Identification of specialists and abundance-occupancy relationships among intestinal bacteria of *Aves*, *Mammalia*, and *Actinopterygii*

Hyatt C. Green[a]#, Jenny C. Fisher[b], Sandra L. McLellan[b], Mitchell L. Sogin[c], and Orin C. Shanks[d]

Running Head: Abundance-occupancy relationship among gut bacteria

SUNY-ESF, Syracuse, NY[a]; Univ. of Wisconsin, Milwaukee, WI[b]; Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA[c]; U.S. EPA, Cincinnati, OH[d]

# Address correspondence to Hyatt C. Green, hgreen@esf.edu

Abstract

The coalescence of next generation DNA sequencing methods, ecological perspectives, and bioinformatics analysis tools is rapidly advancing our understanding of the evolution and function of vertebrate-associated bacterial communities. Delineating host-microbial associations has applied benefits ranging from clinical treatments to protecting our natural waters. Microbial communities follow some broad-scale patterns observed for macro-organisms, but it remains unclear how specialization of intestinal vertebrate-associated communities to a particular host environment influences broad-scale patterns in microbial abundance and distribution. We analyzed the V6 region of 16S rRNA gene amplified from 106 fecal samples spanning *Aves*, *Mammalia*, and *Actinopterygii* (ray-finned fish). The interspecific abundance-occupancy relationship—where widespread taxa tend to be more abundant than narrowly distributed taxa—among operational taxonomic units (OTUs) was investigated within and among host species. In a separate analysis, specialists OTUs that were highly abundant in a single host and rare in all other hosts were identified using a multinomial model without excluding under-sampled OTUs *a priori*. We also show that intestinal microbes in humans and other vertebrates studied follow a similar interspecific abundance-occupancy relationship compared to plants and animals, as well as microbes in ocean and soil environments; but because intestinal host-associated communities have undergone intense specialization, this trend is violated by a disproportionately large number of specialist taxa. Although it is difficult to distinguish the effects of dispersal limitations, host selection, historical contingency, and stochastic processes on community assembly, results suggest bacterial taxa can be shared among diverse vertebrate hosts in ways similar to those of ‘free-living’ bacteria.

Introduction

Because the structure and composition of intestinal host-associated communities (microbiota) have both beneficial and detrimental effects on the physiology of animals, especially vertebrates (1), the factors that shape these communities have been the subject of intense research (2-6). Microbes help maintain homeostasis by exchanging signals with mammalian immune, circulatory, digestive, and neuroendocrine organ systems (7, 8); and although such specific interactions have only been investigated in a few model species, they are thought to play important roles in physiologies of most vertebrates. In turn, host type, immune system, diet (9), age (5, 10), association with cohabitants (11), and other factors shape the microbial community structure in a host-specific fashion (12). Discovery of such intimate associations between host and microbe have led to the indirect treatment of infected individuals through direct manipulation of their microbiota (6, 13). Effective management of microbial communities to improve health requires a better understanding of factors that control the community assembly of microbes through ecological approaches (14).

Ecological drift and intense selection pressure from the host are thought to drive the divergence and co-diversification of some intestinal microbes into host-associated assemblages. Physiological and genomic evidence from some bacteria suggest host-specialization, where fine-tuning to the host environment often results in increased fitness and abundance (15, 16). In contrast, it is possible that some bacteria may generalize in a wide range of hosts by utilizing a large range of substrates or by processing them more efficiently. This classic trade-off between two lifestyles has been thoroughly investigated and discussed for “large” organisms and some microbial communities, but less so for intestinal host-associated microbial communities. Additionally, identification and targeting of specialists outside the host in environmental samples, such as recreational water bodies or drinking water supplies, can help identify dangerous sources of fecal pollution by serving as host-specific indicators.

Distribution patterns for intestinal microbes have been studied less than those for microbes in other habitats (10). Taxa-area and distance-decay relationships for microbes have been observed in salt marsh sediments (17, 18) and treeholes (19, 20). The interspecific positive relationship between abundance and occupancy (also called the abundance-range or abundance-distribution relationship), whereby more abundant taxa also tend to be more widespread, has been observed for microbes in aquatic habitats (21-23) and soils (24-26). In such environments where this relationship holds true there may be a lack of a fitness trade-off between specialist and generalist lifestyles. Given the potential for intestinal microbes to be somewhat dispersal limited and under intense selection pressure from the host, it is unclear if this fitness trade-off exists for microbes in the gut environment.

In this study, we analyzed sequences from the bacterial V6 hypervariable region of the 16S rRNA gene from 106 animal fecal samples to investigate bacterial distributions within microbial communities sampled from members of *Aves* (birds), *Mammalia* (mammals), and *Actinopterygii* (ray-finned fish). We used a multinomial species classification method to identify locally abundant specialist bacterial taxa without discarding OTU observations through data normalization. We also characterized the relationship between OTU abundance and distribution between host of similar and different species.

Methods

Sample collection. The overall dataset was generated by combining data from newly collected fecal samples and previously published datasets using similar methods (Table 1). Host species common names are used throughout the manuscript for brevity. Fresh fecal samples were collected aseptically using sterile gloves, sterile disposable spatulas, and sterile 50 ml conical tubes. Fish gut contents were removed surgically and stored in 1.7 ml microcentrifuge tubes. All fresh samples were stored on ice immediately after sampling and at -80°C upon arrival to the United States Environmental Protection Agency laboratory in Cincinnati, OH.

DNA extraction, quantification, and sequencing. All DNA extractions were performed with the FastDNA Kit (Q-Biogene, Carlsbad, CA) according to the manufacturer's instructions as previously described (4). Prior to fecal DNA extraction, GITC buffer (5 M guanidine isothiocyanate, 100 mM EDTA [pH 8.0], 0.5% Sarkosyl) was mixed with approximately 1 g wet weight of fecal material to create a fecal slurry. Eight hundred microliters of each fecal slurry was bead homogenized at 4.0 m/s for 30 seconds using a MP FastPrep-24 instrument (MP Biomedicals, LLC Solon, OH). DNA was eluted in 100 µl elution buffer and stored at -20°C until further analysis. DNA yield and quality were ascertained using a NanoDrop® 2000 (Thermo Scientific, Wilmington, DE). DNA extracts (5-25 ng per reaction) were amplified using previously published primer sets and conditions (27), which are described in detail elsewhere (http://vamps.mbl.edu/resources/faq.php#tags). Pyrosequencing was performed as described previously (4).

Sequence data pre-processing, pre-clustering, clustering, and taxonomic assignment. Quality filtering and trimming of sequences was performed using the Visualization and Analysis of Microbial Population Structures interface (VAMPS, http://vamps.mbl.edu) (28) as done previously (29). Sequence data were binned according to their barcodes, trimmed, and pre-clustered to minimize the impact of sequencing errors on sample richness. Data were then downloaded from VAMPS. Both chimera removal and read clustering (97% similarity threshold) were performed simultaneously with the USEARCH function *cluster_otus* (30). Taxonomic assignment was performed using the RDP Classifier (31) using the Silva v111 RefNR reference alignment (32). Sequence data is stored on VAMPS and can be viewed without special permission using the generic guest login.

Construction of community data matrices and diversity estimates. OTU distributions were analyzed using the *vegan* (33) and *pvclust* (34) packages, as well as custom scripts in R and Python. From the original dataset, three smaller community data matrices were constructed and formed the basis for all analyses. A human only dataset (859,510 reads) was created by combining data from 33 human samples from the obese/lean dataset (35) and an additional three samples taken at initial time points from a study by Dethlefsen and colleagues (2). The cattle only dataset (629,299 reads) was composed of data from 30 samples (ten from Colorado fed processed grain (CO1 & CO2), five from Ohio fed unprocessed grain (DK), five from Georgia fed forage (USDA), and ten from Nebraska (five fed forage, NE1; and five fed unprocessed grain, NE2)) (4). The vertebrate dataset (75 samples, 1,775,995 reads) was composed of 12 samples from the obese/lean dataset, three initial time points from the study by Dethlefsen and colleagues, the cattle dataset (except CO1 & CO2), and all newly sequenced samples listed in Table 1. After removal of OTUs observed only once (‘singletons’) each dataset was then subsampled randomly without replacement to the minimum sample library size (4986, 8017, and 14844 for vertebrate, human, and cattle datasets, respectively) for 100 iterations to construct community matrices. This normalization procedure resulted in OTU counts less than one in many cases because some OTUs were not represented in all subsamples taken at each iteration. Normalized counts were used only for abundance-occupancy relationship investigations. Non-normalized datasets were used to identify specialists with CLAM tests. Pooled diversities and their standard errors were estimated using the function *vegan::estimateR*.

Community clustering and analysis of variance. Square root transformations followed by Wisconsin transformation (see *?vegan::decostand*) on the vertebrate dataset were used to generate Canberra distances for nonmetric multidimensional scaling (NMDS) using the *vegdist* function. Canberra distances have performed well on datasets whose OTUs may be arranged in clusters as opposed to gradients (36). Unweighted pair group method with arithmetic mean

(UPGMA) clustering was used to group samples and create a dendrogram according to their community similarity using the function *pvclust::pvclust*. The dendrogram was pruned to the total number of host groups in the dataset (n=13) before plotting to remove edges linking different host groups. Variation in microbial communities attributable to host taxonomy was estimated using permutational multivariate analysis of variance using distance matrices (ADONIS).

Abundance-occupancy relationship. Occupancy was defined as the proportion of all host groups (n=13) in which the OTU was observed. Additionally, within-species occupancy was calculated for human and cattle datasets by calculating the proportion of samples in which an OTU was observed within like species. Hereinafter, *within-species occupancy* refers to the occupancy among individuals from a common host species group (either human or cattle). All abundance measures were estimated after summing counts for each OTU or phylum within each host species group. Two measures of abundance were used. One, the highest count observed within any host species group for each OTU was used to estimate maximum abundance. Two, the local mean for each OTU was estimated by averaging the OTU counts for host species in which the OTU was observed (i.e., unoccupied species were omitted). Additionally, within-species abundance measures were calculated analogous to within-species occupancy (described above) whereby the maximum abundance and local mean abundance were estimated using OTU distributions within samples instead of host species. The relationship between abundance and occupancy was investigated with both non-parametric (Loess) and parametric methods (simple linear regression).

One concern was that the tendency of abundant OTUs to be more easily detected could inflate their observed occupancies relative to rare OTUs resulting in what appeared to be an abundance-occupancy relationship, but would actually be an effect of ascertainment bias, or insufficient sampling of rare taxa. To assess the effects of ascertainment bias on the abundance-occupancy relationship in the human dataset, the average increase in observed occupancy was estimated over a series of random subsampling depths ranging from 500 to 7500 OTUs.

Multinomial Species Classification (CLAM test). Non-normalized community data were used to identify specialists using CLAM tests by successive pairwise group comparisons (37) resulting in the identification of OTUs that were specialists for each host species group. As part of the CLAM tests, sample coverage correction based on the number of observed singletons was applied to rare OTUs whose counts were below ten sequences per group. Relative OTU abundances were used above this threshold. After these corrections, OTUs were classified as specialists if ≥ 90% of their occurrences were within a specified group. A significance cut-off of 0.05 was used for individual tests. The specialists identified for each host group were used for descriptive analysis while OTUs classified as generalists, “too rare” to classify, or specialists outside the group of interest were disregarded in this analysis.

Results

General Dataset Description. General descriptions of human and cattle datasets are provided elsewhere (2, 4, 35). Normalization and exclusion of singletons for abundance-occupancy analysis resulted in discarding 52.9%, 56.6%, and 51.8% of vertebrate, human, and cattle OTUs, respectively. Diversity estimates indicate we observed roughly half of all OTUs present in samples (Table 2). Percentages of observed OTUs were lowest in gull samples (30.4%) and highest in horse samples (52.5%).

Community Taxonomic Composition. The taxonomic composition of host communities differed greatly between species; however, all species contained a large proportion of bacteria belonging to the *Firmicutes* phyla (Figure 1). *Firmicutes* composed the largest portion of bacterial communities in birds and mammals, while *Proteobacteria* composed the largest portion in fish. *Bacteroidetes* composed the sixth, second, and third most abundant groups in birds, mammals, and fish, respectively. Deer, dog, and horse bacterial communities stood out among all mammals mainly because of their unusually large proportions of *Proteobacteria*, *Fusobacteria*, and *Lentisphaerae*, respectively.

Microbiota Dissimilarity. NMDS plots arranged animal groups into distinct clusters that agreed well with hierarchical clustering (Figure 2). Samples of domestic or agricultural origin clustered by host type well while there was less agreement among bacterial communities sampled from wildlife. Despite cattle being the same species, diet largely determined cattle sample clustering by NMDS. ADONIS indicated that host taxonomic species and class were able to explain only 29.5% and 5.6% ($p<0.001$) of the variation in vertebrate microbial communities, respectively.

Abundance-Occupancy Relationship. Even after the exclusion of singletons and normalization, the distribution of OTUs was highly skewed towards occupancy in a single host (70.3% of OTUs appeared in a single host type only). Only four OTUs were found in at least one sample from each species: two unclassified *Enterobacteriaceae*, one *Ralstonia* OTU, and one *Clostridium* XI OTU. One concern was that data normalization obscured the presence of some OTUs by exclusion and artificially decreased observed occupancy; however, occupancy analysis with non-normalized counts suggested a similar, highly skewed trend with the majority of OTUs occupying a single host and only a few widespread OTUs (data not shown). In addition, 100% of OTUs present in more than one host species group with the non-normalized dataset were also identified in the normalized dataset confirming that normalization did not strongly influence occupancy estimates. Loess curves on plots of abundance versus occupancy suggested that the relationship between maximum OTU abundance and occupancy was positive and linear within a range (Figure 3). Assuming linearity, regression analysis indicated that the interspecific abundance-occupancy relationship among all OTUs within human, cattle, and other vertebrate datasets in the study was significantly positive ($p < 10^{-4}$) with a high degree of error ($R^2 < 0.1$). A similar trend was observed when local mean abundance was used as the abundance measure instead of maximum abundance (data not shown).

Of 75 OTUs with proportional occupancies ≥ 0.5 (observed in more than 6 host species) 62.7% were classified as *Firmicutes* and 34.7% were classified as *Proteobacteria*. Fifty were widely shared between the three host taxonomic classes and all 75 were shared between *Aves* and *Mammalia.*

The observed relationship between abundance and occupancy could not solely be due to ascertainment bias. Successive subsampling trials in the human dataset showed that for each log (base 10) increase in mean local abundance, there was an average increase in proportional occupancy of 0.5 and a similar increase of 0.3 when maximum abundance was used. In contrast, subsampling a log higher number of OTUs from 500 to 5,000 results in only about a 0.02 increase in proportional occupancy suggesting that increasing the depth of sequencing would result in only a small increase in observed occupancy—not enough to explain the observed relationship.

Another concern was that observed abundance-occupancy relationships could be due to over-clustering of distinct ecotypes within the same OTU. To investigate this possibility we performed an entirely separate analysis on OTUs clustered at the 99% similarity threshold in hopes of minimizing clustering of sequences that may have co-evolved in distant hosts. Because of the more stringent clustering criteria this process resulted in 124,563, 27,020, and 54,854 OTUs for the vertebrate, human, and cattle datasets, respectively. There was no discernable difference in the abundance-occupancy relationships between the two clustering thresholds (data not shown), suggesting that OTU clustering threshold parameters, at least those most commonly used, cannot explain our observations.

CLAM tests. CLAM tests on 54,666 OTUs resulted in the identification of 10,663 (19.4%) specialist OTUs (Table 3). Taking into account abundance, specialist OTUs accounted for 89..4% of all sequence reads. The taxonomic identities of specialists fell roughly in line with the overall community composition with the dominant taxa making up a large portion of the specialist population within each host group (Figure 4). Clustering OTUs at 99% instead of 97% similarity produced similar types and distributions of specialist OTUs (data not shown). *A priori* exclusion of OTUs via normalization prior to CLAM tests decreased the total number of OTUs identified as specialists (data not shown), presumably because the majority of OTUs veiled in the normalization process were at low abundance and low occupancy. This presumption was supported by both the low occupancy of a large portion of OTUs and the observation of a positive abundance-occupancy relationship.

Discussion

The rules governing the distribution of microbes have long been debated (38); and while there are trends shared not only between bacterial communities from different habitats, but also between macro- and microorganisms, the factors structuring the communities likely differ (39). Microbial communities native to the guts of animals are a special case because their current state may be strongly influenced by the ecology and evolution of their hosts. The degree to which microbes invest in particular host-specific lifestyles can be studied by asking how they fit well-studied macro-ecological patterns, if at all. Our work shows that intestinal bacteria present both within and among vertebrate host species follow a similar abundance-occupancy relationship, which we cannot precisely explain in light of previous explanations, but because the increase in occupancy as a function of abundance far outweighs that attributable to sampling depth, it is unlikely that the relationship is due solely to ascertainment bias. We also found that OTU clustering parameters had little effect on the abundance-occupancy relationships. Similar relationships have been observed previously in the human microbiome by comparison of rank-abundance and rank-prevalence (40). Although it is difficult to compare such relationships based on proportions of host species or individuals occupied to those based on ranges from other studies (e.g., latitudes, distances), the observation that similar patterns occur reinforce the idea that host-associated intestinal microbial communities may operate under a similar set of principles as ‘free-living’ communities to a degree (41).

Although some intestinal bacteria have developed mechanisms for survival under oxic and oligotrophic or otherwise harsh conditions, many are not fit for such conditions, limiting survival outside the host to a matter of days (42) and restricting re-colonization in distant suitable habitats (i.e., dispersal limitation). Isolation contributes to community dissimilarity through ecological drift (43). Selective pressure, most of which is mediated by the host immune system or other factors, such as diet and out-competition from highly specialized community members, can restrict successful colonization of the gut from outside members and further contribute to the isolation and divergence of these host-associated communities. The narrow range of abundant specialists suggests that host selection and drift through ecological isolation may have caused a significant portion of intestinal bacteria to deviate from the nearly universal abundance-occupancy relationship. In contrast to previous studies that show clear abundance-occupancy relationships in large organisms, which suggest a lack of a fitness trade-off between generalist and specialist niches, these results confirm that bacterial taxa can and have benefited significantly by acquiring specialist lifestyles.

Because the CLAM test is relatively robust to biases caused by different sampling depths between samples and stochastic sampling of rare taxa (37), we were able to identify thousands of specialist OTUs without the exclusion of a large amount of data *a priori* through normalization by subsampling. Typically, host-associated specialist taxa are identified through comparative 16S rRNA gene sequence analysis or enrichment methods (44-47) followed by testing for their presence in other sources with more sensitive methods, such as PCR, which can take years to complete (48-50). Although our methods, like most, cannot confirm the absence of specific taxa, the comparison of OTUs between multiple host-associated communities simultaneously resulted in the identification of both previously identified and potentially new specialist groups in a single step. Independent studies have also identified members of *Enterococcaceae* that dominate *Larus* spp. (gulls) sampled over a wide geographic range, but not found in other species at significant concentrations (50-52). The relative abundance of *Lachnospiraceae* in mammalian guts has been noted previously and genomic analysis suggests the group's ability to form endospores, produce butyrate, a compound thought to be important in host physiology, and encode genes important for protein interactions and signal transduction play prominently in the group's ability to evolve host-specific preferences (53). Similar mechanisms likely exist for other specialists taxa identified in this study. *Erysipelotrichaceae*, *Porphyromonadaceae*, and *Spirochaetaceae* may represent previously unidentified canine, porcine, and equine specialist groups, respectively. In dogs, the abundance of *Erysipelotrichaceae* drops significantly in diseased states while no significant change occurs for most other bacterial taxa (54), which suggests that canine OTUs within this group may have specialized not only to the canine gut environment, but also to a healthy host state within this environment. Such taxonomic groups identified by the CLAM test may represent potential host-associated targets for PCR- or sequencing-based (55) fecal pollution identification methods and further investigation into their distribution and growth/persistence in the environment is warranted.

There are many considerations and caveats when interpreting CLAM test results. Pre-treatment of the input data (e.g., normalization) and alternate user-defined values (e.g. statistical significance threshold) changed the number of specialist bacteria identified within each host. As noted by the authors of the test, normalization leads to a larger proportion of taxa classified as “too rare” to classify (37). The range, type, and physiological states of host species sampled and their grouping by the analyst (e.g. regarding or disregarding diet regimes) also influence the identification of specialist taxa. Future studies should be directed at describing this variation among hosts to an extent we could not achieve with such small sample sizes. These methods do not distinguish between specialists that have a high abundance within a small proportion of hosts from lower abundance specialists found in a large proportion hosts. Such information may be useful when trying to distinguish dispensable from essential community members or the degree of association between two organisms (56).

While such tests help prioritize bacterial groups for future study, a deeper understanding of the ecological and physiological roles that contribute to patterns of abundance and occupancy are needed to fully understand the extent of host-microbe relationships and to test the widespread assumption that the most abundant bacteria also play the most important physiological roles within the host. Functional metagenomic analysis may provide a more accurate picture of the overall community metabolic capability, while single cell isolation and genome sequencing techniques may be more useful in linking functional capacity to 16S rRNA data such as those produced in this study. A more detailed comparison of host genetics, perhaps through the comparison of mitochondrial genomes or whole nuclear genomes, may provide a host phylogenetic “landscape” on which to study the effects of other environmental factors such as host diet, habitat, or inter-population social interactions, on microbial communities.

Acknowledgements

Information has been subjected to U.S. EPA peer and administrative review and has been approved for external publication. Any opinions expressed in this paper are those of the authors

and do not necessarily reflect the official positions and policies of the U.S. EPA. Any mention of trade names or commercial products does not constitute endorsement or recommendation for use.

References

1. **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.
2. **Dethlefsen L, Huse S, Sogin ML, Relman DA.** 2008. The pervasive effects of an antibiotic on the human gut microbiota, as revealed by deep 16S rRNA sequencing. PLoS Biol **6:**e280.
3. **Ley RE, Backhed F, Turnbaugh P, Lozupone CA, Knight RD, Gordon JI.** 2005. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A **102:**11070-11075.
4. **Shanks OC, Kelty CA, Archibeque S, Jenkins M, Newton RJ, McLellan SL, Huse SM, Sogin ML.** 2011. Community structures of fecal bacteria in cattle from different animal feeding operations. Appl Environ Microbiol **77:**2992-3001.
5. **Shanks OC, Kelty CA, Peed L, Sivaganesan M, Mooney T, Jenkins M.** 2014. Age-related shifts in the density and distribution of genetic marker water quality indicators in cow and calf feces. Appl Environ Microbiol **80:**1588-1594.
6. **Murphy EF, Cotter PD, Hogan A, O'Sullivan O, Joyce A, Fouhy F, Clarke SF, Marques TM, O'Toole PW, Stanton C, Quigley EM, Daly C, Ross PR, O'Doherty RM, Shanahan F.** 2013. Divergent metabolic outcomes arising from targeted manipulation of the gut microbiota in diet-induced obesity. Gut **62:**220-226.
7. **Tremaroli V, Backhed F.** 2012. Functional interactions between the gut microbiota and host metabolism. Nature **489:**242-249.
8. **McFall-Ngai M, Hadfield MG, Bosch TC, Carey HV, Domazet-Loso T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealson K, Pierce NE, Rawls JF, Reid A, Ruby EG, Rumpho M, Sanders JG, Tautz D, Wernegreen JJ.** 2013. Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A **110:**3229-3236.
9. **Ridaura VK, Faith JJ, Rey FE, Cheng J, Duncan AE, Kau AL, Griffin NW, Lombard V, Henrissat B, Bain JR, Muehlbauer MJ, Ilkayeva O, Semenkovich CF, Funai K, Hayashi DK, Lyle BJ, Martini MC, Ursell LK, Clemente JC, Van Treuren W, Walters WA, Knight R, Newgard CB, Heath AC, Gordon JI.** 2013. Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science **341:**1241214.
10. **Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI.** 2012. Human gut microbiome viewed across age and geography. Nature **486:**222-227.
11. **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.

12. **Rawls JF, Mahowald MA, Ley RE, Gordon JI.** 2006. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. Cell **127:**423-433.

13. **Vrieze A, Van Nood E, Holleman F, Salojarvi J, Kootte RS, Bartelsman JF, Dallinga-Thie GM, Ackermans MT, Serlie MJ, Oozeer R, Derrien M, Druesne A, Van Hylckama Vlieg JE, Bloks VW, Groen AK, Heilig HG, Zoetendal EG, Stroes ES, de Vos WM, Hoekstra JB, Nieuwdorp M.** 2012. Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. Gastroenterology **143:**913-916.e917.

14. **Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman DA.** 2012. The application of ecological theory toward an understanding of the human microbiome. Science **336:**1255-1262.

15. **Martens EC, Chiang HC, Gordon JI.** 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host Microbe **4:**447-457.

16. **Xu J, Bjursell MK, Himrod J, Deng S, Carmichael LK, Chiang HC, Hooper LV, Gordon JI.** 2003. A Genomic View of the Human-Bacteroides thetaiotaomicron Symbiosis. Science **299:**2074-2076.

17. **Horner-Devine MC, Lage M, Hughes JB, Bohannan BJ.** 2004. A taxa-area relationship for bacteria. Nature **432:**750-753.

18. **Martiny JB, Eisen JA, Penn K, Allison SD, Horner-Devine MC.** 2011. Drivers of bacterial beta-diversity depend on spatial scale. Proc Natl Acad Sci U S A **108:**7850-7854.

19. **Bell T.** 2010. Experimental tests of the bacterial distance-decay relationship. ISME J **4:**1357-1365.

20. **Bell T, Ager D, Song JI, Newman JA, Thompson IP, Lilley AK, van der Gast CJ.** 2005. Larger islands house more bacterial taxa. Science **308:**1884.

21. **Östman O, Drakare S, Kritzberg ES, Langenheder S, Logue JB, Lindstrom ES.** 2010. Regional invariance among microbial communities. Ecol Lett **13:**118-127.

22. **Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P, Tunlid A, Hagstrom A.** 2007. Global patterns of diversity and community structure in marine bacterioplankton. Mol Ecol **16:**867-880.

23. **Amend AS, Oliver TA, Amaral-Zettler LA, Boetius A, Fuhrman JA, Horner-Devine MC, Huse SM, Welch DBM, Martiny AC, Ramette A, Zinger L, Sogin ML, Martiny JBH, Lambshead J.** 2013. Macroecological patterns of marine bacteria on a global scale. Journal of Biogeography **40:**800-811.

24. **Fulthorpe RR, Roesch LF, Riva A, Triplett EW.** 2008. Distantly sampled soils carry few species in common. Isme j **2:**901-910.

25. **Spain AM, Krumholz LR, Elshahed MS.** 2009. Abundance, composition, diversity and novelty of soil Proteobacteria. Isme j **3:**992-1000.

26. **Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, Knight R.** 2011. Global patterns in the biogeography of bacterial taxa. Environ Microbiol **13:**135-144.

27. **Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences **103:**12115-12120.

28. **Huse SM, Mark Welch DB, Voorhis A, Shipunova A, Morrison HG, Eren AM, Sogin ML.** 2014. VAMPS: a website for visualization and analysis of microbial population structures. BMC Bioinformatics **15:**41.

29. **Huse S, Huber J, Morrison H, Sogin M, Welch D.** 2007. Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology **8:**R143.

30. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics doi:10.1093/bioinformatics/btq461.

31. **Wang Q, Garrity GM, Tiedje JM, Cole JR.** 2007. Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol **73:**5261-5267.

32. **Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO.** 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research **41:**D590-D596.

33. **Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Wagner H.** 2013. vegan: community ecology package, vR package version 2.0-10. http://CRAN.R-project.org/package=vegan.

34. **Suzuki R, Shimodaira H.** 2011. pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling., vR package version 1.2-2. http://CRAN.R-project.org/package=pvclust.

35. **Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI.** 2009. A core gut microbiome in obese and lean twins. Nature **457:**480-484.

36. **Kuczynski J, Liu Z, Lozupone C, McDonald D, Fierer N, Knight R.** 2010. Microbial community resemblance methods differ in their ability to detect biologically relevant patterns. Nat Methods **7:**813-819.

37. **Chazdon RL, Chao A, Colwell RK, Lin S-Y, Norden N, Letcher SG, Clark DB, Finegan B, Arroyo JP.** 2011. A novel statistical method for classifying habitat generalists and specialists. Ecology **92:**1332-1343.

38. **de Wit R, Bouvier T.** 2006. 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? Environ Microbiol **8:**755-758.

39. **Martiny JB, Bohannan BJ, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreas L, Reysenbach AL, Smith VH, Staley JT.** 2006. Microbial biogeography: putting microorganisms on the map. Nat Rev Microbiol **4:**102-112.

40. **Huse SM, Ye Y, Zhou Y, Fodor AA.** 2012. A core human Microbiome as viewed through 16S rRNA sequence clusters. PLoS ONE **7:**e34242.

41. **Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI.** 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat Rev Microbiol **6:**776-788.

42. **Green HC, Shanks OC, Sivaganesan M, Haugland RA, Field KG.** 2011. Differential decay of human faecal Bacteroides in marine and freshwater. Environ Microbiol **13:**3235-3249.

43. **Condit R, Pitman N, Leigh EG, Jr., Chave J, Terborgh J, Foster RB, Nunez 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.

44. **Dick LK, Simonich MT, Field KG.** 2005. Microplate subtractive hybridization to enrich for *Bacteroidales* genetic markers for fecal source identification. Appl Environ Microbiol **71:**3179-3183.

45. **Shanks OC, Santo Domingo JW, Lamendella R, Kelty CA, Graham JE.** 2006. Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. Appl Environ Microbiol **72:**4054-4060.

46. **Shanks OC, Domingo JW, Lu J, Kelty CA, Graham JE.** 2007. Identification of bacterial DNA markers for the detection of human fecal pollution in water. Appl Environ Microbiol **73:**2416-2422.

47. **Green HC, White KM, Kelty CA, Shanks OC.** 2014. Development of rapid canine fecal source identification PCR-based assays. Environ Sci Technol **48:**11453-11461.

48. **Bernhard AE, Field KG.** 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl Environ Microbiol **66:**1587-1594.

49. **Shanks OC, Kelty CA, Sivaganesan M, Varma M, Haugland RA.** 2009. Quantitative PCR for genetic markers of human fecal pollution. Appl Environ Microbiol **75:**5507-5513.

50. **Green HC, Dick LK, Gilpin B, Samadpour M, Field KG.** 2012. Genetic markers for rapid PCR-based identification of gull, Canada goose, duck, and chicken fecal contamination in water. Appl Environ Microbiol **78:**503-510.

51. **Lu J, Santo Domingo JW, Lamendella R, Edge T, Hill S.** 2008. Phylogenetic diversity and molecular detection of bacteria in gull feces. Appl Environ Microbiol **74:**3969-3976.

52. **Koskey AM, Fisher JC, Traudt MF, Newton RJ, McLellan SL.** 2014. Analysis of the gull fecal microbial community reveals the dominance of Catellicoccus marimammalium in relation to culturable Enterococci. Appl Environ Microbiol **80:**757-765.

53. **Meehan CJ, Beiko RG.** 2014. A phylogenomic view of ecological specialization in the Lachnospiraceae, a family of digestive tract-associated bacteria. Genome Biol Evol **6:**703-713.

54. **Suchodolski JS, Markel ME, Garcia-Mazcorro JF, Unterer S, Heilmann RM, Dowd SE, Kachroo P, Ivanov I, Minamoto Y, Dillman EM, Steiner JM, Cook AK, Toresson L.** 2012. The fecal microbiome in dogs with acute diarrhea and idiopathic inflammatory bowel disease. PLoS ONE **7:**e51907.

55. **Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, Sogin ML.** 2013. Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol Evol **4**.

56. **Bluthgen N, Menzel F, Hovestadt T, Fiala B, Bluthgen N.** 2007. Specialization, constraints, and conflicting interests in mutualistic networks. Curr Biol **17:**341-346.

57. **Dethlefsen L, Eckburg PB, Bik EM, Relman DA.** 2006. Assembly of the human intestinal microbiota. Trends Ecol Evol **21:**517-523.

TABLES

Table 1. Fecal samples used in study.

Common name	Scientific name	Location	n	Reference
Chicken	*Gallus gallus domesticus*	Georgia	5	(This study)
Cow	*Bos taurus*	Colorado, Georgia, Ohio, Nebraska	30	(4)
Deer	*Odocoileus virginianus*	Wyoming	4	(This study)
Dog	*Canis lupus familiaris*	Ohio	4	(This study)
Duck	*Anas platyrhynchos*	Ohio	3	(This study)
Goose	*Branta canadensis*	Ohio	3	(This study)
Gull	*Larus delawarensis*	Wisconsin	4	(This study)
Horse	*Equus ferus caballus*	Georgia	5	(This study)
Human	*Homo sapiens*	*	36	(35, 57)
Perch	*Perca flavescens*	Wisconsin	4	(This study)
Swine	*Sus scrofa domesticus*	Georgia	5	(This study)
Trout	*Oncorhynchus mykiss*	Wisconsin	3	(This study)

*Sample origin not in original manuscripts

Table 2. Pooled diversity measures in the vertebrate dataset.

	n_reads	OTUs	Chao	Chao.se	Per
Chicken	44863	1829	5332	311	34.3
Cow_F	208473	11358	25100	508	45.3
Cow_UG	210593	7465	18817	493	39.7
Deer	79540	5448	13066	351	41.7
Dog	117715	2615	6896	302	37.9
Duck	60795	3128	9486	412	33.0
Goose	58483	3740	10298	385	36.3
Gull	161585	3859	12712	532	30.4
Horse	145594	10100	19251	339	52.5
Human	372632	8516	24921	660	34.2
Perch	148380	2184	6249	324	35.0
Swine	106881	6443	13120	307	49.1
Trout	60461	993	3101	258	32.0

“_F” and “_UG” represent cattle fed forage and unprocessed grain, respectively.

Table 3. Number of specialist OTUs within each host group belonging to each bacterial phyla.

	Host Species (Common name)												
	Aves				***Mammalia***							***Actinopterygii***	
	Chicken	Duck	Goose	Gull	Cow_F	Cow_UG	Deer	Dog	Horse	Human	Swine	Perch	Trout
Acidobacteria	0	44	0	7	9	7	6	8	28	0	15	1	1
Actinobacteria	21	49	183	191	169	30	38	9	111	48	64	14	3
Bacteroidetes	10	22	50	32	193	288	120	23	328	67	140	18	11
Chlamydiae	1	0	2	2	10	0	3	0	5	1	2	0	0
Chloroflexi	1	0	1	10	3	1	3	0	21	1	3	0	0
Chrysiogenetes	0	0	4	0	8	1	3	0	4	1	0	3	1
Deferribacteres	0	2	2	0	6	1	0	1	9	1	1	1	0

Deinococcus-Thermus	0	0	0	4	9	2	0	0	4	0	5	0	0
Firmicutes	169	147	302	324	1211	375	319	287	1085	611	484	182	46
Fusobacteria	0	4	1	3	4	1	2	28	9	0	2	2	2
Lentisphaerae	0	0	0	0	4	2	4	0	9	1	2	0	0
Planctomycetes	2	5	0	32	6	4	1	0	38	1	3	1	0
Proteobacteria	54	109	82	264	400	78	132	14	474	63	117	142	38
Spirochaetes	0	1	1	2	20	4	5	1	30	1	14	0	0
Synergistetes	1	1	1	3	7	2	1	0	6	1	0	0	0
Tenericutes	0	1	3	2	49	9	19	1	48	4	26	0	0
Verrucomicrobia	3	13	0	17	10	3	12	4	34	1	4	0	0

Aquificae, *Armatimonadetes*, BRC1, *Caldiserica*, *Chlorobi*, *Elusimicrobia*, *Fibrobacteres*, *Gemmatimonadetes*, *Nitrospira*, OP11, *Thermosulfobacteria*, *Thermotogae*, TM7, and WS3 were represented by 1-20 specialist OTUs and were omitted from the table. "_F" and "_UG" represent cattle fed forage and unprocessed grain, respectively.

FIGURE LEGENDS

Figure 1. Phylum-level taxonomic composition of host-associated intestinal microbial communities. The total height of each stacked bar corresponds to all reads from a sample while shorter, color-coded bars correspond to the proportion those reads falling within major bacterial phyla. "Cow_UG" and "Cow_F" indicate cattle fed unprocessed grain and forage, respectively.

Figure 2. NMDS plot of samples based on microbial community profiles (stress=0.158). A) Samples connected by lines were collected from the same population. B) UPGMA host species group clustering over-layed onto NMDS ordination. Both underlying ordinations are identical and were solved using all OTUs in the vertebrate dataset. UPGMA tree was pruned to the number of host species groups. "Cow_UG" and "Cow_F" indicate cattle fed unprocessed grain and forage, respectively. Perch and trout cluster on top of one another.

Figure 3. The interspecific abundance-occupancy relationship in vertebrate, human, and cattle microbial communities. Within-species occupancy is represented on the x-axis for human and cattle datasets while occupancy is represented on the x-axis for the vertebrate dataset. Lowess curves were estimated using all data from each dataset while regression lines ("Linear mod") were estimated after exclusion of single-occupancy OTUs. Blue shading represents the two-dimensional kernel density of the data. Artificial variance was added post-lowess and -regression analysis on x-axes for plot clarity.

Figure 4. CLAM test results for A) Fimicutes and B) Bacteroidetes over-layed onto a NMDS ordination (stress=0.158). Both underlying ordinations are identical and were solved using all OTUs in the vertebrate dataset. Species bubbles are placed according to their abundance weighted average ordination scores and is descriptive of which host species they are most associated with. The five families containing the most specialist OTUs within each phylum are

displayed. Numerical values in the lower left legend represent the total number of specialist OTUs within each family. The diameter of each filled circle is proportional to specialist OTU relative abundance as a proportion of all summed OTU counts within each host group.