ values were given a normal prior with mean equal to 0 and a separate variance value for each phylogenetic class. Variance parameters were given a gamma(1,1/10) prior.

Reads that were annotated by Greengenes as coming from chloroplast genomes were removed. Samples that contained fewer than 1,000 sequence reads were removed. We investigated potential chimeric reads generated by amplification and sequencing using ChimeraSlayer and USEARCH (29, 30). ChimeraSlayer uses BLAST alignment against representative aligned sequences to identify potential chimera parents. USEARCH performs *de novo* chimera detection based on abundances as well as reference-based chimera detection for raw sequence reads. ChimeraSlayer found 1,265 potential chimeras of 779,732 representative OTU sequences. There was little association between these chimeras and singleton OTUs (95% confidence interval of odds ratio [CI OR], 0.84 to 1.05; Fisher's exact $P = 0.28$) or reads unassigned to the Greengenes database (95% CI OR, 0.85 to 1.13; Fisher's exact $P = 0.8$). USEARCH found 776,531 potential chimeras among 22,538,324 total sequence reads. Among USEARCH chimeras, there was a slight depletion of singleton reads (95% CI, 0.69 to 0.70; Fisher's exact $P < 10^{-6}$) and of reads unassigned to Greengenes (95% CI, 0.59 to 0.60; Fisher's exact $P < 10^{-6}$) but the association was modest.

Code used in this analysis is archived on Zenodo at https://doi.org/10.5281/zenodo.1194331.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.00319-18.

FIG S1, PDF file, 1 MB.
FIG S2, PDF file, 2.8 MB.
FIG S3, PDF file, 0.3 MB.
FIG S4, PDF file, 0.3 MB.
FIG S5, PDF file, 0.2 MB.
FIG S6, PDF file, 1 MB.
TABLE S1, XLSX file, 0.1 MB.
TABLE S2, XLSX file, 8.4 MB.
TABLE S3, XLSX file, 0.01 MB.
TABLE S4, XLSX file, 0.05 MB.

ACKNOWLEDGMENTS

REFERENCES

1. Arrhenius O. 1921. Species and area. J Ecol 9:95–99. https://doi.org/10.2307/2255763.
2. Lomolino MV. 2000. Ecology's most general, yet protean pattern: the species-area relationship. J Biogeogr 27:17–26. https://doi.org/10.1046/j.1365-2699.2000.00377.x.
3. McGuinness KA. 1984. Species-area curves. Biol Rev 59:423–440. https://doi.org/10.1111/j.1469-185X.1984.tb00711.x.
4. Kieft TL, Simmons KA. 2015. Allometry of animal-microbe interactions and global census of animal-associated microbes. Proc Biol Sci 282. https://doi.org/10.1098/rspb.2015.0702.
5. Godon JJ, Arulazhagan P, Steyer JP, Hamelin J. 2016. Vertebrate bacterial gut diversity: size also matters. BMC Ecol 16:12. https://doi.org/10.1186/s12898-016-0071-2.
6. Hubbell SP. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Oxford, United Kingdom.
7. Condit R, Pitman N, Leigh EG, Jr, Chave J, Terborgh J, Foster RB, Núñez P, Aguilar S, Valencia R, Villa G, Muller-Landau HC, Losos E, Hubbell SP. 2002. Beta-diversity in tropical forest trees. Science 295:666–669. https://doi.org/10.1126/science.1066854.
8. Li L, Ma ZS. 2016. Testing the neutral theory of biodiversity with human microbiome datasets. Sci Rep 6:31448. https://doi.org/10.1038/srep31448.
9. Stevens CE, Hume ID. 1995. Comparative physiology of the vertebrate digestive system, 2nd ed. Cambridge University Press, New York, NY.
10. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, Turner P, Parkhill J, Loman NJ, Walker AW. 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol 12:87. https://doi.org/10.1186/s12915-014-0087-z.
11. Kim D, Hofstaedter CE, Zhao C, Mattei L, Tanes C, Clarke E, Lauder A, Sherrill-Mix S, Chehoud C, Kelsen J, Conrad M, Collman RG, Baldassano R, Bushman FD, Bittinger K. 2017. Optimizing methods and dodging pitfalls in microbiome research. Microbiome 5:52. https://doi.org/10.1186/s40168-017-0267-5.
12. Yun JH, Roh SW, Whon TW, Jung MJ, Kim MS, Park DS, Yoon C, Nam YD, Kim YJ, Choi JH, Kim JY, Shin NR, Kim SH, Lee WJ, Bae JW. 2014. Insect gut bacterial diversity determined by environmental habitat, diet, developmental stage, and phylogeny of host. Appl Environ Microbiol 80:5254–5264. https://doi.org/10.1128/AEM.01226-14.
13. Hird SM, Sánchez C, Carstens BC, Brumfield RT. 2015. Comparative gut microbiota of 59 neotropical bird species. Front Microbiol 6:1403. https://doi.org/10.3389/fmicb.2015.01403.
14. Clayton JB, Vangay P, Huang H, Ward T, Hillmann BM, Al-Ghalith GA, Travis DA, Long HT, Tuan BV, Minh VV, Cabana F, Nadler T, Toddes B, Murphy T, Glander KE, Johnson TJ, Knights D. 2016. Captivity humanizes the primate microbiome. Proc Natl Acad Sci U S A 113:10376–10381. https://doi.org/10.1073/pnas.1521835113.
15. Delsuc F, Metcalf JL, Wegener Parfrey L, Song SJ, González A, Knight R. 2014. Convergence of gut microbiomes in myrmecophagous mammals. Mol Ecol 23:1301–1317. https://doi.org/10.1111/mec.12501.
16. Givens CE, Ransom B, Bano N, Hollibaugh JT. 2015. Comparison of the gut microbiomes of 12 bony fish and 3 shark species. Mar Ecol Prog Ser 518:209–223. https://doi.org/10.3354/meps11034.
17. Muegge BD, Kuczynski J, Knights D, Clemente JC, González A, Fontana L, Henrissat B, Knight R, Gordon JI. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science 332:970–974. https://doi.org/10.1126/science.1198719.
18. Sanders JG, Beichman AC, Roman J, Scott JJ, Emerson D, McCarthy JJ, Girguis PR. 2015. Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. Nat Commun 6:8285. https://doi.org/10.1038/ncomms9285.
19. Parfrey LW, Walters WA, Lauber CL, Clemente JC, Berg-Lyons D, Teiling

C, Kodira C, Mohiuddin M, Brunelle J, Driscoll M, Fierer N, Gilbert JA, Knight R. 2014. Communities of microbial eukaryotes in the mammalian gut within the context of environmental eukaryotic diversity. Front Microbiol 5:298. https://doi.org/10.3389/fmicb.2014.00298.
20. McKenna P, Hoffmann C, Minkah N, Aye PP, Lackner A, Liu Z, Lozupone CA, Hamady M, Knight R, Bushman FD. 2008. The macaque gut microbiome in health, lentiviral infection, and chronic enterocolitis. PLoS Pathog 4:e20. https://doi.org/10.1371/journal.ppat.0040020.
21. Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. Nature 486:207–214. https://doi.org/10.1038/nature11234.
22. Wu GD, Chen J, Hoffmann C, Bittinger K, Chen YY, Keilbaugh SA, Bewtra M, Knights D, Walters WA, Knight R, Sinha R, Gilroy E, Gupta K, Baldassano R, Nessel L, Li H, Bushman FD, Lewis JD. 2011. Linking long-term dietary patterns with gut microbial enterotypes. Science 334:105–108. https://doi.org/10.1126/science.1208344.
23. Dudek NK, Sun CL, Burstein D, Kantor RS, Aliaga Goltsman DS, Bik EM, Thomas BC, Banfield JF, Relman DA. 2017. Novel microbial diversity and functional potential in the marine mammal oral microbiome. Curr Biol 27:3752–3762.e6 https://doi.org/10.1016/j.cub.2017.10.040.
24. Hedges SB, Marin J, Suleski M, Paymer M, Kumar S. 2015. Tree of life reveals clock-like speciation and diversification. Mol Biol Evol 32:835–845. https://doi.org/10.1093/molbev/msv037.
25. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–563. https://doi.org/10.1038/nature12820.
26. Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. 2008. Evolution of mammals and their gut microbes. Science 320:1647–1651. https://doi.org/10.1126/science.1155725.
27. Zeng Q, Sukumaran J, Wu S, Rodrigo A. 2015. Neutral models of microbiome evolution. PLoS Comput Biol 11:e1004365. https://doi.org/10.1371/journal.pcbi.1004365.
28. Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci U S A 103:12115–12120. https://doi.org/10.1073/pnas.0605127103.
29. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. https://doi.org/10.1093/bioinformatics/btq461.
30. Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, Ciulla D, Tabbaa D, Highlander SK, Sodergren E, Methe B, DeSantis TZ; Human Microbiome Project Consortium, Petrosino JF, Knight R, Birren BW. 2011. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res 21:494–504. https://doi.org/10.1101/gr.112730.110.
31. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJ, Holmes SP. 2016. DADA2: high-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869.
32. MacArthur RH, Wilson EO. 1967. The theory of island biogeography. Princeton University Press, Princeton, NJ.
33. Engel P, Moran NA. 2013. The gut microbiota of insects—diversity in structure and function. FEMS Microbiol Rev 37:699–735. https://doi.org/10.1111/1574-6976.12025.
34. Albenberg L, Esipova TV, Judge CP, Bittinger K, Chen J, Laughlin A, Grunberg S, Baldassano RN, Lewis JD, Li H, Thom SR, Bushman FD, Vinogradov SA, Wu GD. 2014. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. Gastroenterology 147:1055–1063.e8. https://doi.org/10.1053/j.gastro.2014.07.020.

mBio®

35. McNulty NP, Yatsunenko T, Hsiao A, Faith JJ, Muegge BD, Goodman AL, Henrissat B, Oozeer R, Cools-Portier S, Gobert G, Chervaux C, Knights D, Lozupone CA, Knight R, Duncan AE, Bain JR, Muehlbauer MJ, Newgard CB, Heath AC, Gordon JI. 2011. The impact of a consortium of fermented milk strains on the gut microbiome of gnotobiotic mice and monozygotic twins. Sci Transl Med 3:106ra106. https://doi.org/10.1126/scitranslmed.3002701.

36. Klatt NR, Canary LA, Sun X, Vinton CL, Funderburg NT, Morcock DR, Quiñones M, Deming CB, Perkins M, Hazuda DJ, Miller MD, Lederman MM, Segre JA, Lifson JD, Haddad EK, Estes JD, Brenchley JM. 2013. Probiotic/prebiotic supplementation of antiretrovirals improves gastrointestinal immunity in SIV-infected macaques. J Clin Invest 123:903–907. https://doi.org/10.1172/JCI66227.

37. Kim SW, Suda W, Kim S, Oshima K, Fukuda S, Ohno H, Morita H, Hattori M. 2013. Robustness of gut microbiota of healthy adults in response to probiotic intervention revealed by high-throughput pyrosequencing. DNA Res 20:241–253. https://doi.org/10.1093/dnares/dst006.

38. Shen TC, Albenberg L, Bittinger K, Chehoud C, Chen YY, Judge CA, Chau L, Ni J, Sheng M, Lin A, Wilkins BJ, Buza EL, Lewis JD, Daikhin Y, Nissim I, Yudkoff M, Bushman FD, Wu GD. 2015. Engineering the gut microbiota to treat hyperammonemia. J Clin Invest 125:2841–2850. https://doi.org/10.1172/JCI79214.

39. McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Benecha HK, Dornelas M, Enquist BJ, Green JL, He F, Hurlbert AH, Magurran AE, Marquet PA, Maurer BA, Ostling A, Soykan CU, Ugland KI, White EP. 2007. Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. Ecol Lett 10:995–1015. https://doi.org/10.1111/j.1461-0248.2007.01094.x.

40. Pietri JE, DeBruhl H, Sullivan W. 2016. The rich somatic life of Wolbachia. Microbiologyopen 5:923–936. https://doi.org/10.1002/mbo3.390.

41. Song SJ, Lauber C, Costello EK, Lozupone CA, Humphrey G, Berg-Lyons D, Caporaso JG, Knights D, Clemente JC, Nakielny S, Gordon JI, Fierer N, Knight R. 2013. Cohabiting family members share microbiota with one another and with their dogs. Elife 2:e00458. https://doi.org/10.7554/eLife.00458.

42. Pannaraj PS, Li F, Cerini C, Bender JM, Yang S, Rollie A, Adisetiyo H, Zabih S, Lincez PJ, Bittinger K, Bailey A, Bushman FD, Sleasman JW, Aldrovandi GM. 2017. Association between breast milk bacterial communities and establishment and development of the infant gut microbiome. JAMA Pediatr 171:647–654. https://doi.org/10.1001/jamapediatrics.2017.0378.

43. Beck JM, Schloss PD, Venkataraman A, Twigg H, III, Jablonski KA, Bushman FD, Campbell TB, Charlson ES, Collman RG, Crothers K, Curtis JL, Drews KL, Flores SC, Fontenot AP, Foulkes MA, Frank I, Ghedin E, Huang L, Lynch SV, Morris A, Palmer BE, Schmidt TM, Sodergren E, Weinstock GM, Young VB; Lung HIV Microbiome Project. 2015. Multicenter comparison of lung and oral microbiomes of HIV-infected and HIV-uninfected individuals. Am J Respir Crit Care Med 192:1335–1344. https://doi.org/10.1164/rccm.201501-0128OC.

44. Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, Schmidt TM. 2015. Application of a neutral community model to assess structuring of the human lung microbiome. MBio 6:e02284-14. https://doi.org/10.1128/mBio.02284-14.

45. Hill DA, Hoffmann C, Abt MC, Du Y, Kobuley D, Kirn TJ, Bushman FD, Artis D. 2010. Metagenomic analyses reveal antibiotic-induced temporal and spatial changes in intestinal microbiota with associated alterations in immune cell homeostasis. Mucosal Immunol 3:148–158. https://doi.org/10.1038/mi.2009.132.

46. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. 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. https://doi.org/10.1073/pnas.1000080107.

47. Liu Z, Lozupone C, Hamady M, Bushman FD, Knight R. 2007. Short pyrosequencing reads suffice for accurate microbial community analysis. Nucleic Acids Res 35:e120. https://doi.org/10.1093/nar/gkm541.

48. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336. https://doi.org/10.1038/nmeth.f.303.

49. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J 6:610–618. https://doi.org/10.1038/ismej.2011.139.

50. Lewis JD, Chen EZ, Baldassano RN, Otley AR, Griffiths AM, Lee D, Bittinger K, Bailey A, Friedman ES, Hoffmann C, Albenberg L, Sinha R, Compher C, Gilroy E, Nessel L, Grant A, Chehoud C, Li H, Wu GD, Bushman FD. 2015. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric Crohn's disease. Cell Host Microbe 18:489–500. https://doi.org/10.1016/j.chom.2015.09.008.

51. Liu W, Sundararaman SA, Loy DE, Learn GH, Li Y, Plenderleith LJ, Ndjango JB, Speede S, Atencia R, Cox D, Shaw GM, Ayouba A, Peeters M, Rayner JC, Hahn BH, Sharp PM. 2016. Multigenomic delineation of plasmodium species of the laverania subgenus infecting wild-living chimpanzees and gorillas. Genome Biol Evol 8:1929–1939. https://doi.org/10.1093/gbe/evw128.

52. Liu W, Li Y, Learn GH, Rudicell RS, Robertson JD, Keele BF, Ndjango JB, Sanz CM, Morgan DB, Locatelli S, Gonder MK, Kranzusch PJ, Walsh PD, Delaporte E, Mpoudi-Ngole E, Georgiev AV, Muller MN, Shaw GM, Peeters M, Sharp PM, Rayner JC, Hahn BH. 2010. Origin of the human malaria parasite Plasmodium falciparum in gorillas. Nature 467:420–425. https://doi.org/10.1038/nature09442.

53. Keele BF, Van Heuverswyn F, Li Y, Bailes E, Takehisa J, Santiago ML, Bibollet-Ruche F, Chen Y, Wain LV, Liegeois F, Loul S, Ngole EM, Bienvenue Y, Delaporte E, Brookfield JF, Sharp PM, Shaw GM, Peeters M, Hahn BH. 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science 313:523–526. https://doi.org/10.1126/science.1126531.

54. Sundararaman SA, Liu W, Keele BF, Learn GH, Bittinger K, Mouacha F, Ahuka-Mundeke S, Manske M, Sherrill-Mix S, Li Y, Malenke JA, Delaporte E, Laurent C, Mpoudi Ngole E, Kwiatkowski DP, Shaw GM, Rayner JC, Peeters M, Sharp PM, Bushman FD, Hahn BH. 2013. Plasmodium falciparum-like parasites infecting wild apes in southern Cameroon do not represent a recurrent source of human malaria. Proc Natl Acad Sci U S A 110:7020–7025. https://doi.org/10.1073/pnas.1305201110.

55. Li Y, Ndjango JB, Learn GH, Ramirez MA, Keele BF, Bibollet-Ruche F, Liu W, Easlick JL, Decker JM, Rudicell RS, Inogwabini BI, Ahuka-Mundeke S, Leendertz FH, Reynolds V, Muller MN, Chancellor RL, Rundus AS, Simmons N, Worobey M, Shaw GM, Peeters M, Hahn BH. 2012. Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. J Virol 86:10776–10791. https://doi.org/10.1128/JVI.01498-12.

56. Etienne L, Locatelli S, Ayouba A, Esteban A, Butel C, Liegeois F, Aghokeng A, Delaporte E, Mpoudi Ngole E, Peeters M. 2012. Noninvasive follow-up of simian immunodeficiency virus infection in wild-living non-habituated western lowland gorillas in Cameroon. J Virol 86:9760–9772. https://doi.org/10.1128/JVI.01186-12.

57. Ochman H, Worobey M, Kuo CH, Ndjango JB, Peeters M, Hahn BH, Hugenholtz P. 2010. Evolutionary relationships of wild hominids recapitulated by gut microbial communities. PLoS Biol 8:e1000546. https://doi.org/10.1371/journal.pbio.1000546.

58. Moeller AH, Caro-Quintero A, Mjungu D, Georgiev AV, Lonsdorf EV, Muller MN, Pusey AE, Peeters M, Hahn BH, Ochman H. 2016. Cospeciation of gut microbiota with hominids. Science 353:380–382. https://doi.org/10.1126/science.aaf3951.

59. Roman J, Nevins J, Altabet M, Koopman H, McCarthy J. 2016. Endangered right whales enhance primary productivity in the Bay of Fundy. PLoS One 11:e0156553. https://doi.org/10.1371/journal.pone.0156553.

60. Whitman WB (ed). 2015. Bergey's manual of systematics of archaea and bacteria. John Wiley & Sons, Hoboken, NJ.

61. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

62. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005.

63. Lozupone CA, Hamady M, Kelley ST, Knight R. 2007. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 73:1576–1585. https://doi.org/10.1128/AEM.01996-06.

64. Prado PI, Miranda MD, Chalom A. 2017. sads: maximum likelihood models for species abundance distributions. R package version 0.4.0. https://cran.r-project.org/package=sads.

65. MacArthur RH. 1957. On the relative abundance of bird species. Proc Natl Acad Sci U S A 43:293–295. https://doi.org/10.1073/pnas.43.3.293.

66. Alonso D, Ostling A, Etienne RS. 2008. The implicit assumption of symmetry and the species abundance distribution. Ecol Lett 11:93–105. https://doi.org/10.1111/j.1461-0248.2007.01127.x.

67. Fisher RA, Steven Corbet A, Williams CB. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. J Anim Ecol 12:42–58.

68. Alonso D, Mckane AJ. 2004. Sampling Hubbell's neutral theory of biodiversity. Ecol Lett 7:901–910. https://doi.org/10.1111/j.1461-0248.2004.00640.x.

69. Pueyo S. 2006. Diversity: between neutrality and structure. Oikos 112:392–405. https://doi.org/10.1111/j.0030-1299.2006.14188.x.

70. Bulmer MG. 1974. On fitting the Poisson lognormal distribution to species-abundance data. Biometrics 30:101–110. https://doi.org/10.2307/2529621.

71. Preston FW. 1948. The commonness, and rarity, of species. Ecology 29:254–283. https://doi.org/10.2307/1930989.

72. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. 2017. Stan: a probabilistic programming language. J Stat Soft 76. https://doi.org/10.18637/jss.v076.i01.