

Convergence of gut microbiomes in myrmecophagous mammals

FRÉDÉRIC DELSUC,^{*,†‡} JESSICA L. METCALF,[‡] LAURA WEGENER PARFREY,[‡] SE JIN SONG,^{‡§} ANTONIO GONZÁLEZ[‡] and ROB KNIGHT^{†‡¶**}

^{*}Institut des Sciences de l'Évolution, UMR 5554-CNRS-IRD, Université Montpellier 2, Montpellier, France, [†]Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA, [‡]Biofrontiers Institute, University of Colorado, Boulder, CO 80309, USA, [§]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA, [¶]Department of Computer Science, University of Colorado, Boulder, CO 80309, USA, ^{**}Howard Hughes Medical Institute, University of Colorado, Boulder, CO 80309, USA

Abstract

Mammals have diversified into many dietary niches. Specialized myrmecophagous (ant- and termite-eating) placental mammals represent a textbook example of evolutionary convergence driven by extreme diet specialization. Armadillos, anteaters, aardvarks, pangolins and aardwolves thus provide a model system for understanding the potential role of gut microbiota in the convergent adaptation to myrmecophagy. Here, we expand upon previous mammalian gut microbiome studies by using high-throughput barcoded Illumina sequencing of the 16S rRNA gene to characterize the composition of gut microbiota in 15 species representing all placental myrmecophagous lineages and their close relatives from zoo- and field-collected samples. We confirm that both diet and phylogeny drive the evolution of mammalian gut microbiota, with cases of convergence in global composition, but also examples of phylogenetic inertia. Our results reveal specialized placental myrmecophages as a spectacular case of large-scale convergence in gut microbiome composition. Indeed, neighbour-net networks and beta-diversity plots based on UniFrac distances show significant clustering of myrmecophagous species (anteaters, aardvarks and aardwolves), even though they belong to phylogenetically distant lineages representing different orders. The aardwolf, which diverged from carnivorous hyenas only in the last 10 million years, experienced a convergent shift in the composition of its gut microbiome to become more similar to other myrmecophages. These results confirm diet adaptation to be a major driving factor of convergence in gut microbiome composition over evolutionary timescales. This study sets the scene for future metagenomic studies aiming at evaluating potential convergence in functional gene content in the microbiomes of specialized mammalian myrmecophages.

Keywords: 16S rRNA sequencing, convergence, gut microbiome, mammals, microbial diversity

Received 20 April 2013; revision received 14 August 2013; accepted 15 August 2013

Introduction

The radiation of extant mammals has resulted in more than 5000 living species, which diversified into a wide variety of diet niches ranging from broadly generalized

to highly specialized (Feldhamer 2007). Host-associated microbiota play a major role in diet specialization across vertebrates (Ley *et al.* 2008a; Karasov *et al.* 2011) to the point that they might be considered an integral part of the phenotype (McFall-Ngai *et al.* 2013). In mammals, high-throughput 16S rRNA gene sequencing of faecal samples from a diversity of species has shown that both diet and phylogeny have driven the evolution

Correspondence: Rob Knight, Fax: +1-303-492 7744; E-mail: rob.knight@colorado.edu

of the gut microbiome (Ley *et al.* 2008b). At large taxonomic scales, diet appears to be a major driving factor, as gut microbiomes have evolved convergently in mammals sharing the same feeding habits (Muegge *et al.* 2011). In particular, large differences in gut microbial communities have been demonstrated among carnivores, omnivores and herbivores. Within herbivores, a clear divide can be seen between foregut fermenters, which consist essentially of artiodactyls, and the phylogenetically diverse hindgut-fermenting taxa, which include horses, capybaras, rabbits and elephants (Ley *et al.* 2008b; Muegge *et al.* 2011). Cases of convergence driven by the herbivorous diet are also found among vertebrates with the folivorous stinkbird or hoatzin (Godoy-Vitorino *et al.* 2011) and herbivorous fish (Sullam *et al.* 2012) showing similarities in gut microbiome composition to ruminant mammals and in invertebrates with ants in which herbivory has also driven the convergence of gut symbionts (Russell *et al.* 2009; Anderson *et al.* 2012). At shallower taxonomic scales, such as within bears (Ley *et al.* 2008b), great apes (Ochman *et al.* 2010) and between bat families (Phillips *et al.* 2012), host phylogeny and microbiome composition seem to have codiverged. For instance, despite its exclusive bamboo diet, the giant panda still hosts a gut microbiome similar to other bears (Ley *et al.* 2008b).

Mammalian myrmecophages (anteater and termite eaters) provide an opportunity for further understanding the mechanisms that drive the evolution of the gut microbiome by disentangling the effects of diet and phylogeny. Whereas more than 200 mammalian species include a significant portion of ants and/or termites in their diet, only 22 of them can be considered specialized myrmecophages eating more than 90% of ants and/or termites (Redford 1987). Most of these species feed opportunistically on both types of social insects, with only a few species specialized on either ants or termites. In fact, the aardwolf is the only true termite-eating specialist. Ant-eating specialists are restricted to arboreal species such as the pygmy anteater (*Cyclopes didactylus*), which is the only anteater species fully specialized on ants (Miranda *et al.* 2009), and the long-tailed pangolin (*Manis tetradactyla*), which probably feeds specifically on arboreal ant species (Redford 1987). Mammalian myrmecophages can thus be considered as a highly specialized group of insectivores (i.e. a term generally used to cover the eating of all terrestrial invertebrates) feeding exclusively on social insects of low nutritional value (Redford & Dorea 1984).

Myrmecophagous mammals represent a textbook example of phenotypic evolutionary convergence (Feldhamer 2007; McGhee 2011) and therefore provide a model system in which to characterize the taxonomic composition of the gut microbiome in convergently

evolved species sharing the same highly specialized diet. Lineages specialized to eat exclusively ants and/or termites have independently evolved multiple times in mammalian evolutionary history: in monotremes, the short-beaked echidna (*Tachyglossus aculeatus*); in marsupials, the numbat (*Myrmecobius fasciatus*); in placental mammals across five different orders, tolpeutine armadillos (*Priodontes*, *Cabassous* and *Tolypeutes*), the three anteater genera (*Cyclopes*, *Myrmecophaga* and *Tamandua*), pangolins (*Manis*), the aardvark (*Orycteropus afer*) and the aardwolf (*Proteles cristata*) (Redford 1987). These five placental lineages represent different degrees of morphological specialization towards myrmecophagy, probably reflecting both phylogenetic constraints and the time since this peculiar feeding habit evolved. Giant armadillos, anteaters aardvarks, and pangolins constitute the most extreme examples of specialized myrmecophagous phenotypes (Reiss 2001). These animals have developed similar but convergent morphological adaptations such as the reduction in or loss of teeth, an elongated muzzle with an extensible tongue, viscous saliva produced by hypertrophied salivary glands and powerful claws used to dig into ant and termite nests. Placental myrmecophages also share a relatively low metabolic rate due to their nutritionally poor diet (McNab 1984). Indeed, exclusively feeding on social insects imposes strong energetic constraints because most of the protein value is locked in their recalcitrant chitin exoskeleton (Redford & Dorea 1984). Chitinase genes are found in mammalian genomes, but their exact role in digestion is still unclear (Bussink *et al.* 2007). However, chitinolytic bacteria are ecologically widespread (Gooday 1990), and some species have been identified in the mammalian digestive tract (Simůnek *et al.* 2001). This raises the hypothesis that specialized myrmecophagous mammals might rely on symbiotic bacteria for degrading chitin exoskeletons to optimize their protein nutritional intake.

Recent molecular phylogenetic advances have unambiguously demonstrated that myrmecophagous placentals are anciently diverged, independent lineages with more than 80 million years separating myrmecophagous xenarthrans (tolpeutine armadillos and anteaters), aardvarks, pangolins and aardwolves (Springer *et al.* 2003; Delsuc *et al.* 2004; Meredith *et al.* 2011). Specialized myrmecophages therefore provide an especially good model for studying convergence of the gut microbiome over large evolutionary timescales. In this work, we expand upon previous studies of placental gut microbiome evolution using high-throughput Illumina barcoded 16S rRNA amplicons of faecal samples from representatives of all placental myrmecophagous lineages and their close relatives. We characterize the taxonomic composition of their gut microbiota with the objective of

answering several fundamental ecological and evolutionary questions: 'How distinct are the microbiomes of ant-eating mammals in terms of taxonomic composition compared with other mammals?' 'Have entire gut bacterial communities been affected by convergence towards myrmecophagy or do they still show signs of phylogenetic constraints?' 'Did myrmecophagous host species independently recruit similar gut microbes?' 'Can we identify bacterial taxa specific to myrmecophages that may be potential chitin degraders?'

Materials and methods

Data acquisition

Faecal samples from myrmecophagous placentals and phylogenetically related species were provided by European (Colchester UK, London UK, Leipzig DE, and Montpellier FR) and US (Atlanta GA, Cincinnati OH, Houston TX, and San Diego CA) zoos. We also used faecal samples of 30 giant anteaters (*Myrmecophaga tridactyla*) and five giant armadillos (*Priodontes maximus*) collected in 2006 in Emas National Park (Brazil) under IBAMA licence no. 02001.00215/07-21 provided by the Brazilian Institute on Environment and Natural Resources for a study using scat detection dogs (Vynne *et al.* 2011). Faeces were also collected in 2012 from 10 nine-banded armadillos (*Dasyurus novemcinctus*) previously captured by staff of the Merritt Island National Wildlife Refuge in Florida (USA). Finally, two wild pichis (*Zaedyus pichiy*) and a pink fairy armadillo (*Chlamyphorus truncatus*) were sampled in 2009 with Research Permit 339/08 issued by the Dirección de Recursos Naturales Renovables of Mendoza Province (Argentina). These last three samples were lyophilized and shipped immediately after collection. For zoo samples, faeces were usually swabbed within 48 h of defaecation. For field samples, swabs were made from previously frozen or lyophilized faecal material. A total of 93 swab samples from 15 species were collected (Table S1, Supporting information).

DNA extraction, PCR amplification and amplicon preparation for sequencing followed the protocols described in Caporaso *et al.* (2012) and can be found on the Earth Microbiome Project (EMP; Gilbert *et al.* 2010) web page (<http://www.earthmicrobiome.org/emp-standard-protocols/>). Briefly, faecal swabs were extracted using the PowerSoil DNA isolation kit (MoBio Laboratories, Carlsbad, CA, USA). Total genomic DNA was subjected to PCR amplification targeting a ~300-bp fragment of the 16S rRNA variable region 4 (V4) using the universal bacterial primer set 515F/806R, which amplifies bacterial and archaeal 16S genes near universally (e.g. Walters *et al.* 2011; Caporaso *et al.* 2012). Three

replicate PCRs were performed for each DNA sample, and amplicons generated from each set of three reactions were subsequently pooled and quantified using PicoGreen. Negative controls included no template controls for DNA extraction and PCR amplification. Finally, all barcoded amplicons were pooled in equal concentrations for sequencing. The amplicon pool was purified using the MoBio UltraClean PCR Clean-up kit and sequenced on the Illumina MiSeq sequencing platform (MiSeq Control Software 2.0.5 and Real-Time Analysis software 1.16.18) at the BioFrontiers Institute Next-Generation Genomics Facility at University of Colorado, Boulder, USA. We analysed the single-end sequencing read from the 515f primer (GTGCCAGCMGCCGCG GTAA).

16S rRNA data processing and taxonomic assignment

Raw 16S rRNA amplicon sequences were processed using the QIIME suite of software tools (version 1.6.0-dev) (Caporaso *et al.* 2010a). Sequences were demultiplexed and quality-filtered according to default parameters within QIIME. These sequences were then clustered into operational taxonomic units (OTUs) with a sequence similarity threshold of 97% with UCLUST (Edgar 2010) within QIIME. We assigned sequences to OTUs in two ways, first with an open-reference protocol that captures the full diversity within our data set and second with a closed-reference protocol that enables comparison with previously published studies.

For the open-reference approach, we followed the subsampling open-reference protocol with default parameters in QIIME (http://qiime.org/tutorials/open_reference_illumina_processing.html). Briefly, sequence reads were initially clustered against the October 2012 release of the Greengenes (DeSantis *et al.* 2006; McDonald *et al.* 2012) 97% reference data set (<http://greengenes.secondgenome.com>). The majority of sequences, 85%, matched the reference database, and these OTUs received the taxonomic classification standardized in Greengenes. Sequences that did not match the Greengenes data set at 97% were subsequently clustered into de novo OTUs at 97% similarity with UCLUST. The representative sequences of all OTUs were then aligned to the Greengenes reference alignment using PyNAST (Caporaso *et al.* 2010b), and this alignment was used to construct a phylogenetic tree using FastTree (Price *et al.* 2010) within QIIME. This tree was used subsequently for phylogenetically informed diversity analyses. Sequences that did not align to Greengenes with a 70% similarity threshold were assumed to be non-16S and thus artefactual and removed from further analysis. These de novo OTUs were then assigned their

taxonomies to the finest level possible with the RDP classifier (Wang *et al.* 2007) retrained on the Greengenes October 2012 reference data set using an 80% confidence threshold.

The second OTU picking approach utilized closed-reference OTU picking against the Greengenes 97% reference database from 4 February 2011 (Files available at http://qiime.org/home_static/dataFiles.html). Subsequent analyses used the Greengenes reference tree and taxonomy assignments (McDonald *et al.* 2012). These new data were then combined with previously published data sets. Because all closed-reference data sets are processed using the same method and the same references for taxonomic assignments, using a closed-reference approach allows for the combination of different data sets and comparisons across different studies. Our data as well as previously published data are available in the EMP/QIIME database (<http://www.microbio.me/qiime/> and <http://www.microbio.me/emp/>). For our 93 samples, the closed-reference and open-reference data sets were compared with Procrustes analysis to assess whether we were recovering the same beta-diversity patterns with each data set. This analysis rotates or transforms the points in one PCoA plot to try to match the corresponding points in the second plot while still preserving the relative distances between points within each plot. The goodness of fit (M^2) and statistical significance of the goodness of fit (P) were then measured to determine the level of correspondence between the two sets. Finally, we also compared proportions of phyla that were not assigned to Greengenes by comparing taxa summary charts. For both OTU picking pipelines, low-abundance OTUs (OTUs representing <0.00005% of the total reads in the data set) were filtered out as recommended for Illumina-generated sequence data (Bokulich *et al.* 2013).

Comparative analyses with other mammals

Source tracking. Before combining our closed-reference data set with additional mammalian data, we controlled for potential contamination by soil and other confounding environmental factors. We suspected that some samples, in particular field-collected samples, may have been contaminated postdefaecation with soil or other environmental bacteria. Also, it has been reported that faecal samples of armadillo species can contain high percentages of soil particles (Anacleto 2007; Vaz *et al.* 2012), and some myrmecophagous animals such as pangolins are fed with an insectivore diet mixed with soil in captivity (Yang *et al.* 2007). Because we are not able to confidently distinguish between these two scenarios, we filtered out samples that either displayed evidence of contamination or assigned to a low number

of taxa previously described from mammal gut communities. To detect potential contamination, we used a Bayesian approach to estimate the proportion of each sample derived from a priori-defined source communities (Knights *et al.* 2011). As source data sets, we included 16S rRNA data from a set of 42 soil samples (Ramirez *et al.* 2010; Eilers *et al.* 2012) sequenced using the same 16S rRNA V4 region primers (Table S2, Supporting information) and a representative and diverse set of mammal gut communities (Muegge *et al.* 2011). We excluded samples with more than 10% assignment to our representative soil community. Because our sample set focused on myrmecophagous mammals, which includes host species with potentially highly unique gut microbiomes, we used a fairly liberal filtering threshold (i.e. very low percentage) and removed samples for which <0.01% of the community assigned to our representative source of mammal gut communities. Our goal with these filtering thresholds was to remove samples that were probably not representative of the host's gut microbiome.

Statistical analyses. We performed comparative analyses of myrmecophagous, insectivorous, omnivorous, folivorous, herbivorous and carnivorous mammals by combining our filtered data set with 39 available samples that were originally sequenced as part of the Muegge *et al.*'s. (2011) study. We resequenced the V4 16S region for 16 of these samples on the Illumina platform using the 515f/806r primer set following EMP protocols as described above (Table S3, Supporting information). Previous work has shown that sequence reads of this length amplified over the V2 or V4 region are sufficient for accurate and repeatable taxonomic identification to at least the family level and for global community characterization (Liu *et al.* 2008; Caporaso *et al.* 2011). Because we had both V2 and V4 16S region sequence data for 16 samples, we compared these data sets to confirm their similarity using a Procrustes analysis of PCoA scores based on unweighted UniFrac distances.

Because we included 23 samples sequenced on the 454 platform (Muegge *et al.* 2011), we rarefied our closed-reference comparative data set at a level of 1100 sequences per sample to avoid biases caused by differences in sequencing depth of samples. The composition of each sample was summarized at various taxonomic levels using QIIME. The Greengenes reference tree was used to perform beta-diversity comparisons by computing phylogeny-based UniFrac distances (Lozupone & Knight 2005) between samples within QIIME. For exploring relationships between host diet and host phylogeny, we then used both principal coordinate analysis (PCoA) on UniFrac distance matrices within

QIIME and phylogenetic network analysis of community similarity (Parks & Beiko 2012) by reconstructing networks from UniFrac distance matrices with Splits-Tree4 (Huson & Bryant 2006) using the neighbour-net agglomerative method (Bryant & Moulton 2004). To assess the relative importance of diet vs. host phylogeny in shaping the gut microbiota, we performed two-factor crossed analysis of similarity (ANOSIM) with the factors diet and host order using the software package PRIMER v6 (Clarke & Gorley 2006). To address the uneven sampling across diet and host order categories, we randomly subsampled even numbers of samples for each category and ran ANOSIM on diet and host order separately within QIIME. This subsampling procedure was repeated 100 times.

Finally, we identified bacterial genera significantly associated with an ant-eating diet using the script `otu_category_significance.py` in QIIME. We performed this analysis using a related-taxon approach where myrmecophagous species were compared with their nonmyrmecophagous relatives, allowing us to better control for the potential effect of phylogenetic distance

between certain myrmecophagous and nonmyrmecophagous species. Four separate comparisons were possible given the taxonomic breadth of our data set: (i) aardvark vs. rock hyrax, (ii) aardwolf vs. nonmyrmecophagous species in the order Carnivora, (iii) giant anteater vs. sloths and (iv) southern tamandua vs. sloths. For each comparison, we used an analysis of variance (ANOVA) to identify bacterial genera that are found at significantly higher relative abundance in myrmecophagous species compared with their nonmyrmecophagous relatives. After correction for multiple comparisons, this provided a list of the taxa potentially associated with myrmecophagous mammals (Table 1), for which we calculated the mean relative abundance found across the different diet classes (Table S5, Supporting information). To ensure that microbial taxa potentially unique to this study were not overlooked, these analyses were performed on both the closed-reference (i.e. only sequences matching those in the Greengenes database are retained) and open-reference (i.e. nonmatching sequence clusters are also retained) data sets, respectively, rarefied to 1100 and 12 000 sequences per sample.

Table 1 Bacterial genera significantly higher in mean relative abundance (RA) in the guts of myrmecophagous (Myr) species than closely related nonmyrmecophagous (NM) species. Taxa in bold are those detected in more than one comparison

Myr species	NM related species	Bacterial genus (Family)	Myr RA	NM RA	P^{\dagger}
Aardvark	Rock hyrax	None	NA	NA	NA
Aardwolf	Hyena, lion, bush dog, polar bear	<i>Prevotella</i> (Prevotellaceae)	0.1041	0.0025	0.0001
		<i>Streptococcus</i> (Streptococcaceae)	0.1327	0.0005	0.0341
		<i>Dialister</i> (Veillonellaceae)	0.0055	0	0.0021
		<i>Klebsiella</i> (Enterobacteriaceae)	0.01	0	0.0147
		Unclassified (Erysipelotrichaceae)	0.1123	0.0007	0.0111
		Unclassified (Clostridiaceae)	0.1173	0.0216	0.0182
		<i>Faecalibacterium</i> (Ruminococcaceae)	0.025	0.0027	0.0478
		<i>Eubacterium</i> (Lachnospiraceae)	0.0186	0	0.0489
Giant anteater	Linnaeus two-toed sloth, Hoffmann's two-toed sloth	Unclassified (Clostridiaceae)	0.0286	0.0063	0.007
		<i>Streptococcus</i> (Streptococcaceae)	0.3031	0.0023	0.001
		<i>Eubacterium</i> (Erysipelotrichaceae)*	0.0566	0.0036	0.0019
		<i>Bifidobacterium</i> (Bifidobacteriaceae)	0.0251	0.0025	0.039
		<i>Blautia</i> (Lachnospiraceae)*	0.0241	0.0009	0.0086
		Unclassified (Streptococcaceae)*	0.0128	0.0001	0.0009
		<i>Lactococcus</i> (Streptococcaceae)*	0.0124	0.0001	0
		<i>Lactobacillus</i> (Lactobacillaceae)*	0.0052	0.0004	0.0083
		<i>Peptococcus</i> (Peptococcaceae)	0.0048	0	0.0243
		<i>Turcibacter</i> (Erysipelotrichaceae)	0.0075	0.0014	0.0331
		<i>Sarcina</i> (Clostridiaceae)*	0.0013	0	0
		<i>Catenibacterium</i> (Erysipelotrichaceae)	0.0283	0	0.0081
Southern tamandua	Linnaeus two-toed sloth, Hoffmann's two-toed sloth	<i>Bulleidia</i> (Erysipelotrichaceae)	0.0201	0.0002	0.0095
		<i>Blautia</i> (Lachnospiraceae)	0.0173	0.0005	0
		<i>Roseburia</i> (Lachnospiraceae)*	0.0105	0.0003	0.0451
		<i>Peptococcus</i> (Peptococcaceae)	0.0017	0	0.0029
		<i>Collinsella</i> (Coriobacteriaceae)*	0.0015	0	0

*Additional taxa detected using the open-reference data set.

[†]FDR corrected.

Results

We generated 16S rRNA gene sequences from a total of 93 new faecal samples from 15 mammalian species representing all major ant-eating lineages, including ant-eaters, aardvarks, pangolins and aardwolves, and closely related species such as armadillos and sloths (Table S1). The MiSeq run resulted in 2 030 814 sequence reads of 115–145 bp after quality filtering with default QIIME parameters. One sample failed to amplify (wild giant armadillo sample Pri.max.Seat.05007 with five sequences in total), resulting in a data set of 92 successfully sequenced new mammal gut samples with an average of 22 074 sequences per sample.

Comparison of OTU picking methods

We compared the results from the two different OTU picking protocols to determine what impact they had on the biological conclusions reached from our data set. These methods were the open-reference method (using the Greengenes reference alignment, but allowing for new OTU clusters) and the closed-reference OTU picking (discarding sequences that do not cluster with 97% similarity with the Greengenes reference set). Open-reference OTU picking utilizes more of the sequence data as reads that do not match the reference data set with 97% similarity are clustered into new OTUs, classified to the finest taxonomic level possible and retained in the data set. However, this approach precludes the comparison of data sets sequenced with different primer sets such as the data set from Muegge *et al.* (2011). Closed-reference OTU picking, on the other hand, allows for comparison of data sets sequenced using nonoverlapping primer regions because only the reads that cluster with the reference data set are retained in the analysis. As the reference data set is composed of full-length 16S rRNA sequences, reads generated from different primer sets can be assigned to the same OTU.

Open-reference OTU picking resulted in a total data set of 1 814 980 sequences being assigned to OTUs. The number of assigned sequences per sample ranged from 11 466 to 35 420 with a median of 19 516. Closed-reference OTU picking resulted in ~82% of the open-reference sequences (i.e. 1 481 672 sequences), aligning with Greengenes OTU's. The number of sequences per sample ranged from 8196 to 31 701 with a median of 15 512. Of the 188 336 sequence reads that did not assign to the reference data set, most sequences were assigned to either Firmicutes (42%) or Bacteroidetes (32%) (Fig. S1, Supporting information). The two OTU picking methods nevertheless yielded very similar bacterial communities (Mantel test comparing UniFrac distance matrices from each OTU picking method

$R^2 = 0.95$, $P = 0.001$). A plot of the Procrustes analysis comparing the PCoA based on unweighted UniFrac distances obtained from open-reference and closed-reference OTU assignment methods for the 92 new samples again shows that these are similar (Procrustes $M^2 = 0.041$, $P = 0.001$) and illustrates that the global sample clustering patterns projected on the first two axes of the PCoA match almost exactly (Fig. 1). These results suggest that the closed-reference data set can be used for downstream comparative analyses incorporating critical samples from previously published data sets that also used Greengenes as a reference set (Muegge *et al.* 2011). Thus, we utilized the closed-reference data set for the majority of our analyses.

Filtering out samples that were not probably representative of the host's gut microbiome

Of the 92 new samples successfully sequenced as part of this study, 56 samples resulted in an assignment to a mammalian gut source community of more than 0.01% (Table S4 and Fig. S2, Supporting information). Among these, one wild giant armadillo sample (Pri.max.Seat.04J09) also had a 10% assignment to a soil source community, leaving 55 samples of the initial 92 for

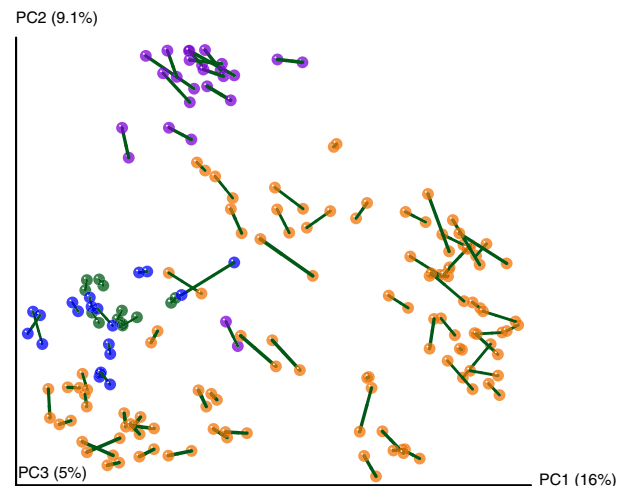


Fig. 1 Procrustes analysis comparing unweighted UniFrac distances for open-reference and closed-reference OTU assignment methods of the 92 mammalian gut samples. This kind of analysis attempts to stretch and rotate the points in one matrix, here obtained by PCoA, to be as close as possible to the points in the other matrix, thus preserving the relative distances between points within each matrix. Circles represent bacterial communities of samples characterized by either open- or closed-reference OTU assignments coloured by diet (blue: omnivore; green: folivore; purple: insectivore; orange: myrmecophage). Each samples' two circles are connected to each other by lines representing the distortion between the two matrices.

subsequent analyses. In total, 5 of the 92 samples resulted in a substantial proportion of their bacterial community assigning to soil. These were field-collected samples from the giant anteater (*Myrmecophaga tridactyla*), giant armadillo (*Priodontes maximus*) and aardvark (*Orycteropus afer*) samples from the Colchester Zoo, for which we suspected soil contamination during sampling. Notably, one of the giant armadillo samples (Pri.max.Seat.3.04L11) assigned 56% of its bacterial community to the source soil data set. Most samples with a low assignment to a mammalian gut community also assigned a high proportion of their community to an 'unknown' source, including 27 samples with 100% assignment to an 'unknown' source. Of these, 19 were giant anteater and giant armadillo faeces located by scat detection dogs in Emas National Park (Brazil). These faecal samples potentially were exposed to a number of contamination sources, including rain and soil. Although we included soil as a source community in our source-tracking analysis, our soil sample set was representative of major soil types and not specific to the field sites or zoo soils associated directly with faecal samples included in this study. Also included in the samples assigning to an 'unknown' source were five aardvark samples from the Colchester Zoo, for which we suspected problems during shipment (i.e. the samples had thawed and stayed at ambient temperature for almost a week), and a wild pichi (*Zaedyus pichiy*) sample. Unfortunately, the only two Chinese pangolin (*Manis pentadactyla*) samples that we managed to collect from the Leipzig Zoo also belonged to this category and were therefore also excluded. Because this source-tracking analysis depended on a closed-reference data set that included a wide taxonomic range of mammals, it is unlikely that full assignment to 'unknown' was due to samples having an entirely unique mammalian gut microbiome, but rather likely due to contamination from sources not included in our source-tracking analysis.

Combining closed-reference mammalian gut data sets

We combined our 55 samples with those from 39 mammals in Muegge *et al.*'s study. (2011), including 23 originally sequenced for the V2 region and 16 samples resequenced for the V4 region. Through a Procrustes analysis, these 16 resequenced samples were confirmed to have similar patterns despite the use of different 16S regions (V2 and V4) and different sequencing technologies (454 and Illumina) (Fig. S3, Supporting information). Therefore, we assembled a comparative data set containing mammalian gut samples from 94 individuals and used only V2 data in cases where no V4 data were available.

We show the taxa relative abundance of the corresponding microbial communities in the phylogenetic context of the 47 mammalian species considered (Fig. 2). Taxa summary charts at the phylum level confirm that mammalian gut microbial communities are dominated by members of a reduced number of bacterial phyla, with Firmicutes, Bacteroidetes, Proteobacteria and Tenericutes being the most abundant in our data set. Large variations in the proportion of these bacterial phyla are seen across the mammalian phylogeny. For example, large proportions of Firmicutes and Bacteroidetes characterize gut microbiomes of Artiodactyla, whereas Firmicutes are mostly predominant in Carnivora. Xenarthran gut microbiomes seem to have relatively high proportions of Proteobacteria, which are especially abundant in anteaters. Within Xenarthra, sloths are characterized by a high abundance of Bacteroidetes, as it is also the case in elephants and hyraxes within Afrotheria.

The comparison of taxa summary plots computed at the family level and ordered by either host taxonomy or host diet (Fig. S4, Supporting information) highlights the confounding effects of phylogeny and diet on mammalian gut microbiome evolution that have been shown in previous studies (Ley *et al.* 2008b; Muegge *et al.* 2011). The two taxa charts display evident cases of clustering of similar gut microbial community composition in species belonging to the same mammalian order or sharing the same diet. For example, herbivorous and folivorous species belonging to different orders (Diprotodontia, Hyracoidea, Proboscidea, Lagomorpha, Perissodactyla and Artiodactyla for herbivores, and Pilosa (sloths) and primates (colobus monkey) for folivores) appear to have similar gut microbiota compositions characterized by high proportions of Bacteroidales and Ruminococcaceae. These phylogenetically diverse species sharing similar diets seem to have converged on gut microbiomes of similar taxonomic compositions confirming herbivory as a major driver of gut microbiome evolution in mammals (Ley *et al.* 2008b).

To further explore the roles of phylogeny and diet in shaping mammal gut microbial communities, we compared beta-diversity based on UniFrac distances computed between the 94 samples. The neighbour-net network reconstructed from the unweighted UniFrac distance matrix revealed interesting patterns of clustering by both host taxonomy and diet (Fig. S5, Supporting information). However, we also noticed in this network a potential effect of zoo vs. wild sampling, with some field-collected samples of giant anteaters, giant armadillos and pichi clustering with the wild nine-banded armadillo samples. This effect, potentially driven by the distinctiveness of the numerous nine-banded armadillo samples, is best visualized in a PCoA of unweighted

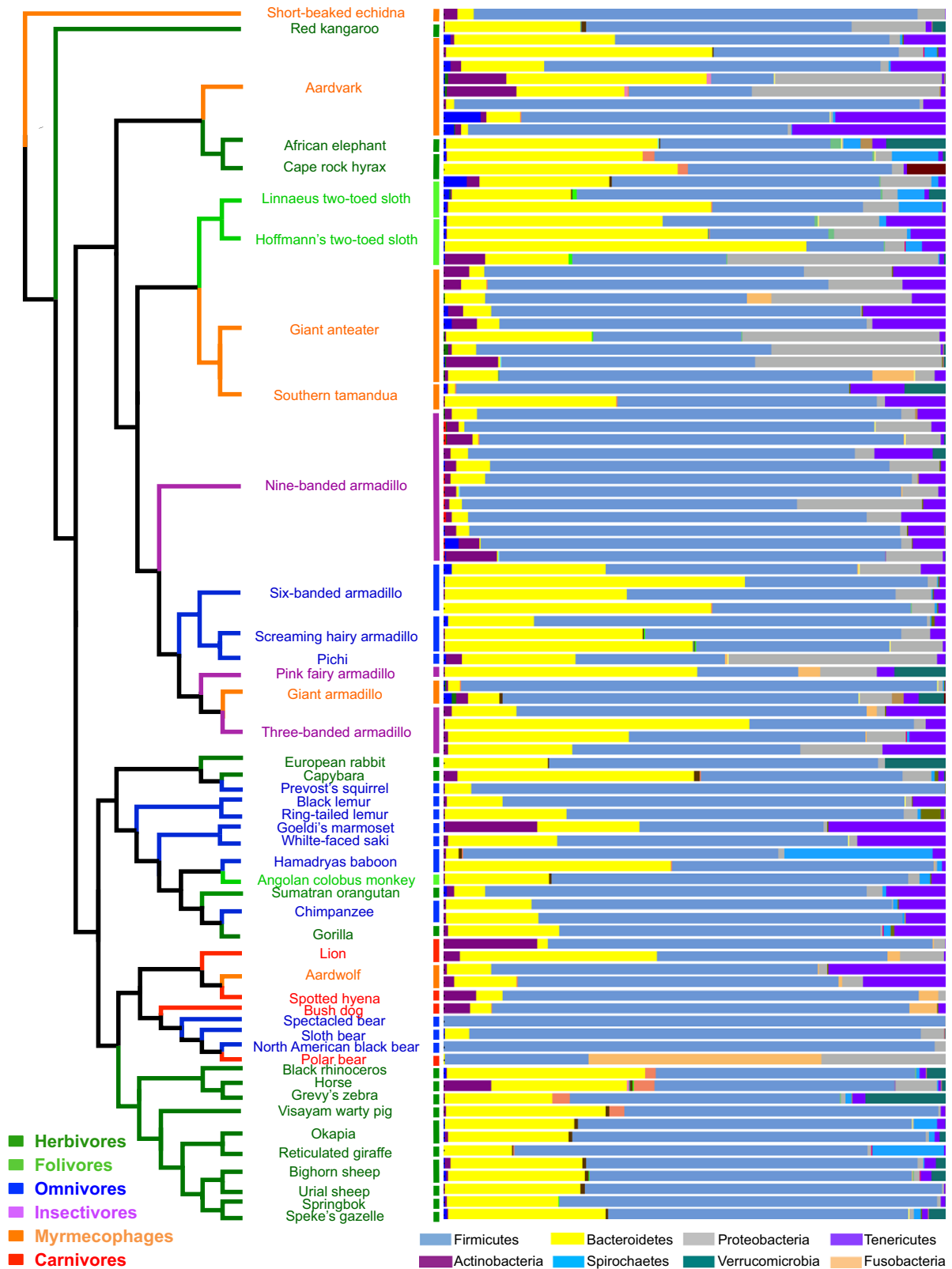


Fig. 2 Phylogeny of the 47 mammalian species represented by their corresponding gut microbiome compositions estimated at the phylum level from a total of 94 faecal samples. The mammalian phylogenetic tree coloured by diet (see inset legend) follows Delsuc *et al.* (2012) for xenarthrans and Meredith *et al.* (2011) for other mammals. The eight most represented bacterial phyla are indicated.

UniFrac distances (Fig. S6, Supporting information). Although this effect appears less important than clustering by taxonomy (ANOSIM host order: $R = 0.628$, $P = 0.001$) or by diet (ANOSIM host diet: $R = 0.394$, $P = 0.001$; Table S6, Supporting information), it is nevertheless statistically significant (ANOSIM captive vs. wild $R = 0.178$, $P = 0.002$). As most samples in the mammalian comparative data set were of captive origins, we excluded field-collected samples to prevent the influence of this potentially confounding variable in downstream analyses. This resulted in a final data set of 69 captive mammals, with all diet categories still represented including myrmecophages, insectivores, carnivores, omnivores, folivores and herbivores.

The neighbour-net network reconstructed from the unweighted UniFrac distance matrix on this reduced data set still shows a clear pattern of clustering by both host diet and taxonomy (Fig. 3). In particular, all

herbivores except the gorilla form a well-defined cluster within which we retrieved the classic divide into two distinct groups corresponding to foregut (artiodactyls) and hindgut (horses, rhinos, capybaras, hyraxes and elephants) fermenters. This analysis also revealed the distinctiveness of the gut microbiomes of the folivorous two-toed sloths (*Choloepus didactylus* and *Choloepus hoffmanni*), which form a well-defined group clearly separated from other mammals, but nevertheless close to herbivores. Primates (lemurs, sakis, baboons, gorillas and chimps) that are primarily omnivorous also group together. Similarly, most members of Carnivora (hyena, bush dog, lions and bears) also belong to a distinct cluster with the notable exceptions of the sloth bear and aardwolves that are well separated. Aardwolves are in fact part of a well-defined cluster of myrmecophagous species together with the large majority of aardvark and anteater samples. Among myrmecophages, only two

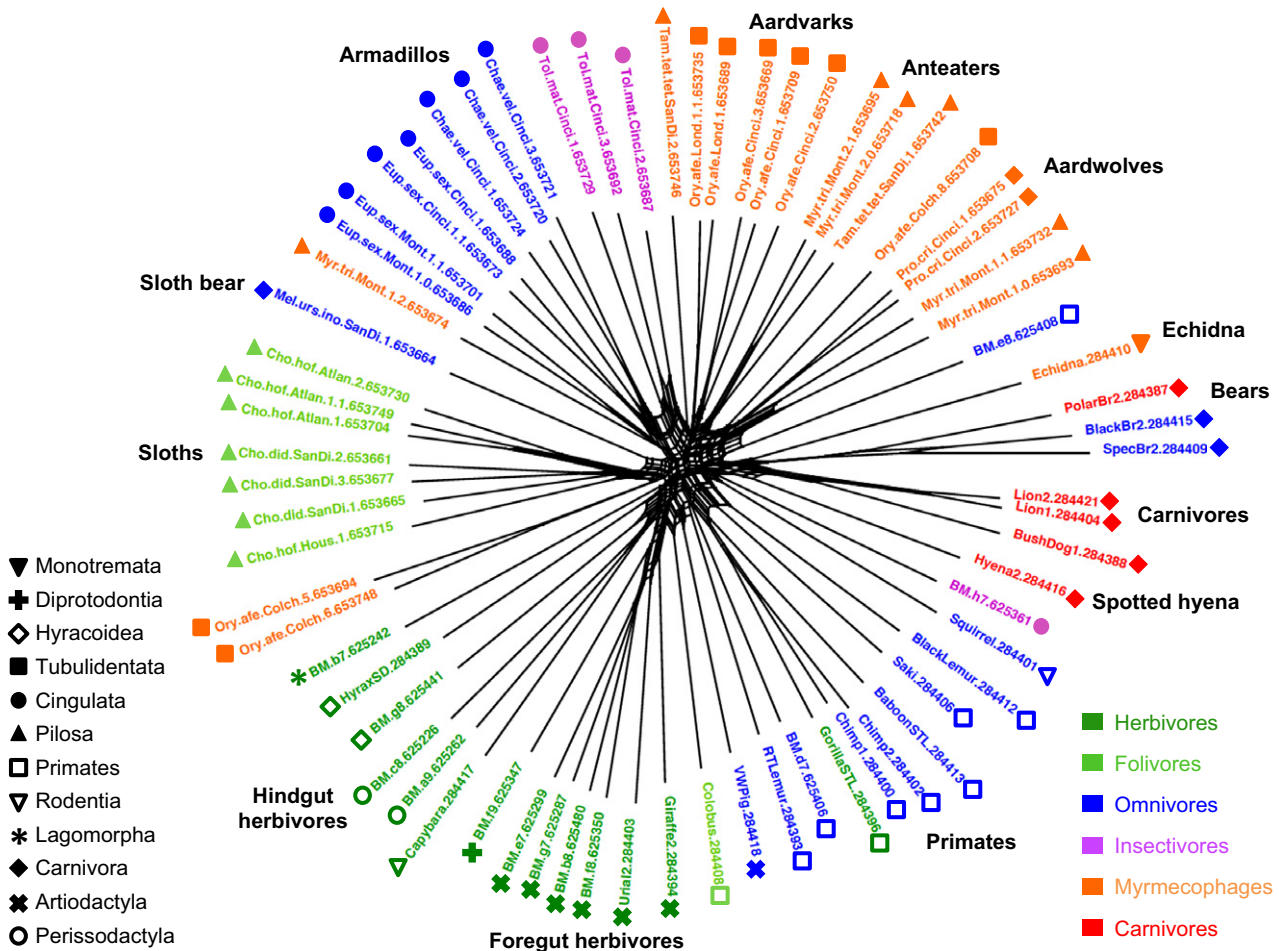


Fig. 3 Neighbour-net network reconstructed from unweighted UniFrac distances between 69 mammalian gut samples collected from captive animals. Samples are coloured by diet, and symbols indicate mammalian orders (see inset legend). Neighbour-net is an agglomerative method based on the neighbour-joining algorithm and is useful for representing the overall structure of the data as a network while accounting for uncertainty.

aardvarks, a giant anteater sample, and the distantly related echidna (*Tachyglossus aculeatus*) appear separated from the other species. Finally, armadillos form another obvious group divided into two distinct clusters corresponding to omnivorous (hairy and six-banded armadillos) and insectivorous (three-banded armadillos) species and it appears close to myrmecophages in the network.

Beta-diversity analysis of this 69-sample captive data set using PCoA of unweighted UniFrac distances confirmed a strong signal for sample clustering by both diet and host order (Fig. 4). The major clusters previously identified in the neighbour-net network are visually evident on the PCoA plot, with herbivores and myrmecophages being distributed at the two extremes of the first axis. Also, sloths appear to form a distinct group, which is nevertheless close to herbivores. This clustering pattern by both diet and host order is significant as assessed by a crossed ANOSIM (diet $R = 0.559$, $P = 0.001$; Table S6, Supporting information; host order $R = 0.62$, $P = 0.001$). However, because our data set contains uneven numbers of diet and host order categories, we randomly subsampled the data so that each category had the same number of samples and reran the ANOSIM 100 times. Our results are robust to uneven sampling as both factors were significant at $P = 0.001$ in all 100 subsample tests. Examining the pairwise ANOSIM results across diet categories shows that myrmecophages are significantly different from most other diets ($P = 0.001$), although a comparison of variances within groups using a Bartlett test (Bartlett 1937) shows that the amount of dispersion among these groups is not equal (Table S6, Supporting information), which can

also lead to statistically significant differences. Insectivores, however, are not significantly different from myrmecophages ($R = 0.196$, $P = 0.137$). Accordingly, in the PCoA (Fig. 4), samples from the insectivorous three-banded armadillo (Order Cingulata, pink circle) appear superimposed with the myrmecophagous aardvark (Order Tubulidentata, orange square) and anteater samples (Order Pilosa, orange triangle). Two other notable results from this beta-diversity analysis are the fact that both the termite-eating aardwolf (*Proteles cristata*) and the sloth bear (*Melursus ursinus*), which includes a significant portion of termites and ants in its diet (Joshi *et al.* 1997), are outliers within Carnivora. The aardwolf firmly clusters with other specialized myrmecophages (anteaters and aardvarks), rather than with other members of Carnivora, including its closest relative the spotted hyena (*Crocuta crocuta*). The sloth bear also appears clearly separated from the other members of Carnivora including its close bear relatives.

Finally, we explored the distinctiveness of myrmecophage gut microbiota composition by performing comparisons with their closest relatives using ANOVA (Table 1). The comparisons of the relative abundance of all bacterial taxa present indicate that the aardwolf gut microbiota is significantly enriched in *Prevotella* (Prevotellaceae), *Streptococcus* (Streptococcaceae), *Dialister* (Veillonellaceae), *Klebsiella* (Enterobacteriaceae), *Faecalibacterium* (Ruminococcaceae), *Eubacterium* (Lachnospiraceae) as well as unclassified genera of Erysipelotrichaceae and Clostridiaceae, as compared to other members of Carnivora. Similar comparisons conducted between anteaters and their sloth relatives also revealed a significant enrichment in *Streptococcus* and other

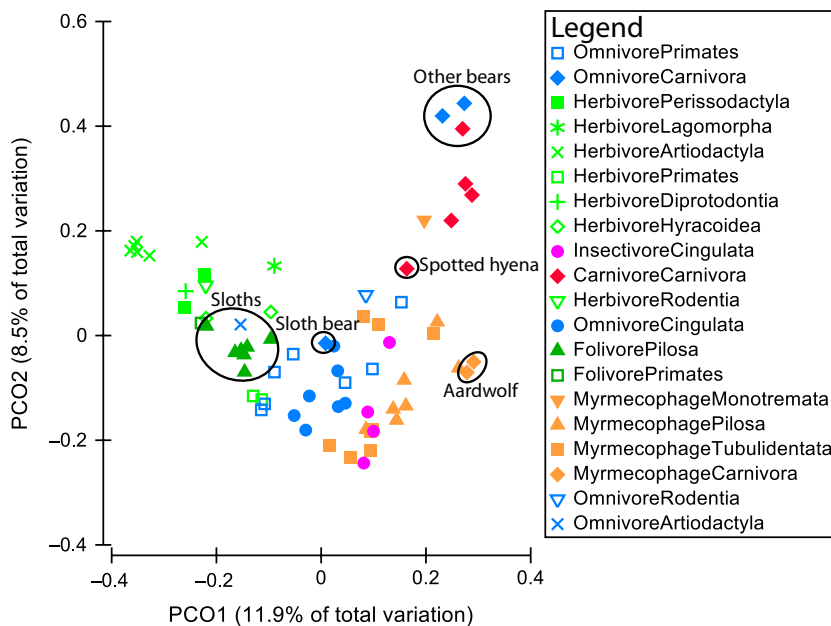


Fig. 4 Mammalian faecal bacterial communities clustered using principal coordinates analysis of the unweighted UniFrac distance matrix. Each point corresponds to a faecal sample coloured according to diet with different symbols corresponding to host order (see legend inset). The respective positions of the sloths, spotted hyena, aardwolf, sloth bear and other bears are indicated. In several cases, host taxa gut microbiomes are clustered by diet. A couple of striking examples include the sloth bear, which has a gut microbiome that is more similar to that of sloths than other bears; and the aardwolf, which has a gut microbiome that clusters more closely with the microbiomes of myrmecophagous host taxa rather than with that of its closest relative, the spotted hyena.

bacterial taxa in anteaters' gut microbiotae. However, when all related-taxon comparisons are considered, very few genera appear to be consistently more abundant in the myrmecophagous than in nonmyrmecophagous species. Furthermore, when mean relative abundances of these bacterial taxa are calculated across all diet groups represented in the data set (Table S5), it becomes less clear as to whether these same taxa may be considered to be uniquely associated with myrmecophagous mammals, with insectivores in general or even more broadly with omnivorous mammals that include insects in their diet. For example, although *Streptococcus* is found to be significantly higher in abundance in aardwolves and giant anteaters compared with their nonmyrmecophagous relatives, it appears to be enriched at similar levels in omnivores as in myrmecophages when compared with the other diet categories (Table S5).

Discussion

Diversity of mammalian gut microbiomes and the problem of wild vs. zoo sampling

Our study is the first large-scale comparative study of the gut microbiome in mammals that includes a significant number of samples collected in the field. Although previous studies reported no effect of captivity (Ley *et al.* 2008b; Muegge *et al.* 2011), we observed systematic differences between wild and zoo specimens. These differences are not altogether unexpected given the large differences in diet that may exist between wild and captive settings. The effect might be especially noticeable in animals for which diet in captivity is markedly different than in nature, which is the case for xenarthrans in general and especially myrmecophagous mammals (Superina 2011). It is also worth noting that of the samples that were detected as containing soil microbes, a large proportion were from wild animals. For example, field-collected samples from giant anteaters and giant armadillos showed signs of soil contamination with notably one of the giant armadillo samples assigning 56% to soil. Although our application of source tracking allowed us to identify faecal samples that may have truly been contaminated with soil, we recognize that a signature of soil microbes may not be an indication of contamination, but rather a natural feature of myrmecophagous mammals feeding on ant and termite nests, as soil is ingested along with their prey (Redford 1987; Anacleto 2007; Vaz *et al.* 2012). For example, soil particles are generally found in most armadillo species faeces (Anacleto 2007) with an estimate of 17% in the nine-banded armadillo (Beyer *et al.* 1994). Also, some ant-eating species kept in captivity, such as the Chinese pangolins we sampled, are main-

tained on a specific diet incorporating additional chitin and soil to facilitate ingestion (Yang *et al.* 2007). Also, we could not exclude the possibility that for field-collected samples of myrmecophages, a certain proportion of the 16S rRNA sequences obtained might actually come from ants and especially termites, which contain a diversity of associated bacteria (Köhler *et al.* 2012).

Diet differences between wild and captive herbivores and omnivores, on the other hand, are probably smaller than for species with more specialized diets such as carnivores and insectivores for which special diets have to be designed in captivity (Yang *et al.* 2007; Superina 2011). Accordingly, only minor differences were reported in studies comparing the gut microbiomes of wild and captive pandas (Zhu *et al.* 2011) and of domestic vs. feral goats (De Jesus-Laboy *et al.* 2011). The few wild samples included in Ley *et al.* (2008b) and Muegge *et al.* (2011) were from omnivorous and herbivorous species, which may explain why the differences were not significant.

Studies of host-associated microbiota in wild vertebrate populations have recently begun to flourish, with host taxa covered to date including primates (Ochman *et al.* 2010; Yildirim *et al.* 2010), North American moose (Ishaq & Wright 2012), bats (Phillips *et al.* 2012), capybaras (Garcia-Amado *et al.* 2012), hoatzins (Godoy-Vitorino *et al.* 2012), iguanas (Lankau *et al.* 2012) and house mice (Linnenbrink *et al.* 2013). These studies have generally revealed diverse gut microbiota in wild populations and the influence of environmental factors such as geography that can affect both interspecific (Phillips *et al.* 2012) and intraspecific variations (Godoy-Vitorino *et al.* 2012; Linnenbrink *et al.* 2013). This suggests that broad sampling across the range of each species might be necessary to gain a full understanding of microbial diversity. For example, in wild populations of house mice, biogeography was identified as the main driving factor of microbiome structure (Linnenbrink *et al.* 2013), and similar patterns have been observed in humans (Yatsunenkov *et al.* 2012), although geographical variation is often also confounded by variation in diet and host genetics.

In our analyses, field-sampled nine-banded armadillos (*Dasypus novemcinctus*) appear to have distinct gut microbiota from other armadillos and from mammals in general. This species is the most widespread xenarthran species ranging from northern Argentina to the southeastern United States and is thus adapted to a wide range of environmental conditions. It is also worth noting that nine-banded armadillos entered Texas only relatively recently (around 1850 A.D.). They quickly become invasive and rapidly spread eastwards to meet founding populations accidentally introduced in Florida later (Taulman & Robbins 1996; Loughry *et al.* 2009). This recent invasion was accompanied by severe genetic

bottlenecks, likely due to successive founder effects (Huchon *et al.* 1999). The individuals we sampled came from the removal programme of an invasive population conducted at Merritt Island National Wildlife Refuge in Florida in 2012. The gut microbiota of these individuals are likely not to be representative of the full range of gut microbiome diversity of this species, as invasive populations have shifted from a mostly myrmecophagous diet in their native range to a more omnivorous diet in the newly colonized northern area (Redford 1986; Sikes *et al.* 1990). There is the possibility that the invasive population may have also shifted its microbiota structure, or, intriguingly, the microbiome may have contributed to its success as an invasive species. Alternatively, the observed distinctiveness of nine-banded armadillo gut microbiota might also reflect a wild vs. captive pattern as suggested by its clustering with other field-collected samples in the neighbour-net analysis (Fig. S5). Our results highlight the need for more comparative sampling in wild specimens to capture the essence of intraspecific variation in the gut microbiome across species range.

Convergence of gut microbiota in myrmecophagous mammals

Our expanded analyses including wild and captive animals and a wider range of diets confirm both diet and phylogeny as major drivers of the microbiome composition as shown in previous studies (Ley *et al.* 2008b; Muegge *et al.* 2011). The previously unstudied myrmecophagous species appear to span a large area of the beta-diversity plots, which highlight marked differences in the composition of their gut bacterial communities. Although the microbiota of specialized myrmecophagous mammals show substantial variation, they cluster significantly with respect to the rest of the mammals. However, we note that in the reduced 69 captive animal data set, statistical significance for comparisons against certain diet types (i.e. myrmecophage vs herbivore and myrmecophage vs carnivore) could be driven by differences in dispersion between groups rather than true clustering (see Table S6, Supporting information). Yet supporting statistical evidence, microbial clustering by diet is also evident in neighbour-net networks where the gut microbiota of myrmecophages appear significantly distinct from that of animals with other diets, except insectivores. Indeed, the gut microbiota of insectivorous three-banded armadillos (*Tolypeutes matacus*) were clustered with that of aardvark and anteaters in the PCoA and appeared close to myrmecophage gut communities in network analyses. The fact that three-banded armadillos are opportunistic insectivores eating substantial quantities of ants and termites (Redford 1985; Bolković *et al.* 1995)

might explain this similarity in gut microbiota composition. However, phylogenetic inertia might also play a role as the gut communities of omnivorous armadillos (*Chaetophractus* and *Euphractus*) cluster with those of the three-banded armadillos in the neighbour-net network. The future inclusion of additional insectivorous taxa from different orders such as bats, shrews and tenrecs will allow disentangling the two effects. These analyses illustrate the utility of network representations for comparing bacterial communities based on measures such as UniFrac as a complement to classical PCoA. They provide a useful alternative to clustering methods such as UPGMA, which do not take into consideration the occurrence of conflicting signals in the data when visualizing the similarity between bacterial communities (Parks & Beiko 2012).

Our results revealed that aardvarks, anteaters and aardwolves possess similar gut microbiota despite representing highly distinct phylogenetic lineages that diverged some 100 million years ago (Mya; Meredith *et al.* 2011). This strongly suggests a major role for diet in driving the convergent composition of the gut microbiome in these specialized myrmecophages, paralleling the case of hindgut fermenter herbivores composed of phylogenetically diverse lineages such as horses, rhinos, capybaras, hyraxes and elephants (Muegge *et al.* 2011). In contrast, groups such as primates and carnivores show signs of phylogenetic inertia, with members of these groups forming well-defined clusters consisting of species with different diets. This clearly is not the case within myrmecophages where anteaters have diverged considerably from their closest relatives in xenarthrans (i.e. sloths), aardvarks appear very distinct from the other afrotherians (i.e. elephants and hyraxes), and the aardwolf has diverged substantially from the other members of Carnivora (hyenas, lions, bush dog and bears). This latter case represents an interesting contrast to the case of the giant panda, which clusters with other Ursidae despite its distinct bamboo diet (Ley *et al.* 2008b).

Both of these cases represent recent transitions to a highly specialized diet in the two distinct groups of Carnivora (Eizirik *et al.* 2010). The aardwolf (*Proteles cristata*) is a strict myrmecophage, almost exclusively feeding on the termite genus *Trinivertermes* with some seasonal variations (Kruuk & Sands 1972; Redford 1987). The aardwolf represents the sister lineage to all other hyenas (Koepfli *et al.* 2006) from which it diverged <10 Mya (Eizirik *et al.* 2010). Despite this relatively recent divergence, this species shows marked morphological and physiological differences from carnivorous hyenas, such as a reduced dentition, elongated muzzle and tongue, and digestive tract modifications such as a muscular stomach and relatively short small intestine, to facilitate the rapid processing of termites (Anderson *et al.* 1992).

The giant panda (*Ailuropoda melanoleuca*) diverged from all other ursids some 19 Mya (Krause *et al.* 2008; Eizirik *et al.* 2010), but the fossil record estimate for its dietary shift to a bamboo-rich diet is at least 7 Mya (Jin *et al.* 2007). Despite the fact that the aardwolf and the giant panda present some parallel aspects in their diet adaptation, the trajectories followed by their gut microbiota nevertheless differ greatly. The panda has retained a carnivore-like gut microbiome (Ley *et al.* 2008b) and uses specific genes from its bacterial taxa to facilitate cellulose and lignin degradation, even though their significance for the host physiology has yet to be demonstrated (Zhu *et al.* 2011; Fang *et al.* 2012). In the case of the aardwolf, its gut microbiome appears to have shifted in overall composition to converge with other myrmecophages. In contrast to the giant panda, the sloth bear (*Melursus ursinus*) is an omnivore specialized on ants and termites when available, but also exhibits seasonal variations in its diet (Joshi *et al.* 1997). The sloth bear diverged from other Ursinae around 7 Mya (Krause *et al.* 2008), and morphological adaptations to its myrmecophagous diet include an elongated snout with bare lips and the lack of upper incisors. It is interesting to note that this species is the only member of Carnivora (excluding the aardwolf) that does not cluster with other bears in neighbour-net networks and beta-diversity plots. This suggests that despite its more recent divergence from other bears and its less specialized diet than the panda, the sloth bear experienced an overall shift in the composition of its gut microbiota.

Finally, our results also provide new insights into the evolution of sloth gut microbiota. Sloths are poorly studied arboreal folivores with multichambered stomachs allowing the fermentation of plant material in a manner analogous to foregut fermenters such as artiodactyls (Foley *et al.* 1995). In our analyses, two-toed sloths (genus *Choloepus*) present distinct gut microbial communities, but nevertheless appear close to herbivores in neighbour-net analyses and in intermediate position between omnivores and herbivores in beta-diversity plots. They do not cluster firmly with the colobus monkey (*Colobus guereza*), the only other folivorous species, nor with artiodactyls or hindgut fermenters. There is thus some degree of convergence in gut microbiota composition between folivorous sloths and mammalian herbivores, but the sloth microbiome seems to be rather distinct suggesting substantial differences in digestion modes in these xenarthrans relative to other mammals.

Potential chitin degraders and prospects for future metagenomic surveys

The myrmecophagous diet imposes strong nutritional challenges to mammals because a large proportion of

the protein value of termites and ants consists of the chitin constituting their exoskeletons. In terms of nutritional values, ants and termites are not especially different from other terrestrial invertebrates; however, their larval and alate forms contain much more fat and are thus the frequent prey of many species of mammals (Redford & Dorea 1984). Apart from a muscular stomach thought to assist in the mechanical processing of large amounts of social insects, myrmecophagous mammalian digestive tracts do not show obvious anatomical adaptations that might reflect an increased potential for chitin degradation (Stevens & Hume 2004). Nevertheless, the macroscopic inspection of faecal samples in diet characterization studies (Kruuk & Sands 1972; Taylor *et al.* 2002; Miranda *et al.* 2009) and the results of some rare functional assays (Cooper & Withers 2004) suggest that substantial variations exist among mammalian myrmecophages with regard to chitin degradation and assimilation.

The degradation of chitin in nature is primarily carried out by bacterial taxa such as pseudomonads, enteric bacteria, gliding bacteria, actinomycetes and members of the genera *Bacillus*, *Vibrio* and *Clostridium* (Gooday 1990). Endogenous chitinases have also been reported in plants, invertebrates, fungi and vertebrates including mammals. Thus, there are three possible sources of chitinolytic enzymes in the digestive system: from the animal itself, from its associated gut microbiome or from the ingested food (Gooday 1990). Chitinase genes are found in vertebrate genomes, but their exact function and role in digestion are not fully understood in mammals (Bussink *et al.* 2007; Funkhouser & Aronson 2007). These enzymes are nevertheless believed to be important for chitin degradation in marine fishes feeding on crustaceans (Gutowska *et al.* 2004), and chitinase activity has been detected in the stomach of insectivorous nine-banded armadillos (Smith *et al.* 1998). It has also been recently demonstrated that these enzymes might also have a digestive activity in the human gastric juice (Paoletti *et al.* 2007). However, there is also evidence that chitinolytic bacteria are an integral part of mammalian gut microbiomes (Simůnek *et al.* 2001). Some studies have suggested that the forestomach of minke whales (*Balaenoptera acutorostrata*) includes bacterial taxa that can digest the chitinous exoskeleton of krill (Mårtensson *et al.* 1994) and that insectivorous bats use symbiotic gut bacteria to assimilate the chitin of their insect prey (Whitaker *et al.* 2004).

A key question is thus whether myrmecophagous mammals possibly use symbiotic bacteria to help them digest the chitin of ant and termite exoskeletons that would account at least partially for the observed convergence in the composition of their gut microbiota. If this process occurs, are the same bacterial taxa

recruited independently in different myrmecophagous lineages or do different bacterial taxa provide the same functions? Although we tentatively identified some bacterial genera whose abundance is significantly increased in myrmecophagous species, among which *Lactococcus* is a promising candidate with species known to degrade chitin such as *L. lactis* (Vaaje-Kolstad *et al.* 2009), we do not have an understanding of the functional gene content associated with these taxa. For example, processes such as gene loss and lateral gene transfer occurring among gut and environmental bacteria (Hehemann *et al.* 2010) could lead to a weak correspondence between the taxonomic classification and the functional role of host-associated microbes. An investigation of the functional gene content of myrmecophagous gut microbial communities through shotgun metagenomic approaches could help elucidate whether these and other taxa provide such functional roles.

A previous shotgun metagenomic survey illustrated the point that mammalian gut microbial communities can differ in their composition, but nevertheless share a functional core of genes ensuring similar functions (Muegge *et al.* 2011). Such functional metagenomic studies of the mammalian gut microbiome are currently expanding, as illustrated by recent surveys conducted in pig (Lamendella *et al.* 2011), Iberian lynx (Alcaide *et al.* 2012) and pygmy loris (Xu *et al.* 2013), providing comparative data for future studies of myrmecophagous mammals. The study on the endangered Iberian lynx (*Lynx pardalis*) suggests the intriguing possibility that the lynx microbiome evolved to digest not only the meat of its exclusive prey (the European rabbit, *Oryctolagus cuniculus*) but also the plants constitutive of its diet (Alcaide *et al.* 2012). For myrmecophagous animals, such studies would focus on genes and pathways involved in chitin degradation, which might have been convergently recruited in ant-eating placentals. Future genomic, metagenomic and functional studies of myrmecophagous placental species and their associated oral and gut microbiomes should help reveal the complex interplay between the host genome and its associated microbiome in the adaptation to the myrmecophagous diet. Testing whether adaptation to myrmecophagy consists primarily of differences in which taxa are present, in which genes supply essential functions, or in gene expression thus remains a key challenge for the field, although the present study provides a substantial advance in our understanding of the overall patterns of microbial communities associated with this extreme dietary adaptation.

Acknowledgements

We would like to thank Gail Ackermann, Catherine Nicholas, Chris Lauber, Donna Berg-Lyons, Matthew J.

Gebert, Gregory Humphrey and Yoshiki Vázquez Baeza for their help in data generation and processing. This study would not have been possible without the contribution of the following people and institutions in facilitating access to biological samples: Carly Vynne, Jim Loughry, Mariella Superina, David Gomis, Cédric Libert and Yann Raulet (Zoo de Lunaret Montpellier), Cynthia Steiner and Josephine Braun (San Diego Zoo), Andreas Bernhard (Leipzig Zoo), Angela Ryan and Amanda Ferguson (Regent's Park, Zoological Society of London), Sarah Forsyth (Colchester Zoo), Terri Roth (Cincinnati Zoo), Joseph Mendelson (Atlanta Zoo), Joseph Flanagan and Lauren Howard (Houston Zoo). We wish to thank the National Ecological Observatory Network (a project sponsored by the National Science Foundation and managed under cooperative agreement by NEON, Inc.) for donation of the soil samples that we used in source-tracking analyses. Finally, we thank the subject editor and three anonymous referees for numerous constructive and thoughtful comments. This work has been financially supported by grants from the France-US Fulbright Commission to FD and from the Howard Hughes Medical Institute to RK. This is publication ISEM 2013-108 of the Institut des Sciences de l'Evolution de Montpellier.

References

- Alcaide M, Messina E, Richter M *et al.* (2012) Gene sets for utilization of primary and secondary nutrition supplies in the distal gut of endangered Iberian lynx. *PLoS One*, **7**, e51521.
- Anacleto TCS (2007) Food habits of four armadillo species in the Cerrado area, Mato Grosso, Brazil. *Zoological Studies*, **46**, 529–537.
- Anderson MD, Richardson PRK, Woodall PF (1992) Functional analysis of the feeding apparatus and digestive tract anatomy of the aardwolf *Proteles cristatus*. *Journal of Zoology*, **228**, 423–434.
- Anderson KE, Russell JA, Moreau CS *et al.* (2012) Highly similar microbial communities are shared among related and trophically similar ant species. *Molecular Ecology*, **21**, 2282–2296.
- Bartlett MS (1937) Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London Series A*, **160**, 268–282.
- Beyer WN, Connor EE, Gerould S (1994) Estimates of soil ingestion by wildlife. *The Journal of Wildlife Management*, **58**, 375–382.
- Bokulich NA, Subramanian S, Faith JJ *et al.* (2013) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, **10**, 57–59.
- Bolković ML, Caziani SM, Protomastro JJ (1995) Food habits of the three-banded armadillo (Xenarthra: Dasypodidae) in the dry Chaco, Argentina. *Journal of Mammalogy*, **76**, 1199–1204.
- Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.
- Bussink AP, Speijer D, Aerts JM, Boot RG (2007) Evolution of mammalian chitinase(-like) members of family 18 glycosyl hydrolases. *Genetics*, **177**, 959–970.

- Caporaso JG, Bittinger K, Bushman FD *et al.* (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*, **26**, 266–267.
- Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.
- Caporaso JG, Lauber CL, Walters WA *et al.* (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences, USA*, **108**(Suppl. 1), 4516–4522.
- Caporaso JG, Lauber CL, Walters WA *et al.* (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal*, **6**, 1621–1624.
- Cooper CE, Withers PC (2004) Termite digestibility and water and energy contents determine the water economy index of numbats (*Myrmecobius fasciatus*) and other myrmecophages. *Physiological and Biochemical Zoology*, **77**, 641–650.
- Clarke KR, Gorley RN (2006) PRIMER v6. *User manual/tutorial. Plymouth routine in multivariate ecological research. Plymouth Marine Laboratory.*
- De Jesus-Laboy KM, Godoy-Vitorino F, Piceno YM *et al.* (2011) Comparison of the fecal microbiota in feral and domestic goats. *Genes*, **3**, 1–18.
- Delsuc F, Vizcaino SF, Douzery EJP (2004) Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: a relaxed molecular clock study within xenarthrans. *BMC Evolutionary Biology*, **4**, 11.
- Delsuc F, Superina M, Tilak MK, Douzery EJP, Hassanin A (2012) Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. *Molecular Phylogenetics and Evolution*, **62**, 673–680.
- DeSantis TZ, Hugenholtz P, Larsen N *et al.* (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied Environmental Microbiology*, **72**, 5069–5072.
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, **26**, 2460–2461.
- Eilers KG, Debenport S, Anderson S, Fierer N (2012) Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. *Soil Biology and Biochemistry*, **50**, 58–65.
- Eizirik E, Murphy WJ, Koepfli KP *et al.* (2010) Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. *Molecular Phylogenetics and Evolution*, **56**, 49–63.
- Fang W, Fang Z, Zhou P *et al.* (2012) Evidence for lignin oxidation by the giant panda fecal microbiome. *PLoS One*, **7**, e50312.
- Feldhamer GA (2007) *Mammalogy: Adaptation, Diversity, Ecology*. John Hopkins University Press, Baltimore, Maryland.
- Foley WJ, Engelhardt WV, Charles-Dominique P (1995) The passage of digesta, particle size, and in vitro fermentation rate in the three-toed sloth *Bradypus tridactylus* (Edentata: Bradypodidae). *Journal of Zoology*, **236**, 681–696.
- Funkhouser JD, Aronson NN (2007) Chitinase family GH18: evolutionary insights from the genomic history of a diverse protein family. *BMC Evolutionary Biology*, **7**, 96.
- Garcia-Amado MA, Godoy-Vitorino F, Piceno YM *et al.* (2012) Bacterial diversity in the cecum of the world's largest living rodent (*Hydrochoerus hydrochaeris*). *Microbial Ecology*, **63**, 719–725.
- Gilbert JA, Meyer F, Jansson J *et al.* (2010) The Earth Microbiome Project: meeting report of the “1 EMP meeting on sample selection and acquisition” at Argonne National Laboratory October 6 2010. *Standards in Genomic Sciences*, **3**, 249–253.
- Godoy-Vitorino F, Goldfarb KC, Karaoz U *et al.* (2011) Comparative analyses of foregut and hindgut bacterial communities in hoatzins and cows. *ISME Journal*, **6**, 531–541.
- Godoy-Vitorino F, Leal SJ, Diaz WA *et al.* (2012) Differences in crop bacterial community structure between hoatzins from different geographical locations. *Research in Microbiology*, **163**, 211–220.
- Gooday GW (1990) The ecology of chitin degradation. *Advances in Microbial Ecology*, **11**, 387–430.
- Gutowska MA, Drazen JC, Robison BH (2004) Digestive chitinolytic activity in marine fishes of Monterey Bay, California. *Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology*, **139**, 351–358.
- Hehemann JH, Correc G, Barbeyron T *et al.* (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature*, **464**, 908–912.
- Huchon D, Delsuc F, Catzeflis FM, Douzery EJP (1999) Armadillos exhibit less genetic polymorphism in North America than in South America: nuclear and mitochondrial data confirm a founder effect in *Dasypus novemcinctus* (Xenarthra). *Molecular Ecology*, **8**, 1743–1748.
- Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.
- Ishaq SL, Wright AD (2012) Insight into the bacterial gut microbiome of the North American moose (*Alces alces*). *BMC Microbiology*, **12**, 212.
- Jin C, Ciochon RL, Dong W *et al.* (2007) The first skull of the earliest giant panda. *Proceedings of the National Academy of Sciences, USA*, **104**, 10932–10937.
- Joshi AR, Garshelis DL, Smith JL (1997) Seasonal and habitat-related diets of sloth bears in Nepal. *Journal of Mammalogy*, **78**, 584–597.
- Karasov WH, Martinez del Rio C, Caviedes-Vidal E (2011) Ecological physiology of diet and digestive systems. *Annual Review of Physiology*, **73**, 69–93.
- Knights D, Kuczynski J, Charlson ES *et al.* (2011) Bayesian community-wide culture-independent microbial source tracking. *Nature Methods*, **8**, 761–763.
- Koepfli KP, Jenks SM, Eizirik E *et al.* (2006) Molecular systematics of the Hyaenidae: relationships of a relictual lineage resolved by a molecular supermatrix. *Molecular Phylogenetics and Evolution*, **38**, 603–620.
- Köhler T, Dietrich C, Scheffrahn RH, Brune A (2012) High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Applied and Environmental Microbiology*, **78**, 4691–4701.
- Krause J, Unger T, Nocon A *et al.* (2008) Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. *BMC Evolutionary Biology*, **8**, 220.
- Kruuk H, Sands WA (1972) The aardwolf (*Proteles cristata*, Sparrman 1783) as predator of termites. *African Journal of Ecology*, **10**, 211–227.

- Lamendella R, Domingo JW, Ghosh S, Martinson J, Oerther DB (2011) Comparative fecal metagenomics unveils unique functional capacity of the swine gut. *BMC Microbiology*, **11**, 103.
- Lankau EW, Hong PY, Mackie RI (2012) Ecological drift and local exposures drive enteric bacterial community differences within species of Galapagos iguanas. *Molecular Ecology*, **21**, 1779–1788.
- Ley RE, Lozupone CA, Hamady M, Knight R, Gordon JI (2008a) Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Reviews Microbiology*, **6**, 776–788.
- Ley RE, Hamady M, Lozupone C *et al.* (2008b) Evolution of mammals and their gut microbes. *Science*, **320**, 1647–1651.
- Linnenbrink M, Wang J, Hardouin EA *et al.* (2013) The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Molecular Ecology*, **22**, 1904–1916.
- Liu Z, DeSantis TZ, Andersen GL, Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*, **36**, e120.
- Loughry WJ, Truman RW, McDonough CM, Tilak MK, Garnier S, Delsuc F (2009) Is leprosy spreading among nine-banded armadillos in the southeastern United States? *Journal of Wildlife Diseases*, **45**, 144–152.
- Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*, **71**, 8228–8235.
- Mårtensson PE, Nordøy ES, Blix AS (1994) Digestibility of krill (*Euphausia superba* and *Thysanoessa* sp.) in minke whales (*Balaenoptera acutorostrata*) and crabeater seals (*Lobodon carcinophagus*). *The British Journal of Nutrition*, **72**, 713–716.
- McDonald D, Price MN, Goodrich J *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME Journal*, **6**, 610–618.
- McFall-Ngai M, Hadfield MG, Bosch TC *et al.* (2013) Animals in a bacterial world, a new imperative for the life sciences. *Proceedings of the National Academy of Sciences, USA*, **110**, 3229–3236.
- McGhee GR (2011) *Convergent Evolution: Limited Forms Most Beautiful*. Massachusetts Institute of Technology Press, Cambridge, Massachusetts.
- McNab BK (1984) Physiological convergence amongst ant-eating and termite-eating mammals. *Journal of Zoology*, **203**, 485–510.
- Meredith RW, Janecka JE, Gatesy J *et al.* (2011) Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science*, **334**, 521–524.
- Miranda F, Veloso R, Superina M, Zara FJ (2009) Food habits of wild silky anteaters (*Cyclopes didactylus*) of São Luis do Maranhão, Brazil. *Edentata*, **8**, 1–5.
- Muegge BD, Kuczynski J, Knights D *et al.* (2011) Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*, **332**, 970–974.
- Ochman H, Worobey M, Kuo CH *et al.* (2010) Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biology*, **8**, e1000546.
- Paoletti MG, Norberto L, Damini R, Musumeci S (2007) Human gastric juice contains chitinase that can degrade chitin. *Annals of Nutrition and Metabolism*, **51**, 244–251.
- Parks DH, Beiko RG (2012) Measuring community similarity with phylogenetic networks. *Molecular Biology and Evolution*, **29**, 3947–3958.
- Phillips CD, Phelan G, Dowd SE *et al.* (2012) Microbiome analysis among bats describes influences of host phylogeny, life history, physiology and geography. *Molecular Ecology*, **21**, 2617–2627.
- Price MN, Dehal PS, Arkin AP (2010) FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*, **5**, e9490.
- Ramirez KS, Lauber CL, Knight R, Bradford MA, Fierer N (2010) Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology*, **91**, 3463–3470.
- Redford KH (1985) Foods habits of armadillos (Xenarthra: Dasypodidae). In: *The Evolution and Ecology of Sloths, Armadillos, and Vermilinguas*, pp. 429–437. Smithsonian Institution Press, Washington, D.C.
- Redford KH (1986) Dietary specialization and variation in two mammalian myrmecophages (variation in mammalian myrmecophagy). *Revista Chilena de Historia Natural*, **59**, 201–208.
- Redford KH (1987) Ants and termites as food: patterns of mammalian myrmecophagy. In: *Current Mammalogy*, vol. 1, pp. 349–399. Plenum Press, New York.
- Redford KH, Dorea JG (1984) The nutritional value of invertebrates with emphasis on ants and termites as food for mammals. *Journal of Zoology*, **203**, 385–395.
- Reiss KZ (2001) Using phylogenies to study convergence: the case of the ant-eating mammals. *American Zoologist*, **41**, 507–525.
- Russell JA, Moreau CS, Goldman-Huertas B *et al.* (2009) Bacterial gut symbionts are tightly linked with the evolution of herbivory in ants. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 21236–21241.
- Sikes RS, Heidt GA, Elrod DA (1990) Seasonal diets of the nine-banded armadillo (*Dasypus novemcinctus*) in a northern part of its range. *American Midland Naturalist*, **123**, 383–389.
- Simůnek J, Hodrová B, Bartonová H, Kopecný J (2001) Chitino-lytic bacteria of the mammal digestive tract. *Folia Microbiologica*, **46**, 76–78.
- Smith SA, Robbins LW, Steiert JG (1998) Isolation and characterization of a chitinase from the nine-banded armadillo, *Dasypus novemcinctus*. *Journal of Mammalogy*, **79**, 486–491.
- Springer MS, Murphy WJ, Eizirik E, O'Brien SJ (2003) Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proceedings of the National Academy of Sciences, USA*, **100**, 1056–1061.
- Stevens CE, Hume ID (2004) *Comparative Physiology of the Vertebrate Digestive System*. Cambridge University Press, Cambridge, UK.
- Sullam KE, Essinger SD, Lozupone CA *et al.* (2012) Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Molecular Ecology*, **21**, 3363–3378.
- Superina M (2011) Husbandry of a pink fairy armadillo (*Chlamyphorus truncatus*): case study of a cryptic and little known species in captivity. *Zoo Biology*, **30**, 225–231.
- Taulman JF, Robbins LW (1996) Recent range expansion and distributional limits of the nine-banded armadillo (*Dasypus novemcinctus*) in the United States. *Journal of Biogeography*, **23**, 635–648.

- Taylor WA, Lindsey PA, Skinner JD (2002) The feeding ecology of the aardvark *Orycteropus afer*. *Journal of Arid Environments*, **50**, 135–152.
- Vaae-Kolstad G, Bunaes AC, Mathiesen G, Eijsink VG (2009) The chitinolytic system of *Lactococcus lactis* ssp. *lactis* comprises a nonprocessive chitinase and a chitin-binding protein that promotes the degradation of alpha- and beta-chitin. *FEBS Journal*, **276**, 2402–2415.
- Vaz VC, Santori RT, Jansen AM, Delciellos AC, D'Andrea PS (2012) Notes on food habits of armadillos (Cingulata, Dasypodidae) and anteaters (Pilosa, Myrmecophagidae) at Serra Da Capivara National Park (Piauí State, Brazil). *Edentata*, **13**, 84–89.
- Vynne C, Skalski JR, Machado RB *et al.* (2011) Effectiveness of scat-detection dogs in determining species presence in a tropical savanna landscape. *Conservation Biology*, **25**, 154–162.
- Walters WA, Caporaso JG, Lauber CL *et al.* (2011) Primer-Prospector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*, **27**, 1159–1161.
- Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, **73**, 5261–5267.
- Whitaker JO Jr, Dannelly HK, Prentice DA (2004) Chitinase in insectivorous bats. *Journal of Mammalogy*, **85**, 15–18.
- Xu B, Xu W, Yang F *et al.* (2013) Metagenomic analysis of the pygmy loris fecal microbiome reveals unique functional capacity related to metabolism of aromatic compounds. *PLoS One*, **8**, e56565.
- Yang CW, Chen S, Chang CY *et al.* (2007) History and dietary husbandry of pangolins in captivity. *Zoo Biology*, **26**, 223–230.
- Yatsunencko T, Rey FE, Manary MJ *et al.* (2012) Human gut microbiome viewed across age and geography. *Nature*, **486**, 222–227.
- Yildirim S, Yeoman CJ, Sipos M *et al.* (2010) Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. *PLoS One*, **5**, e13963.
- Zhu L, Wu Q, Dai J, Zhang S, Wei F (2011) Evidence of cellulose metabolism by the giant panda gut microbiome. *Proceedings of the National Academy of Sciences, USA*, **108**, 17714–17719.

F.D., J.L.M and R.K. designed research. F.D. collected samples. R.K. and A.G. contributed new reagents and analytical tools. F.D., J.L.M., L.W.P and S.J.S. performed research and analysed data. F.D. and J.L.M. wrote the first draft of the manuscript. F.D., J.L.M., L.W.P, S.J.S., A.G. and R.K. participated in revising the final manuscript.

Data accessibility

The new 16S rRNA sequence data produced in this study are part of the Earth Microbiome Project and can

be freely accessed from a dedicated public database at <http://www.microbio.me/emp/> under study ID 1056 “Delsuc_microbi_ant”. The data are also available at EBI under Accession no. ERP003782 <http://www.ebi.ac.uk/ena/data/view/ERP003782>. Finally, mapping files and OTU tables used for this study may be downloaded from DRYAD at doi:10.5061/dryad.390 ng.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Information and basic statistics on the 93 samples newly sequenced in this study.

Table S2 Information on the soil samples included in the source tracking analysis.

Table S3 Information and basic statistics on the 39 additional mammalian samples included in our comparative study.

Table S4 Results from the source-tracking analysis using a previously published set of mammal gut samples, as well as soil samples, as potential sources to determine which of the 92 new samples were potentially contaminated with soil.

Table S5 Mean relative abundance (RA) of the taxa from Table 1 in the six diet groups: myrmecophages, insectivores, omnivores, carnivores, herbivores, and folivores.

Table S6 Results of ANOSIM and the Bartlett test of homogeneity of variances between diet groups for the full (All) and reduced (Captive) data sets, including pairwise comparisons of diet categories.

Fig. S1 Proportion of Phyla represented in sequences that did not hit Greengenes.

Fig. S2 Results from the source-tracking analysis using a previously published set of mammal gut samples, as well as soil samples, as potential sources to determine which of the 92 samples were potentially contaminated with soil.

Fig. S3 Procrustes analysis comparing unweighted UniFrac distances for 16 mammalian samples sequenced for V2 using 454 technology (Muegge *et al.* 2011) and resequenced as part of this study for the V4 region using an Illumina platform.

Fig. S4 Taxa summary plots at the order level for 94 fecal samples representing 47 mammalian species ordered by host diet (upper panel) and by host order (lower panel).

Fig. S5 Neighbor-Net network based on unweighted UniFrac distances between 94 mammalian gut samples colored by diet (see inset legend).

Fig. S6 Principal coordinate analysis (PCoA) of 94 mammalian samples (with potentially contaminated samples removed) based on unweighted UniFrac distances and colored by wild (red) vs. captive (blue) status.