Evolution of Mammals and Their Gut Microbes
Ruth E. Ley, *et al.*
Science **320**, 1647 (2008);
DOI: 10.1126/science.1155725

19. G. G. Simpson, P. P. Dijkwel, V. Quesada, I. Henderson, C. Dean, *Cell* **113**, 777 (2003).
20. K. L. Veraldi *et al*., *Mol. Cell. Biol.* **21**, 1228 (2001).
21. J. Lu *et al*., *Nature* **435**, 834 (2005).
22. M. S. Kumar, J. Lu, K. L. Mercer, T. R. Golub, T. Jacks, *Nat. Genet.* **39**, 673 (2007).
23. We thank members of the Burge and Sharp labs as well as M. Winslow, K. Cante-Barrett, and O. Larsson. Supported by the Knut and Alice Wallenberg Foundation (R.S.); the Cancer Research Institute (J.R.N.); the Gina De Felice and Robert M. Lefkowitz (1975) Fund (A.S.); U.S. Public Health Service grant RO1-GM34277 and National Cancer Institute grants PO1-CA42063 and U19 AI056900 (P.A.S.); Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute; and National Human Genome Research Institute grant R01-HG002439 (C.B.B.). Array data have been deposited in Gene Expression Omnibus (accession number GSE10666).

Supporting Online Material
www.sciencemag.org/cgi/content/full/320/5883/1643/DC1
Materials and Methods
SOM Text
Figs. S1 to S10
Tables S1 to S10
References

18 January 2008; accepted 19 May 2008
10.1126/science.1155390

Evolution of Mammals and Their Gut Microbes

Ruth E. Ley,[1] Micah Hamady,[2] Catherine Lozupone,[1,3] Peter J. Turnbaugh,[1] Rob Roy Ramey,[4] J. Stephen Bircher,[5] Michael L. Schlegel,[6] Tammy A. Tucker,[6] Mark D. Schrenzel,[6] Rob Knight,[3] Jeffrey I. Gordon[1]*

Mammals are metagenomic in that they are composed of not only their own gene complements but also those of all of their associated microbes. To understand the coevolution of the mammals and their indigenous microbial communities, we conducted a network-based analysis of bacterial 16*S* ribosomal RNA gene sequences from the fecal microbiota of humans and 59 other mammalian species living in two zoos and in the wild. The results indicate that host diet and phylogeny both influence bacterial diversity, which increases from carnivory to omnivory to herbivory; that bacterial communities codiversified with their hosts; and that the gut microbiota of humans living a modern life-style is typical of omnivorous primates.

Our "metagenome" is a composite of *Homo sapiens* genes and genes present in the genomes of the trillions of microbes that colonize our adult bodies (*1*). The vast majority of these microbes live in our distal guts. "Our" microbial genomes (microbiomes) encode metabolic functions that we have not had to evolve wholly on our own, including the ability to extract energy and nutrients from our diet. It is unclear how distinctively human our gut microbiota is, or how modern *H. sapiens*' ability to construct a wide range of diets has affected our gut microbial ecology. Here, we address two general questions concerning the evolution of mammals: How do diet and host phylogeny shape mammalian microbiota? When a mammalian species acquires a new dietary niche, how does its gut microbiota relate to the microbiota of its close relatives?

The acquisition of a new diet is a fundamental driver for the evolution of new species. Coevolution, the reciprocal adaptations occurring between interacting species (*2*), produces physiological changes that are often recorded in fossil remains. For instance, although mammals made their first appearance on the world stage in the Jurassic [~160 million years ago (Ma)], most modern species arose during the Quaternary [1.8 Ma to the present (*3*)], when C4 grasslands (dominated by plants that use for photosynthesis the Hatch-Slack cycle rather than the Calvin cycle typical of C3 plants) expanded in response to a fall in atmospheric CO_2 levels and/or climate changes (*4*–*6*). The switch to a C4 plant–dominated diet led to selection for herbivores with high-crowned teeth (*7*, *8*) and longer gut retention times necessary for the digestion of lower-quality forage (*9*). However, these adaptations may not suffice for the exploitation of a new dietary niche. The community of microbes in the gut constitutes a potentially critical yet unexplored component of diet-driven speciation.

Because we cannot interrogate extinct gut microbiotas directly, past evolutionary processes can only be inferred from comparative analyses of extant mammalian gut microbial communities. Therefore, we analyzed the fecal microbial communities of 106 individual mammals representing 60 species from 13 taxonomic orders, including 17 nonhuman primates. To isolate the effects of phylogeny and diet, we included multiple samples from many of the mammalian species, as well as species that had unusual diets compared to their close phylogenetic relatives. For example, the majority of the nonhuman primate species studied were omnivores (12 of 17), but the leaf-eating (folivorous) East Angolan colobus, Eastern black-and-white colobus, Douc langur, and François langur were also sampled. In addition, the herbivorous giant panda and red panda were included from the Carnivora. Most animals were housed at the San Diego Zoo and the San Diego Zoo's Wild Animal Park (n = 15) or the St. Louis Zoo (n = 56). Others were examined in the wild (n = 29) or domesticated (n = 6; table S1). To test the reproducibility of host species–associated gut microbiotas and to gauge the effects of animal provenance, we represented mammalian species by multiple individuals from multiple locales where possible, and chose wild animals to match captive animals. We generated a data set of >20,000 16*S* rRNA gene sequences; for comparison of the human, primate, and nonprimate mammalian gut microbiotas (*10*), the 106 samples also included published fecal bacterial 16*S* rRNA sequences (>3000) from wild African gorilla (*11*), Holstein cattle (*12*), Wistar rats (*13*), and healthy humans of both sexes, ranging in age from 27 to 94, living on three continents and including a strict vegetarian (*14*–*18*) (table S1).

We used network-based analyses to map gut microbial community composition and structure onto mammalian phylogeny and diet, thereby complementing phylogeny-based microbial community comparisons. These analyses were used to bin 16*S* rRNA gene sequences into operational taxonomic units (OTUs) and to display microbial genera partitioning across hosts. Genus-level OTUs (sets of sequences with ≥96% identity) and animal hosts were designated as nodes in a bipartite network, in which OTUs are connected to the hosts in which their sequences were found (Fig. 1A). To cluster the OTUs and hosts in this network, we used the stochastic spring-embedded algorithm, as implemented in Cytoscape 2.5.2 (*19*), where nodes act as physical objects that repel each other, and connections act as a spring with a spring constant and a resting length; the nodes are organized in a way that minimizes forces in the network.

The ensemble of sequences in this study provides an overarching view of the mammal gut microbiota. We detected members of 17 phyla (divisions) of Bacteria (*10*). The majority of sequences belong to the Firmicutes [65.7% of 19,548 classified sequences (*10*)] and to the Bacteroidetes (16.3%); these phyla were previously shown to constitute the majority of sampled human (and mouse) gut-associated phylotypes (*10*, *20*). The other phyla represented were the Proteobacteria (8.8% of all sequences collected; 85% in the Gamma subdivision), Actinobacteria (4.7%), Verrucomicrobia (2.2%), Fusobacteria (0.67%), Spirochaetes (0.46%), DSS1 (0.35%), Fibrobacteres (0.13%), TM7 (0.13%), deep-

[1]Center for Genome Sciences, Washington University School of Medicine, St. Louis, MO 63108, USA. [2]Department of Computer Science, University of Colorado, Boulder, CO 80309, USA. [3]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA. [4]Wildlife Science International Inc., Nederland, CO 80466, USA. [5]St. Louis Zoological Park, St. Louis, MO 63110, USA. [6]Zoological Society of San Diego, San Diego, CA 92112, USA.

*To whom correspondence should be addressed. E-mail: jgordon@wustl.edu

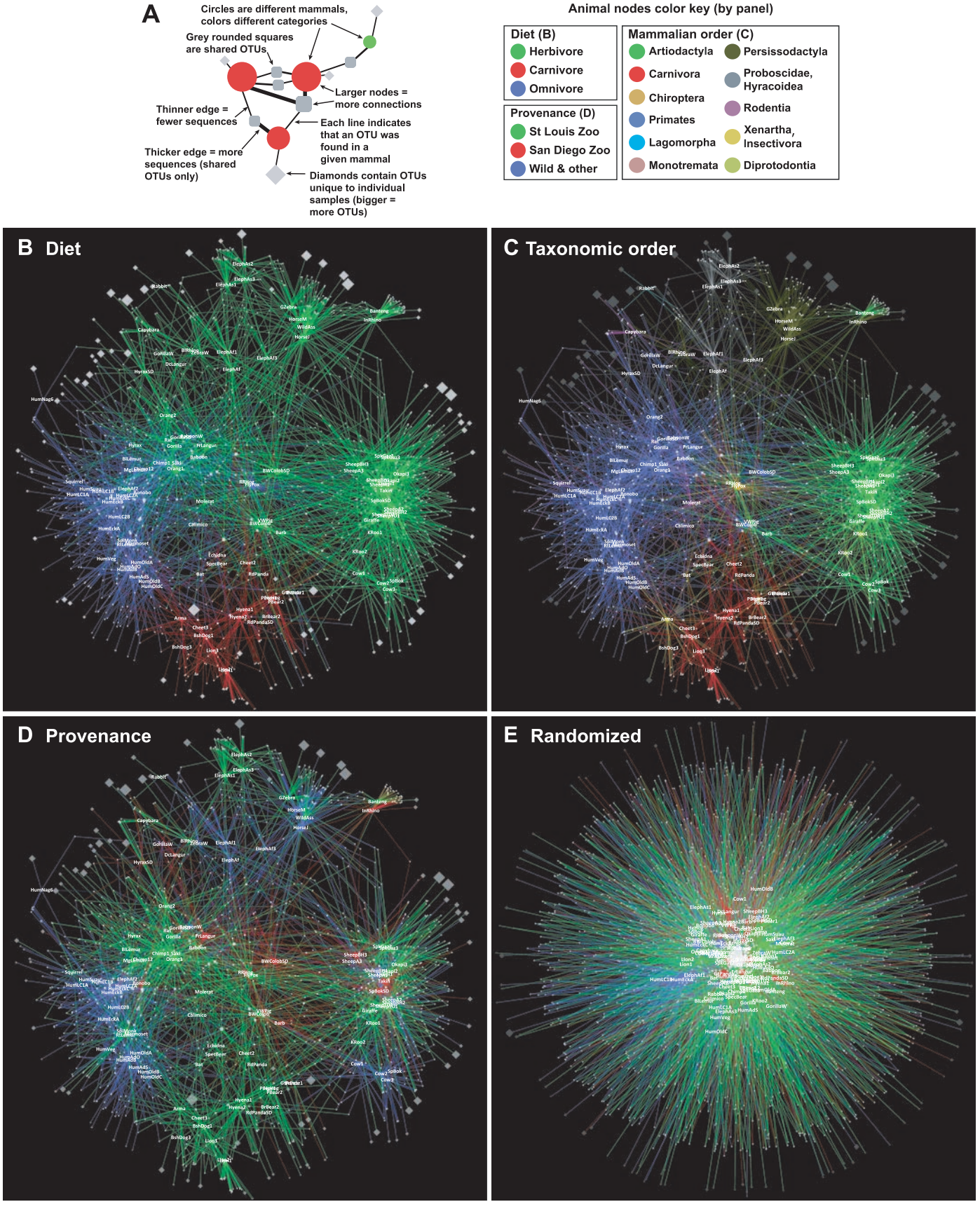

Fig. 1 (opposite page). Network-based analyses of fecal bacterial communities in 60 mammalian species. (**A**) Simplified cartoon illustration of a host-gut microbe network. (**B** to **E**) Network diagrams are color-coded by diet (B), animal taxonomy (C), or animal provenance (D), or represent randomized assignments of OTUs to animal nodes (E). Abbreviations used for animal species (asterisk denotes wild): Asian elephants, ElephAs1–3; baboons, Baboon, *BaboonW; African elephants, *ElephAf1–4; Bwindi gorilla, *GorillaW; Hartmann's mountain zebra, *ZebraW; armadillo, Arma; Argali sheep, *SheepA1–3; babirusa, Barb; Seba's short-tailed bat, Bat; American black bears, BrBear1, 2; bush dogs, BshDog1, 3; banteng, Banteng; bighorn sheep, *SheepBH1, 2 (BH3 not wild); black lemur, BlLemur; bonobo, Bonobo; calimicos (Goeldi's marmoset), Calimico; capybara, Capybara; cheetahs, Cheet2, 3; chimpanzees, Chimp1, 2; Eastern black-and-white colobus, BWColob; East Angolan colobus, BWColobSD; cattle, Cow1–3; Douc langur, DcLangur; echidna, Echidna; flying fox, FlyFox; François langur, FrLangur; giraffe, Giraffe; Western lowland gorillas, Gorilla, GorillaSD; giant panda, GtPanda; Geoffrey's marmoset, Marmoset; Grevy's zebra, GZebra; humans, HumAdB, HumAdO, HumAdS, HumEckA, HumEckB, HumEckC, HumNag6, HumOldA, HumOldB, HumOldC, HumSuau, HumVeg, HumLC1A, HumLC1B, HumLC2A, HumLC2B; hedgehog, HgHog; horses, Horse], HorseM; rock hyraxes, Hyrax, HyraxSD; spotted hyenas, Hyena1, 2; Indian rhinoceros, InRhino; red kangaroos, KRoo1, 2; lions, Lion1–3; mongoose lemur, MgLemur; naked mole rat, Molerat; okapi, Okapi1–3; orangutans, Orang1, 2; polar bears, PBear1, 2; rabbit, Rabbit; Norway rat (Wistar), Rat; black rhinoceros, BlRhino; red pandas, RdPanda, RdPandaSD; Red River hog, RRHog; ring-tailed lemur, RtLemur; white-faced saki, Saki; springboks, *SpBok, SpBokSD; spectacled bear, SpecBear; Speke's gazelles, SpkGaz2, 3; Prevost's squirrel, Squirrel; spider monkey, SpiMonk; takin, Takin; Transcaspian Urial sheep, SheepTU1, 2; Visayun warty pig, VWPig; Somali wild ass, WildAss. See table S1 for additional details.

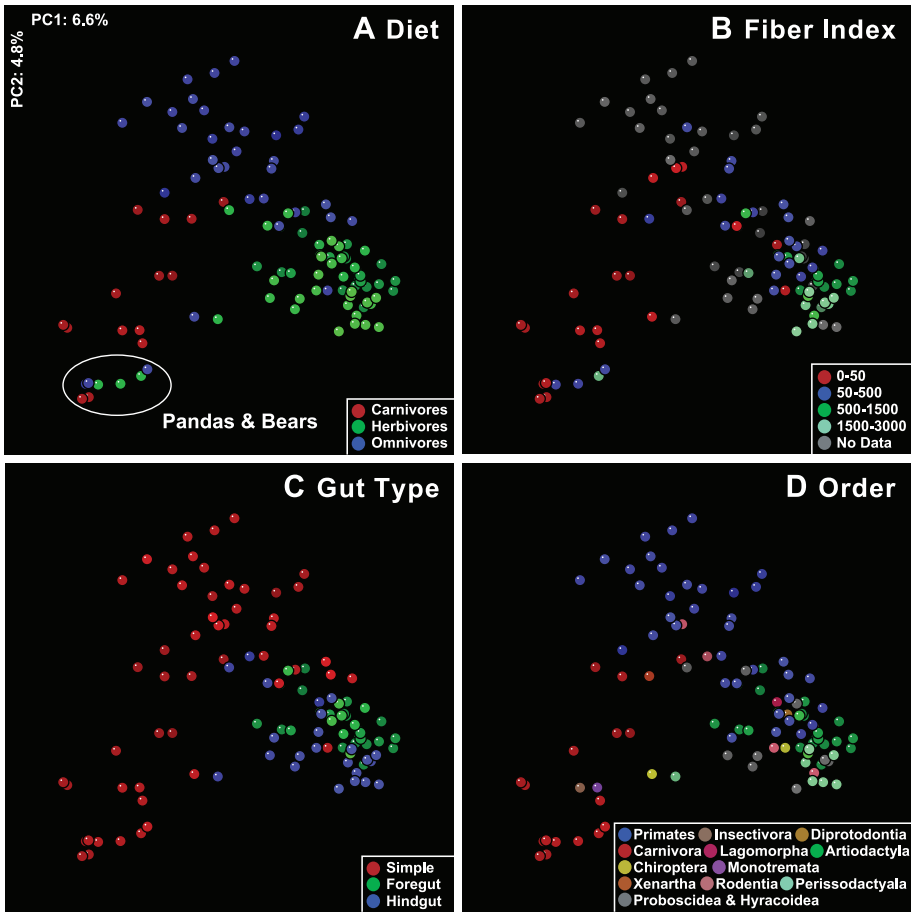

Fig. 2. Mammalian fecal bacterial communities clustered using principal coordinates analysis (PCoA) of the UniFrac metric matrix. PC1 and PC2 are plotted on *x* and *y* axes. Each circle corresponds to a fecal sample colored according to (**A**) diet, (**B**), diet fiber index, (**C**), gut morphology/physiology, and (**D**) host taxonomic order. The same data (samples) are shown in each panel. The percentage of the variation explained by the plotted principal coordinates is indicated on the axes.

rooting Cyanobacteria [0.10%; these are not chloroplasts (*20*)], Planctomycetes (0.08%), Deferribacteres (0.05%), Lentisphaerae (0.04%), and Chloroflexi, SR1, and Deinoccus-Thermus (all 0.005%). We were unable to assign 1985 16*S* rRNA gene sequences that passed a chimera-checking algorithm (*21*) to known phyla on the basis of BLAST searches against the Greengenes database (*22*) and the Ribosomal Database Project taxonomy annotations (*23*). Of the phyla that were detected, only Firmicutes were found in all samples (fig. S1). However, each mammalian host harbored OTUs (96% sequence identity) not observed in any other sample (at this level of sampling, on average, 56% and 62% of OTUs were unique within a sample and species, respectively; table S1).

The network-based analyses disclosed that overall, the fecal microbial communities of same-species (conspecific) hosts were more similar to each other than to those of different host species: Host nodes were significantly more connected within than between species (G test for independence, $G = 11.9$, $P = 0.0005$; Fig. 1B). Shown in fig. S2 is a tree-based analysis where similarity is defined using the UniFrac metric; this metric is based on the degree to which individual communities share branch length on a common (master) phylogenetic tree constructed from all 16*S* rRNA sequences from all communities being compared (*24*, *25*). The results are consistent with the network-based analysis; that is, they show that UniFrac distances are smaller within conspecific hosts than between nonconspecific hosts ($P < 0.005$ by one-tailed t test, confirmed by matrix permutation and corrected for multiple comparisons).

The impact of host species on community composition is most evident when considering conspecific hosts living separately, because cohousing may confound any species effect. For example, the two Hamadryas baboons clustered together (fig. S2), although one is from Namibia and the other from the St. Louis Zoo; similarly, the red pandas housed in different zoos clustered together. All 16 human samples also clustered together. Nonetheless, some conspecifics with different origins did not cluster (e.g., the two Western lowland gorillas), which suggests that diet and other environmental exposures ["legacy effects" (*26*)] play roles in addition to host phylogeny (taxonomic order).

The clustering by diet (herbivore, omnivore, and carnivore) was highly significant in both the tree-based (fig. S2) and network-based analyses (Fig. 1B). In the network-based analysis, host nodes are significantly more connected to other host nodes from the same diet group ($G = 115.8$; $P = 5.1 \times 10^{-27}$) (*10*). Similarly, hosts within the same taxonomic order are more connected in the network to hosts within the same order (Fig. 1C; $G = 356$; $P = 2.1 \times 10^{-79}$). Likewise, UniFrac-based principal coordinates analysis (PCoA) showed clustering by diet (Fig. 2, A and B) and by taxonomic order (Fig. 2D). (UniFrac distances are smaller for within versus between diet categories, and for within versus between orders, $P < 0.005$.) There was no significant clustering according to the provenance of the animals (including humans) in either the network- or UniFrac-based analyses ($P > 0.05$ for both; Fig. 1D and fig. S2, respectively), nor in a randomized network (Fig. 1E).

Classification of the mammals into herbivore, omnivore, and carnivore groups was based on diet records and natural history. Heavy iso-

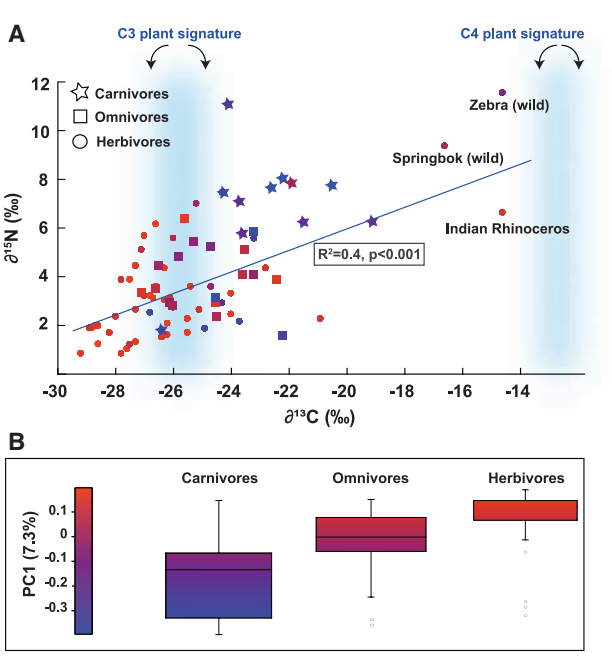

Fig. 3. Markers of trophic level mapped onto the variance in fecal microbial community diversity. (**A**) Stable isotope values for C and N plotted for each fecal sample, presented according to diet group. Symbols are colored according to their PC1 value; PC1 is the first principal coordinate of the PCoA of the unweighted UniFrac metric. $\delta^{13}C$ ranges for C3 and C4 plants [per mil (‰)] are highlighted in blue. R^2 is for $\delta^{13}C$ versus $\delta^{15}N$. (**B**) Box plots are shown for the three diet groups (central line is the mean; box outline equals 1 SD; the bar denotes 2 SD; circles are outliers). The majority of fecal $\delta^{13}C$ values are intermediate between the average for C4 plants (–12.5%) and C3 plants (–26.7%).

topes of carbon and nitrogen bioaccumulate in the food chain (*27*). Therefore, to obtain a more objective marker of diet, we measured stable isotope ratios of carbon and nitrogen, $\delta^{13}C$ and $\delta^{15}N$, in the feces (where $\delta = 1000 \times [(R_{sample} - R_{standard})/R_{standard}]$ and R = ratio of atom percentages $^{13}C/^{12}C$ and $^{15}N/^{14}N$). The results were consistent with the original diet group classification. Heavy isotopes were enriched in the order herbivore < omnivore < carnivore (Fig. 3A). The protein and fat contents of the diets of animals in captivity (obtained from diet records) were positively correlated with $\delta^{13}C$ and $\delta^{15}N$ fecal values (R^2 values for fat versus $\delta^{13}C$ and $\delta^{15}N$ were 0.51 and 0.45, respectively, and for protein, 0.36 and 0.38).

To test for a direct link between diet and microbial community composition, we mapped stable isotope values onto the coordinates that explained the largest proportion of the variance in the microbial communities, as determined by PCoA of the UniFrac distances between hosts (Fig. 3B). Principal coordinate 1 (PC1) separates carnivores from herbivores and omnivores (mean is significantly lower for carnivores than herbivores, which are equivalent to omnivores; $F_{80,2} = 9.9$, $P < 0.001$) and is also correlated with $\delta^{13}C$ and $\delta^{15}N$ values (multiple regression $R^2 = 0.25$, $F_{80,2} = 12.7$, $P < 0.001$). Together, these results support an association between microbial community membership and diet, and provide an independent validation of the dietary clustering observed in the network diagrams that is free of bias in assigning hosts to one of the three diet categories.

Underlying the correlation between bacterial community composition and diet is the partitioning of bacterial phyla among hosts according to diet. Herbivore microbiotas contained the most phyla (14), carnivores contained the fewest (6), and omnivores were intermediate (12) (fig. S1). Phylogenetic trees constructed from 16*S* rRNA sequences from the feces of herbivores also had the greatest amount of total branch length (phylogenetic diversity; fig. S3A). Consistent with this finding, herbivores had the highest genus-level richness, followed by omnivores and carnivores (fig. S3B).

Ancestral mammals were carnivores (*9*). We used an analysis based on the Fitch parsimony algorithm (*10*) to test whether bacterial lineages found in herbivores were derived from lineages found in carnivores. The results did not support this notion; hence, gut bacterial communities required to live largely on a plant-based diet were likely acquired independently from the environment.

Adaptation to a plant-based diet was an evolutionary breakthrough in mammals that resulted in massive radiations: 80% of extant mammals are herbivores, and herbivory is present in most mammalian lineages (*9*). To access the more complex carbohydrates present in plants, such as celluloses and resistant starches, disparate mammalian lineages lengthened gut retention times to accommodate bacterial fermentation; this occurred via enlargement of the foregut or hindgut (*9*). We found that herbivores clustered into two groups that corresponded generally to foregut fermenters and hindgut fermenters: the foregut-fermenting sheep, kangaroo, okapi, giraffe, and cattle clustered together to form herbivore group 1 in fig. S2, whereas the hindgut-fermenting elephant, horse, rhinoceros, capybara, mole rat, and gorilla clustered together in herbivore group 2. The strong impact of gut morphology on bacterial community composition is also evident in PCoA of the UniFrac data: Herbivores separate into fore- and hindgut groups, and omnivores separate into hindgut fermenters and those with simple guts (Fig. 2, C and D).

Differences between the fecal communities of foregut and hindgut fermenters are likely due to host digestive physiology: In foregut fermenters, the digesta is moved into the equivalent of the monogastric stomach after fermentation, so that part of the microbiota is also digested; in hindgut fermenters, the fermentative microbes are more likely to be excreted in the feces. Fermentation requires microbial interactions such as cross-feeding and interspecies hydrogen transfer (*28*). Our results suggest that as mammals underwent convergent evolution in the morphological adaptations of their guts to herbivory, their microbiota arrived at similar compositional configurations in unrelated hosts with similar gut structures.

The diet outliers in our study were folivores. Despite their herbivorous diet, red and giant pandas have simple guts, cluster with other carnivores, and have carnivore-like levels of phylogenetic diversity (figs. S2 and S3). In folivorous primates, the simple gut has evolved pouches for fermentation of recalcitrant plant material (*9*). The fecal microbiota of the two colobus monkeys and the François langur cluster together by UniFrac with the three pig species (Red River hog, Visayun warty pig, babirusa) and the flying fox, baboon, chimpanzee, gorilla, and orangutan, forming a phylogenetically mixed group whose diets include a large component of plant material. This cluster occupies an intermediate position between other primates and herbivorous foregut fermenters in fig. S2. This observation suggests that the colobus monkeys and the François langur harbor microbial lineages typical of omnivores but have a greater representation of the lineages driving the breakdown of a plant-based diet. Such host-level selection of specific members of a microbiota has been demonstrated under laboratory conditions by reciprocal transplantations of gut microbiota from one host species to germ-free recipients of a different species: Groups of bacteria were expanded or contracted in the recipient host to resemble its "normal" microbiota through a process that may have been influenced by diet (*26*).

Coevolution has been hypothesized to occur in animal species whose parental care enables vertical transmission of whole gut communities, and where the properties of the community as a whole confer a fitness advantage to the host (*29*). Although coevolution has been inferred from observations of bacterial host specificity (*30*), these observations could also be explained by dietary preference. Therefore, we searched for evidence of codiversification, a special case of coevolution (*2*) that would be manifest in this case by a clustering of fecal microbial communities that mirrors the mammalian phylogeny. A UniFrac analysis was performed recursively (*10*) on the entire mammalian fecal bacterial tree, using a procedure that had the effect of asking whether the bacterial lineages stemming from each tree node mirrored the mammalian phylogeny (*31*). The results were compared to those using a randomized version of the mammalian phylogeny. The patterns of community similarity matched the mammal phylogeny more often than would

be expected if no codiversification had occurred (fig. S4; $P = 1.79 \times 10^{-11}$; $t = -6.73$, df = 88).

Although mammalian gut microbes are highly adapted to life in this body habitat, and many lineages are extremely rare outside of it (*29*), they appear to be fairly promiscuous between hosts. This could account for the spectacular success of mammals and herbivores in particular: Acquiring a gut microbiota was not a constraint, and morphological and behavioral adaptations were likely far more restrictive. One implication of this work is that the tolerance of the immune system to gut microbes is a basal trait in mammal evolution.

The global success of humans is based in part on our ability to control the variety and amount of food available using agriculture and cookery. These capabilities have not appreciably affected the major bacterial lineages that constitute our gut microbiota: As noted above, fecal samples from unrelated healthy human samples cluster with other omnivores (Fig. 1 and fig. S2), with interpersonal differences (UniFrac distances) being significantly smaller than the distances between humans and all other mammalian species ($G = -47.7$, $P < 0.005$, $n = 106$). Although our interpersonal differences appear to be smaller than interspecies differences among mammals, deeper sampling and analysis will be required to circumscribe the gut microbial diversity inherent to humans. This is one of the early goals of the recently initiated international human microbiome project (*1*).

References and Notes

1. P. J. Turnbaugh *et al.*, *Nature* **449**, 804 (2007).
2. N. A. Moran, *Curr. Biol.* **16**, R866 (2006).
3. A. M. Lister, *Philos. Trans. R. Soc. London Ser. B* **359**, 221 (2004).
4. M. Pagani, J. C. Zachos, K. H. Freeman, B. Tipple, S. Bohaty, *Science* **309**, 600 (2005); published online 16 June 2005 (10.1126/science.1110063).
5. Y. Huang *et al.*, *Science* **293**, 1647 (2001).
6. T. E. Cerling *et al.*, *Nature* **389**, 153 (1997).
7. B. J. MacFadden, *Trends Ecol. Evol.* **20**, 355 (2005).
8. T. E. Cerling, J. R. Ehleringer, J. M. Harris, *Philos. Trans. R. Soc. London Ser. B* **353**, 159 (1998).
9. C. Stevens, I. Hume, *Comparative Physiology of the Vertebrate Digestive System* (Cambridge Univ. Press, Cambridge, ed. 2, 2004).
10. See supporting material on *Science* Online.
11. J. C. Frey *et al.*, *Appl. Environ. Microbiol.* **72**, 3788 (2006).
12. Y. Ozutsumi, H. Hayashi, M. Sakamoto, H. Itabashi, Y. Benno, *Biosci. Biotechnol. Biochem.* **69**, 1793 (2005).
13. S. P. Brooks, M. McAllister, M. Sandoz, M. L. Kalmokoff, *Can. J. Microbiol.* **49**, 589 (2003).
14. P. B. Eckburg *et al.*, *Science* **308**, 1635 (2005); published online 14 April 2005 (10.1126/science.1110591).
15. R. E. Ley, P. J. Turnbaugh, S. Klein, J. I. Gordon, *Nature* **444**, 1022 (2006).
16. H. Hayashi, M. Sakamoto, Y. Benno, *Microbiol. Immunol.* **46**, 535 (2002).
17. H. Hayashi, M. Sakamoto, Y. Benno, *Microbiol. Immunol.* **46**, 819 (2002).
18. H. Hayashi, M. Sakamoto, M. Kitahara, Y. Benno, *Microbiol. Immunol.* **47**, 557 (2003).
19. P. Shannon *et al.*, *Genome Res.* **13**, 2498 (2003).
20. R. E. Ley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11070 (2005).
21. T. Huber, G. Faulkner, P. Hugenholtz, *Bioinformatics* **20**, 2317 (2004).
22. T. Z. DeSantis *et al.*, *Appl. Environ. Microbiol.* **72**, 5069 (2006).
23. Q. Wang, G. M. Garrity, J. M. Tiedje, J. R. Cole, *Appl. Environ. Microbiol.* **73**, 5261 (2007).
24. C. Lozupone, M. Hamady, R. Knight, *BMC Bioinform.* **7**, 371 (2006).
25. C. Lozupone, R. Knight, *Appl. Environ. Microbiol.* **71**, 8228 (2005).
26. J. F. Rawls, M. A. Mahowald, R. E. Ley, J. I. Gordon, *Cell* **127**, 423 (2006).
27. P. L. Koch, M. L. Fogel, N. Tuross, in *Stable Isotopes in Ecology and Environmental Science*, K. Lajtha, R. H. Michener, Eds. (Blackwell, Oxford, 1994).
28. J. B. Russell, J. L. Rychlik, *Science* **292**, 1119 (2001).
29. R. E. Ley, D. A. Peterson, J. I. Gordon, *Cell* **124**, 837 (2006).
30. L. Dethlefsen, M. McFall-Ngai, D. A. Relman, *Nature* **449**, 811 (2007).
31. U. Arnason *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8151 (2002).

32. We thank S. Wagoner for superb technical assistance; L. Fulton, R. Fulton, K. Delahaunty, and our other colleagues in the Washington University Genome Sequencing Center for assistance with 16*S* rRNA gene sequencing; and the Namibian Ministry of Environment and Tourism. Supported by NIH grants DK78669, DK70977, and DK30292, the W. M. Keck Foundation, the Ellison Medical Foundation, and NIH Molecular Biophysics Training Program grant T32GM065103 (M.H.). Sequences (EU458114 to EU475873 and EU771093 to EU779492) were deposited in GenBank.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1155725/DC1
Materials and Methods
Figs. S1 to S14
Table S1
References

28 January 2008; accepted 15 May 2008
Published online 22 May 2008;
10.1126/science.1155725
Include this information when citing this paper.

Ankyrin Repeat Proteins Comprise a Diverse Family of Bacterial Type IV Effectors

Xiaoxiao Pan,[1] Anja Lührmann,[1]* Ayano Satoh,[2]* Michelle A. Laskowski-Arce,[1]† Craig R. Roy[1]‡

Specialized secretion systems are used by many bacteria to deliver effector proteins into host cells that can either mimic or disrupt the function of eukaryotic factors. We found that the intracellular pathogens *Legionella pneumophila* and *Coxiella burnetii* use a type IV secretion system to deliver into eukaryotic cells a large number of different bacterial proteins containing ankyrin repeat homology domains called Anks. The *L. pneumophila* AnkX protein prevented microtubule-dependent vesicular transport to interfere with fusion of the *L. pneumophila*-containing vacuole with late endosomes after infection of macrophages, which demonstrates that Ank proteins have effector functions important for bacterial infection of eukaryotic host cells.

Type IV secretion systems (TFSSs) are molecular machines used by Gram-negative bacteria for protein transfer into recipient cells (*1*). Many bacterial pathogens and endosymbionts use TFSSs to regulate host processes important for survival and replication (*2*), and several of these organisms have a large number of genes encoding proteins with multiple ankyrin repeat homology domains (ARHDs) (*3*–*7*). Infrequently encountered in bacterial proteins but common in eukaryotic proteins, ARHDs form molecular scaffolds that mediate protein-protein interactions (*8*). An *Anaplasma phagocytolyticum* protein containing multiple ARHDs called AnkA (*9*) and several ARHD proteins in strains of *Wolbachia* (*10*, *11*) have been proposed to be delivered into host cells by a TFSS (*12*); however, whether Ank proteins are bona fide TFSS effectors has not been established.

Legionella pneumophila and *Coxiella burnetii* are both intracellular pathogens that encode several proteins containing ARHDs and a TFSS called Dot/Icm (*5*–*7*). To test whether ARHD proteins are TFSS substrates, we measured host cell translocation of four *L. pneumophila* Ank proteins fused to a calmodulin-dependent adenylate cyclase reporter (Cya), using the *L. pneumophila* effector RalF as a positive control (*13*, *14*). These four Ank proteins were delivered into mammalian cells as indicated by a >10-fold increase in adenosine 3′,5′-monophosphate (cAMP) following infection (Fig. 1A). No cAMP increase was observed when the Cya-Ank proteins were produced in the *L. pneumophila* Δ*dotA* mutant lacking a functional TFSS, which indicates that the Dot/Icm system is required for Ank protein delivery into host cells. Thirteen different *C. burnetii* proteins with ARHDs were tested for translocation with the Cya assay. Genetic ma-

[1]Section of Microbial Pathogenesis, Yale University School of Medicine, 295 Congress Avenue, New Haven, CT 06536, USA. [2]Department of Cell Biology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520, USA.

*These authors contributed equally to this work.
†Present address: University of Texas Southwestern Medical Center, 6000 Harry Hines Boulevard, NA5.124, Dallas, TX 75390, USA.
‡To whom correspondence should be addressed. E-mail: craig.roy@yale.edu