Metagenomics uncovers dietary adaptations for chitin digestion in the gut microbiota of convergent myrmecophagous mammals

Sophie Teullet[a,#], Marie-Ka Tilak[a], Amandine Magdeleine[a], Roxane Schaub[b,c], Nora M. Weyer[d], Wendy Panaino[d,e], Andrea Fuller[d], W. J. Loughry[f], Nico L. Avenant[g], Benoit de Thoisy[h,i], Guillaume Borrel[j] and Frédéric Delsuc[a,#]

[a] Institut des Sciences de l'Evolution de Montpellier (ISEM), Univ Montpellier, CNRS, IRD, Montpellier, France

[b] CIC AG/Inserm 1424, Centre Hospitalier de Cayenne Andrée Rosemon, Cayenne, French Guiana

[c] Tropical Biome and immunopathology, Université de Guyane, Labex CEBA, DFR Santé, Cayenne, French Guiana

[d] Brain Function Research Group, School of Physiology, University of the Witwatersrand, Johannesburg, South Africa

[e] Centre for African Ecology, School of Animals, Plant, and Environmental Sciences, University of the Witwatersrand, Johannesburg, South Africa

[f] Department of Biology, Valdosta State University, Valdosta, GA, USA

[g] National Museum and Centre for Environmental Management, University of the Free State, Bloemfontein, South Africa

[h] Institut Pasteur de la Guyane, Cayenne, French Guiana, France

[i] Kwata NGO, Cayenne, French Guiana, France

[j] Institut Pasteur, Université Paris Cité, UMR CNRS 6047, Evolutionary Biology of the Microbial Cell, Paris, France

[#]Corresponding authors
sophie.teullet@umontpellier.fr

frederic.delsuc@umontpellier.fr

Abstract

In mammals, myrmecophagy (ant and termite consumption) represents a striking example of dietary convergence. This trait evolved independently at least five times in placentals with myrmecophagous species comprising aardvarks, anteaters, some armadillos, pangolins, and aardwolves. The gut microbiome plays an important role in dietary adaptation, and previous analyses of 16S rRNA metabarcoding data have revealed convergence in the composition of the gut microbiota among some myrmecophagous species. However, the functions performed by these gut bacterial symbionts and their potential role in the digestion of prey chitinous exoskeletons remain open questions. Using long- and short-read sequencing of fecal samples, we generated 29 gut metagenomes from nine myrmecophagous and closely related insectivorous species sampled in French Guiana, South Africa, and the USA. From these, we reconstructed 314 high-quality bacterial genome bins of which 132 carried chitinase genes, highlighting their potential role in insect prey digestion. These chitinolytic bacteria belonged mainly to the family Lachnospiraceae, and some were likely convergently recruited in the different myrmecophagous species as they were detected in several host orders (i.e., *Enterococcus faecalis*, *Blautia* sp), suggesting that they could be directly involved in the adaptation to myrmecophagy. Others were found to be more host-specific, possibly reflecting phylogenetic constraints and environmental influences. Overall, our results highlight the potential role of the gut microbiome in chitin digestion in myrmecophagous mammals and provide the basis for future comparative studies performed at the mammalian scale to further unravel the mechanisms underlying the convergent adaptation to myrmecophagy.

Importance

Myrmecophagous mammals are specialized in the consumption of ants and/or termites. They do not share a direct common ancestor and evolved convergently in five distinct placental orders raising questions about the underlying adaptive mechanisms involved and the relative contribution of natural selection and phylogenetic constraints. Understanding how these species digest their prey can help answer these questions. More specifically, the role of their gut microbial symbionts in the digestion of the insect chitinous exoskeleton has not been investigated in all myrmecophagous orders. We generated 29 new gut metagenomes from nine myrmecophagous species to reconstruct more than 300 bacterial genomes in which we identified chitin-degrading enzymes. Studying the distribution of these chitinolytic bacteria among hosts revealed both shared and specific bacteria between ant-eating species. Overall, our results highlight the potential role of gut symbionts in the convergent dietary adaptation of myrmecophagous mammals and the evolutionary mechanisms shaping their gut microbiota.

Keywords

convergent evolution; myrmecophagy; mammals; metagenomics; chitinases; genome binning

Introduction

The role played by microbial symbionts in animal evolution is increasingly recognized as they have interacted with their hosts for a long time and coevolved as changes occurred in both partners (McFall-Ngai *et al*, 2013). More specifically, the gut microbiota is involved in many host functions such as digestion, immunity, or behavior (Suzuki, 2017). Several factors such as host diet, physiology, genetics, and phylogeny, but also environmental factors, shape the taxonomic composition and functional structure of the gut microbiota (Alessandri *et al*, 2021), with resulting changes affecting host fitness (Suzuki, 2017; Moeller and Sanders, 2020). The mechanisms driving this coevolution are complex. Host evolutionary history influences how microbial communities assemble and evolve within their host, sometimes leading to closely related species sharing more similar microbiomes (i.e., phylosymbiosis). Phylosymbiosis is particularly strong in non-flying mammals (Song *et al*, 2020) likely due to mammalian-specific physiological traits (Mallott and Amato, 2021), and in internal host compartments such as the gut (Mazel *et al*, 2018). In mammals, the gut microbiota has played a major role in dietary diversification enabling transitions to novel carbon sources (Ley *et al*, 2008a; Moeller and Sanders, 2020). Consequently, many studies have focused on convergent dietary adaptations and the effects of host diet, investigating how host phylogeny shapes the gut microbiota (Ley *et al*, 2008b; Muegge *et al*, 2011; Groussin *et al*, 2017; Nishida and Ochman, 2018; Youngblut *et al*, 2019; Song *et al*, 2020; Thomas *et al*, 2022).

A striking example of convergent dietary specialization is the adaptation to myrmecophagy in placental mammals. Myrmecophagous species are characterized by a diet composed of at least 90% of ants and/or termites (Redford, 1987). A total of 22 placental species evolved toward this diet and are found in five independent orders: Tubulidentata (aardvarks), Pilosa (anteaters), Cingulata (tolypeutine armadillos), Pholidota (pangolins), and Carnivora (aardwolves) (Delsuc *et al*, 2002; Meredith *et al*, 2011; Álvarez-Carretero *et al*, 2021). Their divergent evolutionary histories raise the question of how these species convergently adapted to the myrmecophagous diet. Evolution toward myrmecophagy was accompanied by convergent morphological adaptations linked to this diet. Myrmecophagous species are usually characterized by an elongated skull and a protrusive tongue associated with the production of large amounts of viscous saliva thanks to hypertrophied salivary glands that help them catch their prey that are directly swallowed (Patterson, 1975; Griffiths, 1968; Reiss, 2001; Feirrera-Cardoso *et al*, 2019, 2022). The lack of mastication resulted in teeth reduction or even complete loss (e.g., anteaters, pangolins) and jaw muscle reduction (Patterson, 1975; Griffiths, 1986; Reiss, 2001; Feirrera-Cardoso *et al*, 2020). Additionally, myrmecophagous species need to digest the chitinous exoskeleton of their prey to extract nutrients. Mammals carry chitinase genes in their genomes, which participate in chitin digestion (Emerling *et al*, 2018; Ma *et al*, 2017, 2019; Allio *et al*, 2023), but their gut microbiota might also have played an important role in facilitating prey digestion (Delsuc *et al*, 2014; Ma *et al*, 2018; Cheng *et al*, 2022). Taxonomic analyses based on 16S rRNA sequences have highlighted similarities in the gut microbiota composition of myrmecophagous species (Delsuc *et al*, 2014) compared to their non-myrmecophagous sister-species. Focusing on the gut

archaeome of placentals, *Methanobrevibacter* have been found almost only in myrmecophagous species (Thomas *et al*, 2022). Yet, little is known regarding the role these symbionts play in the digestion of the chitinous exoskeleton of their prey. Chitinolytic bacteria are present in a diversity of environments (i.e., soils, sediments, aquatic environments) (Gooday, 1990; Beier and Bertilsson, 2013; Swiontek Brzezinska *et al*, 2014) including the digestive tract of mammals (Stevens and Hume, 1998; Šimůnek *et al*, 2001; Whitaker *et al*, 2004; Irulan *et al*, 2011; McDonald *et al*, 2014). More specifically, symbiotic chitin-degrading bacteria have been identified, as well as bacterial chitinase genes, in the Malayan pangolin and the giant anteater (Ma *et al*, 2018; Cheng *et al*, 2022) and in other mammals having a chitin-rich diet (Sanders *et al*, 2015; Kohl *et al*, 2022), suggesting other myrmecophagous species might carry chitinolytic gut bacteria that could participate in prey digestion.

As historical contingency and determinism can both play a role in convergent evolution (Losos, 2011; Blount *et al*, 2018), the foregoing raises the question of whether similar adaptive mechanisms were involved in the evolution of convergent ant-eating phenotypes. The different myrmecophagous mammals diverged millions of years ago and they evolved in diverse environments. Chitinolytic microbes might thus have been independently recruited to ensure chitin digestion, but not necessarily similar ones. Assessing whether chitinolytic bacteria present in myrmecophagous mammals are taxonomically and functionally similar will shed light on their origin. Metagenomics offers a valuable tool to answer these questions as it enables the study of both genes and functions carried out by symbionts. For instance, the direct reconstruction of bacterial genomes from metagenomes makes it possible to study complete gene sets present in a specific taxon and reconstruct complex metabolic pathways. Combined with other *-omics* approaches (e.g., metaproteomics, metabolomics) metagenomics has increased our knowledge of the taxonomic and functional diversity of numerous microbiomes from diverse environments (Nayfach *et al*, 2020; Shaffer *et al*, 2022) including the vertebrate gut microbiota (Youngblut *et al*, 2020; Levin *et al*, 2021). Long-read sequencing data can be used to more efficiently reconstruct bacterial genomes, as their longer size can facilitate the assembly process and it has recently enabled the reconstruction of bacterial genomes from diverse microbial communities (e.g., Lavrinienko *et al*, 2020; Arumugam *et al*, 2021; Singleton *et al*, 2021; Cuscó *et al*, 2022; Bickhart *et al*, 2022). To improve the quality of assembled genomes, short reads are still useful to correct long reads through hybrid assembly or for polishing long-read assemblies (e.g., Arumugam *et al*, 2021; Chen *et al*, 2021; Singleton *et al*, 2021).

To study the chitin-degrading potential of the gut microbiota of myrmecophagous mammals, we took advantage of these recent advances to reconstruct high-quality bacterial genomes and identify Glycoside Hydrolases 18 (GH18). This enzyme family belongs to the carbohydrate-active enzymes (CAZymes) (Drula *et al*, 2022) and comprises chitinases and chitin-binding proteins. Since bacterial chitinases mainly belong to this enzyme family (Adrangi and Faramarzi, 2013; Rathore and Gupta, 2015; Shakir *et al*, 2016), we focused on them to assess the potential of the gut microbiota of our focal species to digest chitin. By combining Nanopore long-read and Illumina short-read shotgun metagenomic sequencing,

we generated 29 new gut metagenomes from field-collected fecal samples of nine myrmecophagous and insectivorous species, from which we reconstructed 314 high-quality bacterial genomes harboring a diversity of GH18 chitinases. Identification of GH18 sequences in the same bacterial genomes revealed a potential role of gut symbionts in prey digestion through putative complex metabolic pathways. Both generalist and host-specific bacteria were identified in the different myrmecophagous species, potentially reflecting the divergent evolutionary histories of their hosts, and raising questions about the evolutionary forces (historical contingency and determinism) at play in shaping the gut microbiome of convergently evolved myrmecophagous mammals.

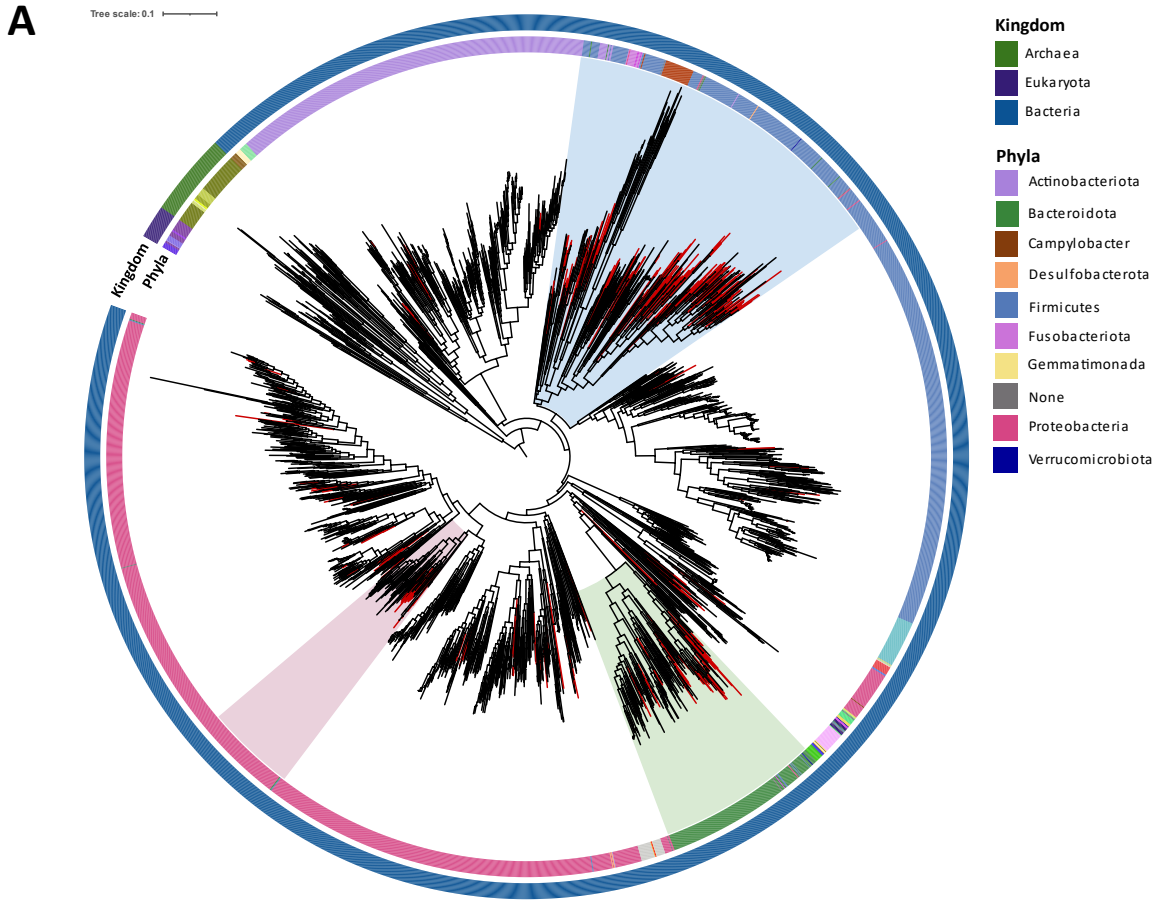
Results

Potential role of the gut microbiota in chitin digestion

Combining long-read and short-read assemblies, we were able to reconstruct 314 high-quality genome bins (156 and 158 from each dataset respectively; see Table S1F) and highlighted the usefulness of using the two methods (see supplementary results part 1). The following analyses were done on this set of high-quality selected bins to assess the presence of chitinolytic bacteria in the gut microbiota of myrmecophagous mammals.

All selected bins were taxonomically assigned to bacterial genomes and their phylogeny is presented in Figure S2. They belonged mainly to the Lachnospiraceae (n = 63), Burkholderiaceae (n = 24), Acutalibacteraceae (n = 18), Ruminococcaceae (n = 12) and Bacteroidaceae (n = 11) families (Table S1F). These 314 selected bins were integrated in a phylogeny of prokaryote reference genomes to confirm these results (Fig 1A). All selected bins were well placed in this phylogeny, suggesting they are similar to known bacterial genomes and not distantly related, even for genome bins for which taxonomic assignment failed. Some of the reconstructed bacterial genome bins clustered together within the Firmicutes (Fig 1B), Bacteroidetes, and Proteobacteria (Figs S1 A and B), in clades not including reference genomes, suggesting that these genome bins could be specific to myrmecophagous species (here called myrmecophagous-specific clades). Within the Firmicutes, six clades containing more than two bins reconstructed from myrmecophagous samples can be defined within the Lachnospiraceae family with two clades including respectively 12 and 11 genome bins (Fig 1B). In the first clade, six bins had no taxonomic assignment below the family level and could represent bacterial genomes not yet described, two belonged to the *Acetatifactor* genus, and four to the *CAG-510* genus. In the other clade, four bins were not taxonomically assigned below the family level, three bins belonged to the *CAG-590* genus, one was assigned to the *CAG-127* genus and two were assigned at the species level (*Frisingiococcus caecimuris* and *Acetivibrio ethanolgignens*) (Fig 1B). These bins were all reconstructed from xenarthran gut metagenomes (Fig 1B). Within the Acutalibacteraceae family, one clade contained 18 genome bins reconstructed mostly from xenarthran gut metagenomes and the three aardvark samples (Fig 1B). Eleven had no taxonomic assignment below the family level, four belonged

to the *Eubacterium* genus, and three to the *UBA1227*, *UBA1691*, and *UBA6857* genera (Fig 1B). Other clades containing only genome bins from myrmecophagous species were found within the Ruminococcaceae, Oscillospiraceae, Anaerovoracaceae, and Erysipelotrichaceae families.



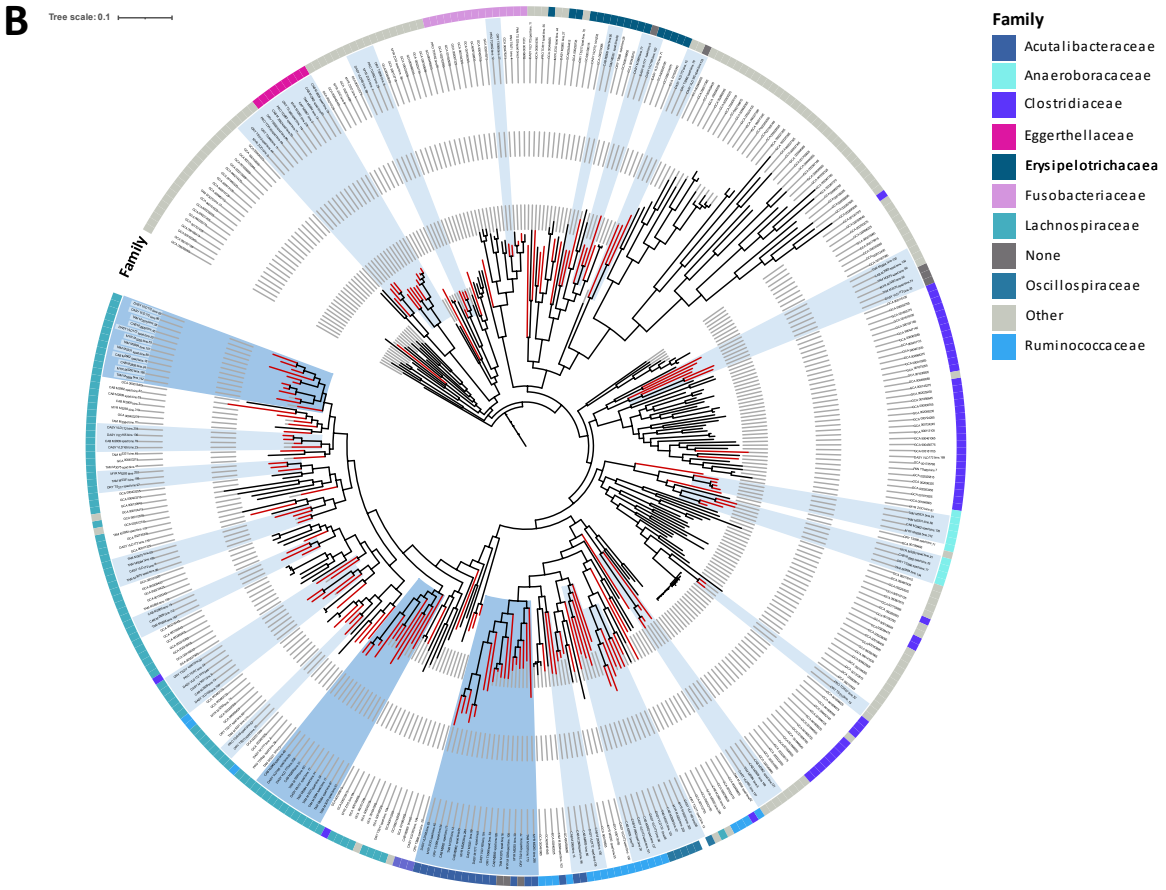


Figure. 1. Phylogenetic position of the 314 high-quality selected bins reconstructed from 29 gut metagenomes of the nine focal myrmecophagous species within a reference prokaryotic phylogeny. A: Phylogeny of the 314 selected bins (red branches) with 2496 prokaryote reference genomes. Circles respectively indicate (from inner to outer circles): the bacterial phyla and kingdom to which these genome bins were assigned based on the Genome Taxonomy Database (Chaumeil *et al*, 2020). Clades, where a subtree was defined, are highlighted in blue for the Firmicutes (Fig 1B), green for the Bacteroidetes, and pink for the Proteobacteria (Figs S1 A and B respectively). **B**: Subtree within Fimircutes showing myrmecophagous-specific clades (blue highlights; dark blue corresponds to the three clades mentioned in the results, light blue to the other clades). The outer circle indicates the bacterial family to which these genome bins were assigned based on the Genome Taxonomy Database.

To understand whether these bacteria could potentially degrade chitin, these high-quality genome bins were then used to search for GH18 enzymes, which include chitinases and chitin-binding enzymes. This resulted in 132 bins containing at least one GH18 sequence (between 0 and 17 bins per sample, 4.93 on average). Having complete genome bins enabled us to identify several GH18 genes in the same bin (between 0 and 17 per bin, 1.26 on average; Table S1F). In total, 394 GH18 sequences were identified (Fig 2) with 237 sequences presenting an

active chitinolytic site (DXXDXDXE) (Fig 2) and distributed among 83 bins (here called chitinolytic bins). These chitinolytic sequences are found in genome bins belonging mainly to the Lachnospiraceae (n = 183), Acutalibacteraceae (n = 76), and Ruminoccocaceae (n = 23) bacterial families (Fig 2). Fifty-three sequences not presenting an active site were found in a clade with sequences similar to Lysin Motif (LysM) domain-containing proteins (Fig 2), which is a 40 amino acid domain involved in peptidoglycan and chitin-binding (Buist *et al*, 2008). Sixty-two were placed in a clade with sequences similar to src Homology-3 (SH3) domain-containing proteins (Fig 2), which is a 50 amino acid domain found in intracellular and membrane proteins involved in the binding of ligands (Musacchio *et al*, 1994). These two clades were used to root the tree. Finally, 278 GH18 sequences formed a clade with sequences similar to known bacterial chitinases (Fig 2). The majority had an active chitinolytic site. These sequences were identified in bins reconstructed from different host species, representative of the five myrmecophagous placental orders. They represent a diversity of chitinase genes, as they are distributed in distinct clades (Fig 2).

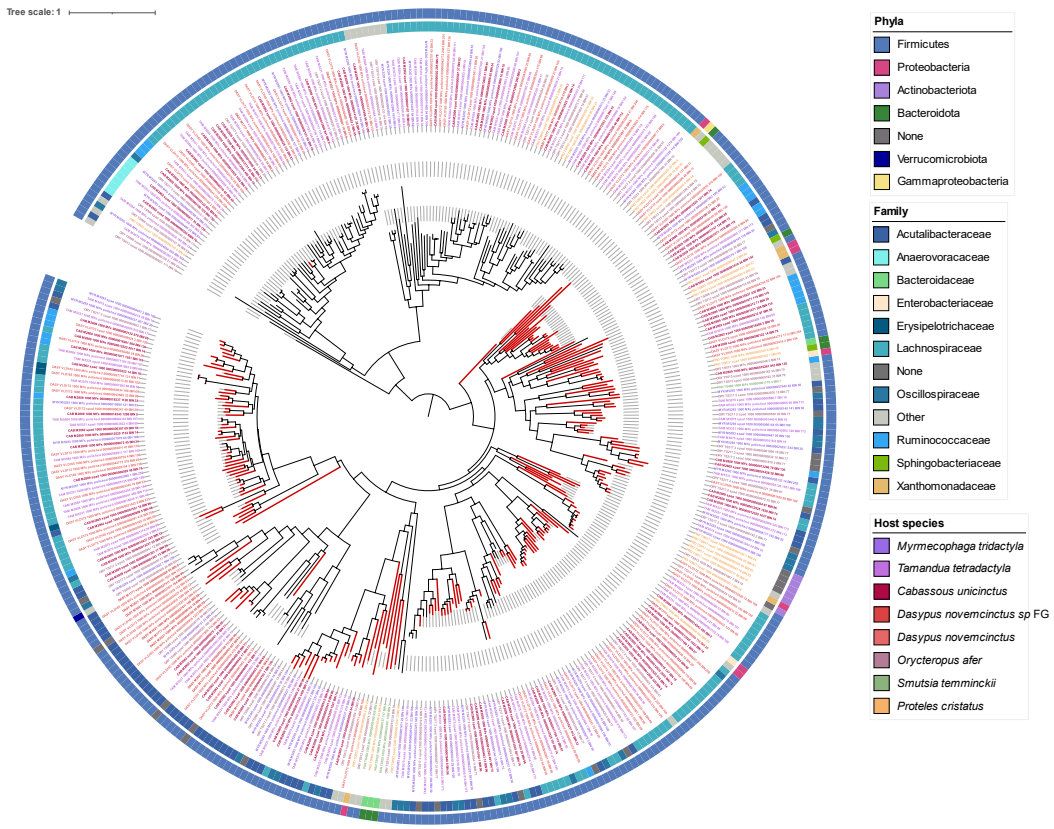


Figure. 2. Phylogeny of the 394 GH18 sequences identified in 132 high-quality selected bins reconstructed from 29 gut metagenomes of the nine focal myrmecophagous species and relatives. Red branches indicate the 237 sequences having an active chitinolytic site (DXXDXDXE). Circles respectively indicate (from inner to outer circles): the bacterial family and phyla of the bin the sequence was retrieved from. Colored sequence names indicate the

host species. Colored circles at certain nodes indicate enzymes to which sequences are similar when blasting them against the NCBI non-redundant protein database.

Distribution of chitinolytic bacteria among myrmecophagous mammals

Numbers of shared and specific selected bins in the nine host species were computed based on a detection threshold of 0.25 (the percentage of the reference covered by at least one read) of the selected bins across samples (Fig 3 and detection table available via Zenodo). According to this threshold, five selected bins were not considered to be detected in any sample. Between 6 and 124 selected bins were detected in each sample (52.62 on average) and 252 selected genomes were shared and detected in samples other than the ones they were reconstructed from (Fig 3). Selected bins were detected in 4.86 different samples on average (between 0 and 18). Selected bins were shared between individuals of the same species but with some intra-specific variability (Fig 3). Conversely, 57 bacterial genomes were detected in only one sample (Fig 3) and 194 bins were detected in more than one host species (between 0 and 7 host species; 2.73 on average), between closely related species or distantly related species (Fig 3). On the other hand, 115 selected bins were specific to a particular host species, meaning they were found in only one of the nine species studied here (Fig 3). Selected bins were detected in 1.80 different host orders on average (between 0 and 5) for a total of 179 bins detected in more than one host order. Finally, 130 bins were detected in only one host order.

There were more shared selected bins between host species carrying GH18 genes than shared bins with no GH18 gene (Fig S3). Among the 194 bins shared between host species, 110 had at least one GH18 gene. These bins belonged to 15 different bacterial families, mainly the Lachnospiraceae (n = 60) and Acutalibacteraceae (n = 10) (Fig S3). The 84 shared selected bins without chitinases belonged to a more diverse range of bacterial families (33 families), mainly Burkholderiaceae (n = 8) and Bacteroidaceae (n = 7) (Fig S3). On the contrary, among bins found in only one host species, there were fewer chitinase-carrying bins. Among the 115 host-species specific bins, 21 had at least one GH18 gene. They were found in 12 different bacterial families (Fig S3). The 94 specific bins with no chitinase genes belonged to 37 different bacterial families, mainly Burkholderiaceae (n = 16) (Fig S3). A similar pattern was observed at the host order level (see supplementary results part 2 available via Zenodo).

Three high-quality selected bins carrying GH18 were shared between species belonging to the five different myrmecophagous orders: one bin of *Bacteroides fragilis* (mean absolute abundance across samples 2.39, see abundance table available via Zenodo) and two bins of Enterobacteriaceae (mean absolute abundances across samples 0.58 and 3.66). Two of these three bins shared across orders were a bit more abundant, on average, than the mean absolute abundance of selected bins across samples of 0.89 (the minimal abundance of a bin across samples was 2.82e-6 and the maximal absolute abundance was 198). Seven selected bins carrying GH18 were shared among four myrmecophagous orders. Two bins belonged to *Enteroccocus faecalis* and were reconstructed from pangolin samples, along with

bins belonging to Lachnospiraceae bacteria (notably one from *Blautia* sp., one from *Faecalimonas* sp. and one from an unknown genus), one from *Bacteroides* sp. (Bacteroidaceae) and one from *Emergencia timonensis* (Anaerovoracaceae) with mean absolute abundances across samples ranging from 0.21 (*E. timonensis*) to 4.74 (*Bacteroides sp*), the latter being found to be abundant in several samples (for example: one *D. novemcinctus* (12.2), one *T. tetradactyla* (21.2), one *O. afer* (41.2)). One bin of *Enteroccus faecalis* also had a mean absolute abundance (3.43) above the mean absolute abundance of selected bins and was found to be abundant in several samples as well (one *D. sp. nov* FG (33.8), three pangolin samples (10.8, 13.3 and 21.8), and one *T. tetradactyla* (15.3)).

At least one chitinolytic bin (a high-quality selected bin carrying at least one GH18 sequence with an active chitinolytic site) was detected in the gut microbiota of each host species (Fig 4). Consistent with our previous results (Fig 2) these bins mainly belonged to the Firmicutes (Fig 4). More chitinolytic bins were detected in gut metagenomes of xenarthran species compared to the aardvark, the southern aardwolf, and the ground pangolin, except for *Dasypus kappleri* for which we only had one sample (Fig 4; Table S2). When compared to the total number of genome bins detected per species, xenarthran species had higher proportions of chitinolytic genome bins (Table S2).

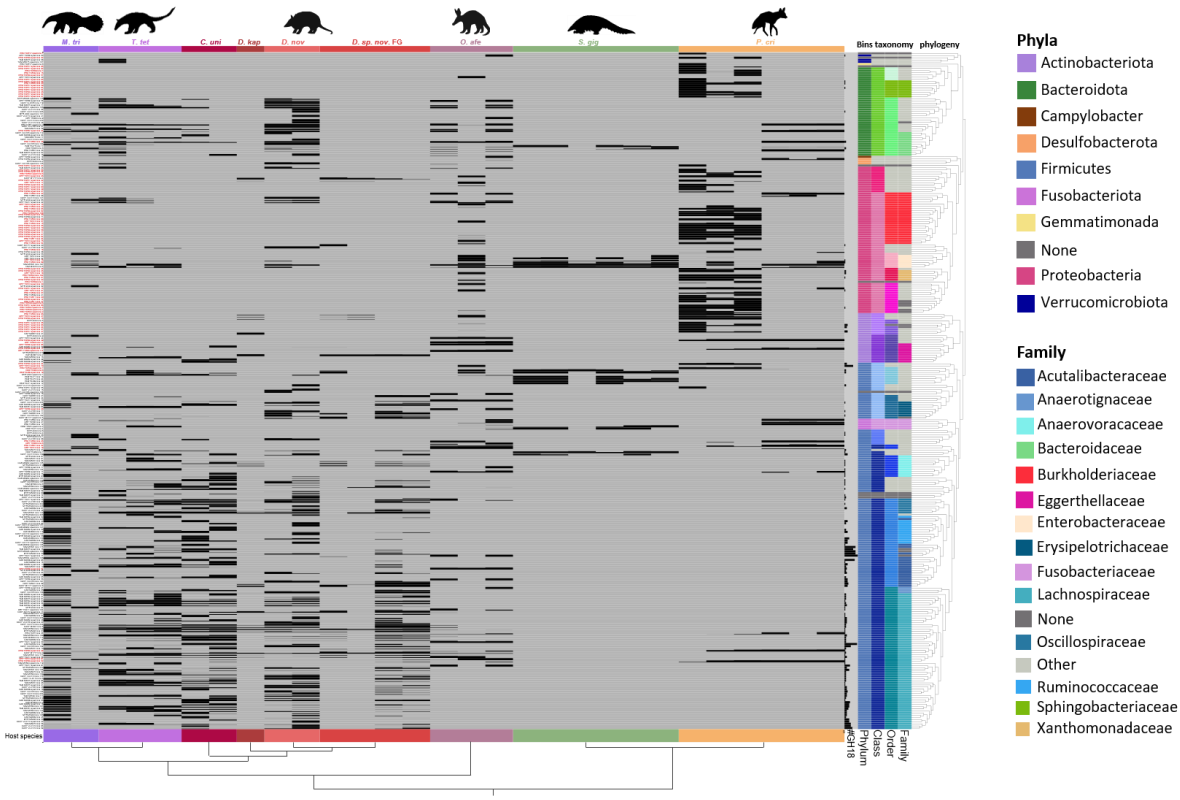


Figure 3. Detection of the 314 high-quality bacterial genomes (lines) **in the 29 gut metagenomes** (columns) **of the nine focal species.** Each square indicates the detection of a genome bin in a sample as estimated by anvi'o v7 (Eren *et al*, 2021). Names of bins are indicated on the left with red indicating bins detected in at least one soil sample (detection > 0.25) (Fig S4 and detection table available via Zenodo). Phylogenetic relationships of host

species distinguished by different color strips, are represented at the bottom of the graph. Columns on the right indicate (from left to right): the number of GH18 sequences identified in each bin (from 0 to 17), the bin's taxonomic phylum, class, order, and family. The phylogeny of the 314 selected bins inferred with PhyloPhlAn v3.0.58 (Asnicar *et al*, 2020) is also represented on the right of the graph (see Fig S2). Silhouettes downloaded from phylopic.org.

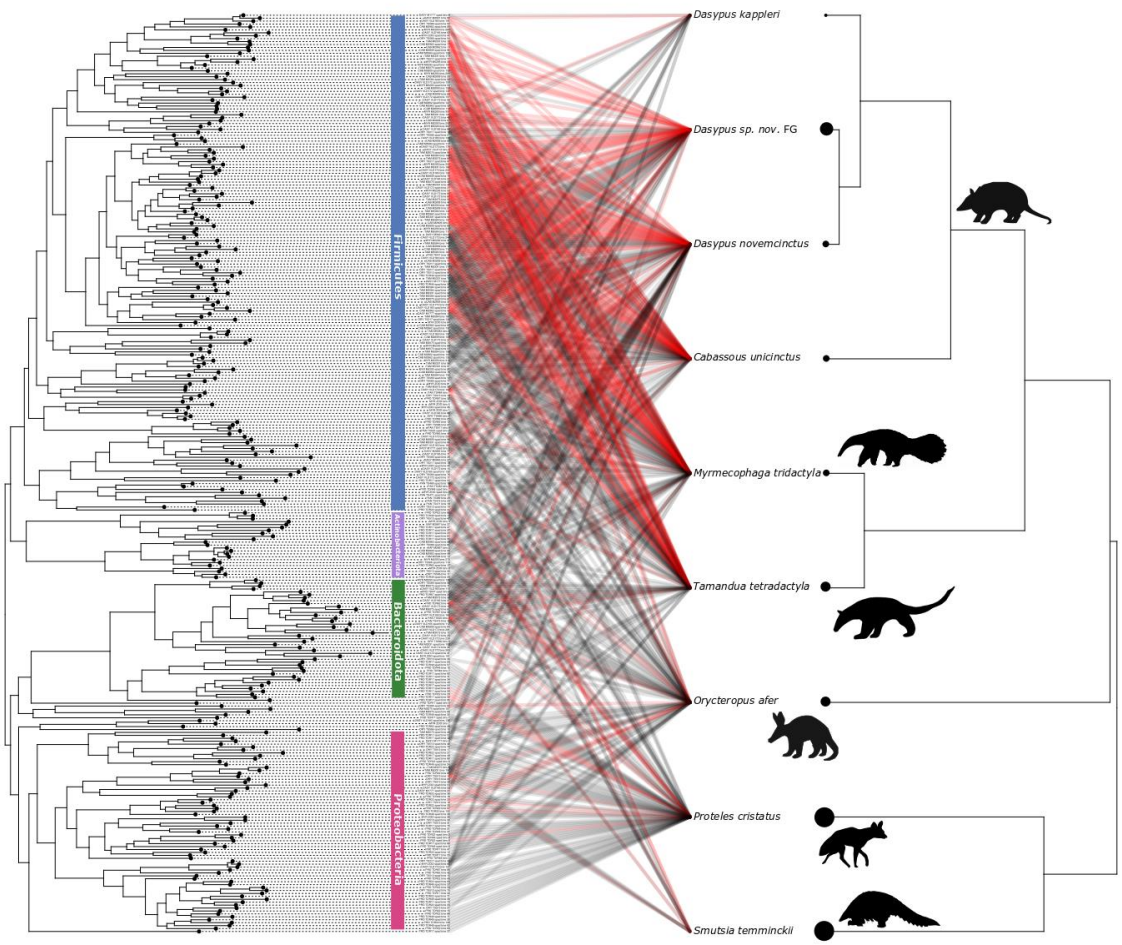


Figure 4. Distribution of chitinolytic selected bins (red links) **among the nine focal myrmecophagous species and relatives.** Phylogenies of the 314 high-quality selected bins (Fig S2) and of the nine host species (downloaded from timetree.org) are represented respectively on the left and the right of the graph. Links illustrate, for each bin, in which host species the bin was detected (detection threshold > 0.25). Red links indicate bins in which at least one GH18 sequence with an active chitinolytic site (DXXDXDXE) was found (chitinolytic bins). The size of the circles at the tips of the host phylogeny is proportional to the number of samples (n = 1 for *D. kap*; n = 2 for *D. nov*, *C. uni* and *M. tri*; n = 3 for *T. tet* and *O. afe*; n = 4 for *D. sp. nov* FG; n = 6 for *P. cri* and *S. tem*). Bins' names are indicated at the tip of the bins' phylogeny and main bacterial phyla are indicated by colored vertical bars. This graph was done with the cophylo R package within the phytools suite (Revell, 2012). Silhouettes downloaded from phylopic.org.

Discussion

Potential role of the gut microbiota in prey digestion in myrmecophagous mammals

The gut microbiota plays an important role in host digestion, which has led to several cases of gut microbiome convergence in distantly related species that share similar diets (Muegge *et al*, 2011; Delsuc *et al*, 2014; Suzuki *et al*, 2017). In animals with a chitin-rich diet, symbiotic chitinolytic bacteria could be involved in prey digestion, as they have been identified in species as diverse as nine-banded armadillos (Smith, 1996), insectivorous bats (Whitaker *et al*, 2004), insectivorous monkeys (McDonald *et al*, 2014), or crustacean-eating whales (Olsen *et al*, 1999). Here, we focused on reconstructing high-quality bacterial genome bins from gut metagenomes of species representative of the five myrmecophagous orders and identified sequences belonging to the main bacterial chitinases (i.e., GH18; Adrangi and Faramarzi, 2013; Shakir *et al*, 2016; Wardman *et al*, 2022). This allowed us to assess the potential of chitin degradation of gut symbionts in myrmecophagous placentals and better understand their convergent adaptation to this highly specialized diet.

Among the high-quality bacterial genome bins reconstructed, several clustered together in myrmecophagous-specific clades. These clades were found mainly within the Lachnospiraceae and Actualibacteraceae families within Firmicutes but also within the Oscillospiraceae and Ruminococcaceae. Several of these genomes had no taxonomic assignment below the family level, suggesting they could represent bacterial taxa not described yet and could be linked with the adaptation to myrmecophagy. Most of these bacterial taxa (i.e., Lachnospiraceae like *Blautia* or *Roseburia* genera or Ruminococcaceae) were previously found to be significantly more abundant in the gut microbiota of the aardwolf, giant anteater, and southern tamandua when compared to their non-myrmecophagous sister-species based on 16S rRNA metabarcoding data (Delsuc *et al*, 2014), which suggests that their presence may be associated with dietary adaptations. Besides, chitin-degrading metabolic pathways were identified in the gut metagenome of the Malayan pangolin (*Manis javanica*) as well as genes encoding chitinases and chitin-binding proteins belonging to the GH18 and GH19 enzyme families (Ma *et al*, 2018). This allowed the identification of chitinolytic symbionts such as *Enterococcus faecalis* (Enterococcaceae), *Clostridium paraputrificum* (Clostridiaceae), and *Bacteroides fragilis* (Bacteroidaceae) (Ma *et al*, 2018), whose genomes were also reconstructed here from Temminck's pangolin (*Smutsia temminckii*) gut metagenomes, confirming their probable role in prey digestion in pangolins.

Numerous GH18 genes were identified within the high-quality bacterial genomes we reconstructed here. We were able to identify 132 genome bins carrying at least one GH18 gene among which 83 had at least one GH18 sequence with an active chitinolytic site (DXXDXDXE; Olland *et al*, 2009; Hamid *et al*, 2013), here called chitinolytic bins, and representing bacteria that might thus play a role in insect prey digestion. Among genome bins found in the three myrmecophagous-specific clades mentioned previously, almost all had at least one GH18 sequence. This further suggests that these genome bins could represent bacterial taxa playing a significant role in chitin digestion in myrmecophagous species.

Studying the distribution of these bacteria in other insectivorous mammals will allow further understanding their role in the adaptation toward these specialized diets.

In addition, some of the bacterial taxa we identified here as being potentially involved in chitin digestion in myrmecophagous species belonged to or were closely related to bacterial taxa known for their chitinolytic properties. For example, *Clostridium* species are known for their chitinolytic activity (Pel *et al*, 1989; Simůnek *et al*, 2002; Sebaihia *et al*, 2007). Chitinases have also been identified and studied in *Enterococcus faecalis* (Leisner *et al*, 2009; Vaaje-Kolstad *et al,* 2012). More generally, Lachnospiraceae, for which we reconstructed the most genomes, are known to degrade complex polysaccharides (Vacca *et al*, 2020) and our results suggest they could be involved in chitin hydrolysis. Moreover, several genomes of Ruminococcaceae bacteria reconstructed here encode a GH18. Within this family, *Ruminococcus* species have been suggested to play a role in chitin digestion in the gut microbiota of insectivorous mammals (La Reau *et al*, 2016). Several GH18 sequences were found in the same bacterial genomes, suggesting that complex chitin-degrading pathways are present in the gut microbiota of the myrmecophagous species, consistent with previous analyses of the chitinolytic properties of certain bacteria. Indeed, several enzymes are involved in the different steps of chitin hydrolysis (Beier and Bertlisson, 2013). For example, two types of chitin-binding and chitinase enzymes have been described in *E. faecalis* (Vaaje-Kolstad *et al*, 2012). This bacteria also carries endo-β-N-acetylglucosaminidases (Bøhle *et al*, 2011) which are part of the chitin hydrolysis pathway and degrade chitin into chitosan (Beier and Bertlisson, 2013). This further highlights the complex chitinolytic machinery of these bacteria.

Finally, some sequences do not have the ability to hydrolyze chitin because no active site was identified, but could still bind chitin. For example, the Lysin-motif (LysM) domain of certain proteins can bind peptidoglycan (Buist *et al*, 2008). It is present in certain eukaryotes chitinases (i.e., algae, nematodes) and participates in the recognition of symbiotic rhizobial bacteria by leguminous plants and is thus thought to bind chitin (Spaink, 2004; Buist *et al*, 2008). The SH3 domain is also known to have binding properties and is involved in protein interactions (Musacchio *et al*, 1194). Therefore, sequences similar to LysM and SH3 domain-containing proteins could still be part of the chitin hydrolysis process by influencing molecular interactions. In addition, enzymes carrying the carbohydrate-binding module 37 (CBM37), which is known to have binding properties notably to chitin, could also be involved in the chitin degrading process. The presence of other enzymes potentially involved in insect digestion could be investigated, such as trehalases, enzymes that break down trehalose, a sugar found in insect blood. The gut microbiota of the Malayan pangolin and giant anteater were found to be significantly enriched in these enzymes when compared to non-myrmecophagous species (Cheng *et al*, 2022). The gut microbiota also plays a role in the detoxification of ingested compounds (Stevens and Hume, 1998; Suzuki, 2017), such as insect toxins. For example, metabolic pathway analyses have revealed the potential of microbial symbionts to detoxify formic acid in some myrmecophagous species (Cheng *et al*, 2022).

Shared and species-specific gut bacteria among myrmecophagous mammals

Morphological and genomic comparative studies have shown that the convergent adaptation to myrmecophagy in mammals involved different mechanisms between the different species and that phylogenetic constraints played a central role (Emerling *et al*, 2018; Ferreira-Cardoso *et al*, 2019, 2020, 2022; Allio *et al*, 2023). As diet is one of the main factors shaping the evolution of the gut microbiota (e.g., Muegge *et al*, 2011), this raises the question of whether the same bacterial symbionts with similar functions were convergently recruited between the different ant-eating species, or whether it was different bacteria with similar functions (functional convergence).

Analysis of the distribution of selected genome bins across samples revealed both shared and host-specific bacteria among myrmecophagous species. Selected genomes were mainly shared between closely related species and mainly among Xenarthra but less so between distantly related orders such as pangolins (Pholidota) and anteaters (Pilosa), highlighting the influence of host evolutionary history. In mammals, phylosymbiosis is particularly strong (Song *et al*, 2020; Perez-Lamarque *et al*, 2023) in part because of mammalian-specific traits such as viviparity or parental care (Mallot and Amato, 2021) but also limited microbial dispersal abilities (Mazel *et al*, 2023) facilitating the vertical transmission of microbes, which may explain these results. Some bacterial genomes were shared between distantly related host species and those shared across more than one order mainly belong to the family Lachnospiraceae. One selected genome of *Bacteroides fragilis* and two of Enterobacteriaceae were shared between species belonging to the five different myrmecophagous orders and carry GH18. Similarly, seven selected genome bins were shared among four host orders and belong to *Enteroccocus faecalis*, the Lachnospiraceae (i.e., *Blautia* sp., *Faecalimonas* sp.), *Bacteroides* sp. (Bacteroidaceae) and *Emergencia timonensis* (Anaerovoracaceae). These shared high-quality selected genome bins carrying GH18 may thus participate in the adaptation to myrmecophagy and highlight the influence of the host diet, consistent with previous studies showing that it is an important factor in shaping the mammalian gut microbiota (e.g., Ley *et al*, 2008a, 2008b; Muegge *et al*, 2011; Youngblut *et al*, 2019). Each host species appears to carry chitinolytic bacteria in its gut microbiota, confirming their potential role in adapting to an insect-based diet. However, differences in the distribution of these chitinolytic selected bins could highlight divergent microbiota adaptations, especially in xenarthran species, which carry more chitinolytic bacteria than aardvark, ground pangolin, and southern aardwolf. Some bins were found to be abundant in some samples raising questions on how the gut microbiota is involved in prey digestion. Having few highly abundant chitinolytic bacterial species producing large amounts of chitinases might be sufficient for a specific host to digest its prey, whereas having diverse chitinolytic bacterial species might have been selected for in other host species.

The distribution of high-quality selected genome bins across samples and species revealed how shared and host-specific bacteria may reflect the divergent evolutionary histories of host species and the effect of adapting to a similar diet despite phylogenetic constraints. This relates to the evolution of the mammalian gut microbiota. From a

mammalian ancestor that probably had an insectivore-like gut microbiota (Perez-Lamarque *et al*, 2023), its composition might have changed as placentals diversified and occupied new niches, constraining its composition in certain species. Chitinolytic bacteria shared by all myrmecophagous orders may represent ancient bacterial lineages inherited from a common ancestor, while bacteria found in only certain host species may reflect more recent adaptations with bacteria acquired in specific lineages.

The environment and biogeography also influence host-associated microbial communities, which could explain the distribution pattern of selected bins across samples, with large differences observed between South Africa and South America. Anteaters and armadillos (Xernarthra) diverged anciently from aardvarks, pangolins, and aardwolves (~80 Ma; Álvarez-Carretero *et al*, 2021) and these lineages have evolved in very distinct environments and biogeographic contexts for most of their evolutionary history. Soil samples from the fecal sampling sites have been collected for the South African samples. Some selected bins reconstructed from gut metagenomes of the ground pangolin, southern aardwolf, and aardvark belonged to bacterial taxa that were indeed not expected to be found in the gut but rather in the environment (e.g., Sphingobacteriaceae, Xanthomonadaceae, Burkholdericaeae). The distribution of selected genome bins reconstructed from gut metagenomes of these species was examined in soil samples and showed that 91 selected bins were also detected in at least one soil sample (detection > 0.25; Fig S4). For example, among the Proteobacteria selected genome bins found in myrmecophagous-specific clades (Fig S1 B), some belonged to Burkholderiaceae bacteria and were detected in at least one soil sample. This may not necessarily reflect environmental contamination, as the center of the feces was specifically sampled to minimize contamination. Rather, it may reflect the fact that these species ingest soil and thus environmental microbes while foraging (Redford, 1987). This could be beneficial to the host (Gull *et al*, 2015), as it may compensate for the lack of mastication and/or help to deal with toxins that might be present in prey (Redford, 1987). Environmental acquisition of microbes has also been suggested in lemurs (Bornbusch *et al*, 2022) and wild echidnas (Perry *et al*, 2022), species that ingest soil during foraging and in which soil bacteria have been identified in their gut microbiota. These environmental bacteria could therefore represent transient bacteria in the gut, reflecting what is ingested by the host. They could also represent resident bacteria that have been recruited from the environment into the gut microbiota of myrmecophagous species. Indeed, some beneficial microorganisms could be acquired from soil microbiomes (Banerjee and van der Heijden, 2023), which could allow the host to better adapt to its environment through horizontal gene transfer of beneficial genes (Hehemann *et al*, 2010). Moreover, it has also been suggested that environmental bacteria could be a source of chitinase genes to digest prey (Gooday, 1990). Thus, their recruitment into the gut microbiota may have been selected to participate in the host digestion and further contribute to the adaptation to its environment. This would highlight the potential influence of the environment on the gut microbiota of myrmecophagous mammals.

Role of the holobiont in prey digestion in myrmecophagous mammals

The influence of the microbiota on host evolution and their coevolution is increasingly recognized as studies of the diverse microbiota of captive and wild animals proliferate. Many scientists now recognize the term holobiont to describe the host and all its associated symbionts, which would be the unit of natural selection as defined by the hologenome theory of evolution (Zilber-Rosenberg and Rosenberg, 2008). Integrative studies (e.g., Zepeda Mendoza *et al*, 2018; Cheng *et al*, 2022) and initiatives such as the "Earth Hologenome Initiative'' (http://www.earthhologenome.org/) to study hologenomic adaptations are now becoming more common. In the specific case of myrmecophagous mammals, a hologenomic approach would help to better understand the adaptive mechanisms involved. Indeed, if chitinolytic bacteria can digest chitin, prey digestion can also be ensured by host-produced chitinases. In mammals, from a placental ancestor that probably carried five functional chitinase paralogs (*CHIA*), some of these paralogs were subsequently lost during placental diversification in non-insectivorous species, leading to a positive correlation between the number of functional chitinase paralogs and the proportion of invertebrates in the diet (Emerling *et al*, 2018); a correlation also observed in primates (Janiak *et al*, 2018). Among myrmecophagous species, this chitinase gene repertoire has evolved differently to ensure chitin digestion, reflecting phylogenetic constraints (Emerling *et al*, 2018; Allio *et al*, 2023). Thus, both endogenous and microbial chitinases may be involved in prey digestion, raising the question of the relative contribution of the host and its symbionts in providing the same function. For example, myrmecophagous mammals with only one functional *CHIA* paralog (*CHIA5*), such as aardwolves and pangolins, may compensate for this by overexpressing their only functional paralog or by relying more on their chitinolytic symbionts to digest their prey. In pangolins, *CHIA5* has been found to be overexpressed in all digestive organs (Allio *et al*, 2023). In the aardwolf, which recently diverged from the other Hyaenidae species (~10 Ma; Eizirik *et al*, 2010), and does not present strong morphological adaptations to myrmecophagy, no host chitinases have been found to be expressed in its salivary glands (Allio *et al*, 2023) and expression of these enzymes in other digestive organs needs further investigation. Therefore, this species could rely more on its gut microbiota to ensure prey digestion. Based on our results, fewer chitinolytic bacteria were found in the gut microbiota of the southern aardwolf compared to other species but they could still ensure chitin digestion, for instance, by overexpressing their chitinases. The combination of the host genomic and transcriptomic data with metagenomic and metatranscriptomic data of its microbiota would allow answering this type of question by comparing host gene repertoires in light of the chitinolytic abilities of the gut microbiota. This will shed light on the adaptation to chitin digestion in placentals and, more generally, on the role of the holobiont in the adaptation to a specific function.

Materials and methods

1. Sampling

Thirty-three fecal samples were collected from nine species representative of the five myrmecophagous placental orders (Table 1). For armadillos and anteaters provided by the JAGUARS collection (Cayenne, French Guiana), fecal samples from roadkill and deceased zoo animals were obtained after unfreezing the specimens and dissecting the lower part of the digestive tract in the lab facilities provided by Institut Pasteur de la Guyane (Cayenne, French Guiana). Roadkill armadillos collected in the USA were also dissected to sample feces. For the aardvark, ground pangolin, and southern aardwolf, fresh fecal samples were collected directly in the field during fieldwork sessions conducted in Tswalu Kalahari and Tussen-die-Riviere reserves (South Africa). The inner part of the feces was sampled with a sterile scalpel blade to avoid soil contamination. In the South African reserves, eight soil samples were also collected near feces sampling sites to serve as a control for potential environmental contamination (Table S3). All fecal and soil samples were stored at -20°C in 96% ethanol before DNA extraction.

Table 1. Detailed sample information for the 33 fecal samples collected.

Sample name	Species name	Common name	Order	Diet	Sex	Age	Wild / Captive	Country of origin	Location	Accession number	
										Illumina	ONT
CAB M2809	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925023	SRR23925022
CAB M2962	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925011	SRR23925000
CAB M3141	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924989	SRR23924978
DASY M1748	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	NA	SRR23924967
DASY M1777	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Juvenile	Wild	French Guiana	coll. JAGUARS	SRR23924962	SRR23924961
DASY M2256	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	NA	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924960	SRR23925021
DASY M2886	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925020	SRR23925019
DASY M3021	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925018	SRR23925017
DASY VLD165	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Female	Adult	Wild	USA	Valdosta (GA)	SRR23925016	SRR23925015
DASY VLD168	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Male	Adult	Wild	USA	Valdosta (GA)	NA	SRR23925014
DASY VLD172	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Female	Adult	Wild	USA	Valdosta (GA)	SRR23925013	SRR23925012
KAP M2867	*Dasypus kappleri*	Greater long-nosed armadillo	Cingulata	Insectivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925010	SRR23925009
MYR M6293	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925008	SRR23925007
MYR M6295	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	NA	Juvenile	Wild	French Guiana	coll. JAGUARS	SRR23925006	SRR23925005
MYR ZOO	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	Female	Adulte	Captive	France	Montpellier zoo	SRR23925004	SRR23925003
ORY T8217	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23925002	SRR23925001
ORY T8513	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924999	SRR23924998
ORY T8688	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924997	SRR23924996
PAN T8471	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924995	SRR23924994
PAN T8476	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924993	SRR23924992
PAN T8482	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924991	SRR23924990
PAN T8488	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924988	SRR23924987
PAN T8489	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924986	SRR23924985
PAN T8626	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924984	SRR23924983
PRO TDR7	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924982	SRR23924981
PRO TDR17	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924980	SRR23924979
PRO TDR22	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924977	SRR23924976
PRO TDR49	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924975	SRR23924974
PRO TDR62	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924973	SRR23924972
PRO TDR67	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924971	SRR23924970
TAM M3075	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924969	SRR23924968
TAM M6331	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924966	SRR23924965
TAM M6584	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924964	SRR23924963

N.B.: Diet was determined based on field observations (i.e, dissections) and the literature (Redofrd, 1985, 1986, 1987; Anacleto, 2007; Vaz *et al*, 2012; Aya-Cuero *et al*, 2019)

2. DNA extraction

Whole DNA was extracted from fecal samples following the optimized protocol of Magdeleine *et al* (2021) using the Genomic DNA from soil kit (NucleoSpin, Macherey-Nagel). Two successive extractions were done and a purification step was added to retrieve high molecular weight DNA suitable for long-read sequencing (Magdeleine *et al*, 2021). The same kit was

used to extract DNA from soil samples. Before the extraction, samples were incubated with 700 µL of the lysis buffer (SL1; Magdeleine *et al*, 2021) and 30 µl of proteinase K at 56°C for30 minutes.

3. Library preparation and DNA sequencing

3.1. Long-read sequencing

Long-read libraries were constructed using the SKQ-LSK109 and SKQ-LSK110 library preparation kits (ONT, Oxford Nanopore Technologies, https://nanoporetech.com/). Shotgun metagenomic sequencing was done using the MinION, MK1C, and GridION devices using one R9 flowcell per sample. Between 150 and 988 ng of DNA were loaded per flowcell, which were run for 48 to 72 hours (Table S1A). For samples sequenced on the GridION, super-accurate basecalling was performed with Guppy v5+ (Qscore = 10) whereas on the MinION and MK1C fast basecalling was used (Qscore = 7). Sequencing output statistics were checked using PycoQC v2.5.2 (Leger and Leonardi, 2019) with a minimum quality score set to 7 or 10 depending on the sequencing device used (Table S1A). Rebasecalling of samples sequenced on the MinION and MK1C was done on a GPU machine with Guppy v5.0.16 (super accurate mode, Qscore = 10, config_file = dna_r9.4.1_450bps_sup.cfg) (ONT).

3.2. Short-read sequencing

Short-read shotgun metagenomic Illumina sequencing was done for all samples to generate data for long-read assemblies polishing. Either the first or second extraction was used for sequencing as they have previously been shown to be both suitable for short-read sequencing (Magdeleine *et al*, 2021). Library preparation and Illumina sequencing on a NovaSeq instrument were outsourced to Novogene Europe (Cambridge, UK) to generate metagenomes using 50 million of 150 bp paired-end reads each (15 Gb of raw data per sample) (Table S1A). Two samples of two *Dasypus* species (DASY M1746 and DASY VLD168) could not be sequenced using short reads due to low quantity of starting material and were not included in the following analyses. Short-read metagenomic sequencing of soil samples was performed following the same protocol.

4. Data filtering

4.1. Long-read data

Sequencing adapters were removed with Porechop v0.2.4 (Wick *et al*, 2017) used with default parameters. Reads shorter than 200 bp were removed using Filtlong v0.2.1 (Wick, 2021). No quality filtering was performed at this stage as it was previously done during basecalling (Qscore > 10).

Host mitogenomes were downloaded for the nine myrmecophagous species of our dataset from the National Center for Biotechnology Information (NCBI) Genbank database (Table S1A). Long-read metagenomes were mapped to the host mitogenome with Minimap2 v2.17 (Li, 2018, 2021) with the ONT preset (*-ax map-ont*) for Oxford Nanopore reads. This allowed us to confirm the host species. Host mitogenomes were then assembled from the

mapping reads with Flye v2.8.3 (Kolmogorov *et al*, 2019) with default parameters for ONT raw reads (*--nano-raw*). Mapping of host mitogenomic reads on host mitogenomes was visually checked with Geneious Prime 2022.0.2 (Kearse *et al*, 2012), and mitochondrial reads were removed from the metagenomes for downstream analyses.

The same was done to remove host nuclear reads by mapping long-read metagenomes against the host genome using Minimap2 v2.17 (Li, 2018, 2021) using the Nanopore read option (*-ax map-ont*) (Table S1A). The *Dasypus novemcinctus* RefSeq genome assembly (Dasnov 3.0; GCF_000208655.1) was downloaded from Genbank. The *Orycteropus afer* genome was downloaded from the DNA Zoo database (Dudchenko *et al*, 2017; HiC assembly based on the draft assembly of Di Palma *et al*, unpublished). *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, *Smutsia gigantea,* and *Proteles cristatus* genomes have been previously reconstructed using long (ONT) and short (Illumina) reads assembled with MaSuRCA v3.2.9 (Zimin *et al*, 2017; see Allio, 2021 for a detailed description of the hybrid assembly process). For *Cabassous unicinctus*, we used the Discovar draft genome assembly generated by Emerling *et al* (2022). For species lacking an available reference genome, the genome of the closest relative in our dataset was used: *Smutsia gigantea* for *S. temminckii* and *Dasypus novemcinctus* for *D. kappleri* and *Dasypus sp. nov.* FG. Host reads were removed from the metagenomes for downstream analyses. One southern naked-tailed armadillo sample (CAB M3141) and one giant anteater sample (MYR M5295) presented a high proportion of host reads (> 95%) and were excluded from the following analyses. The final dataset thus included 29 samples.

Finally, contaminant human reads were removed following the same approach and using the telomere-to-telomere human genome assembly (GCA009914755.3) as reference (Table S1A).

4.2. Short-read data

Illumina sequencing adapters were removed using FASTP v0.20.0 with default quality filtering parameters (Chen *et al*, 2018). Host and human reads were removed using the same approach as for long-read metagenomes but the mapping was done with bowtie2 v2.3.5 (Langmead and Salzberg, 2012) with default parameters (Table S1A).

Samtools v1.7 (Li *et al*, 2009) was used to manipulate long- and short-read mapping files.

5. Metagenome assembly

Long-read metagenomes were assembled for each sample using metaFlye v2.9 (Kolmogorov *et al*, 2020) with default parameters for ONT reads basecalled with Guppy v5+ (*--nano-hq*, recommended mode when < 5% of sequencing errors is expected) and the strain mode (*--keep-haplotypes*), which prevents closely related strains represented by different paths in the assembly graph to be collapsed. Long-read metagenome assemblies were then polished with short-reads using Pilon v1.24 (Walker *et al*, 2014) with default parameters. Short-read metagenomes were assembled using metaSPAdes v3.11.0 (Nurk *et al*, 2017) and MEGAHIT v1.1.2 (Li *et al*, 2015) with default parameters.

Assembly statistics were computed with anvi'o v7 (Eren *et al*, 2021) (Tables S1B and C). Anvi'o first generates a contig database in which k-mer frequencies for each contig are computed (k = 4). Prodigal v2.6.3 (Hyatt *et al*, 2010) is then used to predict archaeal and bacterial open reading frames and estimate the number of genes. Finally, microbial single-copy core genes are searched using HMMER (Eddy, 2011) with the hmmscan v3.2.1 program (default in anvi'o).

6. Reconstruction of bacterial genomes

6.1. Genome binning and bins selection

Prior to binning, reads were mapped to metagenome assemblies to obtain the coverage information needed for binning. This was done using Bowtie2 v2.2.9 (Langmead and Salzberg, 2012) for short reads and Minimap2 v2.17 (Li, 2018, 2021) for long reads. Genome binning was conducted on single long-read polished assemblies and short-read assemblies separately (trimmed to keep contigs longer than 1000 bp) using metaBAT2 v2.15 (Kang *et al*, 2019) with default parameters (minimum size of contigs set to 2500 bp and bin minimal length set to 200 kb).

Completeness and redundancy of bins were estimated using anvi'o v7 (Tables S1D and E), which relies on the detection of single-copy core genes (SCG). Bins with more than 90% completeness and less than 5% redundancy were selected for downstream analyses. Based on these criteria, 239 genome bins were selected from the long-read assemblies and 254 from the short-read assemblies. Genome bins statistics were computed using anvi'o v7 as done for the metagenome assemblies (Table S11 and E).

6.2. Genome bins dereplication

To remove potentially redundant bins, a dereplication step was performed on the set of selected bins using dRep v3.3.0 (Olm *et al*, 2017). dRep first filters genomes based on their length (> 50 000 by default) and their completeness and redundancy as computed with CheckM v1.1.11 (Parks *et al*, 2015) (> 75% completeness, < 25% redundancy by default). dRep then clusters genomes based on an Average Nucleotide Identity (ANI) threshold of 90% using the Mash algorithm v2.3 (Ondov *et al*, 2016). Genomes having at least 90% ANI are then clustered using fastANI v1.33 (Jain *et al*, 2018), here with a secondary ANI threshold of 98% (recommended to avoid mis-mapping of metagenomic reads against bins, for example, to study their distribution across metagenomes). By default, at least 10% of the genome is compared (minimum alignment fraction). Two bins did not pass dRep quality filtering in each dataset and were not included in the analysis. In the long-read dataset, 38 genomes were removed by dRep with an ANI of 98%, resulting in 201 unique genome bins. In the short-read dataset, 48 genomes were removed, resulting in 206 unique genome bins. Selected genome bin statistics were compared between bins reconstructed from long-read and short-read assemblies using anvi'o v7 (see supplementary results part 1 available via Zenodo).

To determine whether similar genome bins were reconstructed using long-read polished assemblies (n = 201) or short-read assemblies (n = 206), dRep was also run on the

set of all selected bins (n = 407) with 98% ANI to remove redundant genome bins (93 genomes removed, resulting in 314 non-redundant high-quality bins). To capture the majority of the bacterial taxa present in our samples, we decided to combine high-quality selected bins reconstructed from both long- and short-read assemblies dereplicated with an ANI of 98% in our final dataset for downstream analyses (n = 314; Table S1F) (see supplementary results part 1 and dRep output results available via Zenodo).

6.3. Taxonomic assignment of selected bins

The taxonomy of selected genome bins reconstructed from long- and short-read assemblies was assessed using anvi'o v7 which uses 22 single-copy core genes (SCGs) and the taxonomy of the genomes defined by the Genome Taxonomy Database (GTDB; Chaumeil *et al*, 2020) from which these genes have been extracted (Table S1F).

The 314 selected bins were placed in a phylogeny of 2566 reference prokaryotic genomes downloaded from Genbank using PhyloPhlAn v3.0.58. PhyloPhlAn first searches universal prokaryotic marker genes (option *-d* phylophlan, a database comprising 400 markers; Segata *et al*, 2013), here using Diamond v2.0.6 (Buchfink *et al*, 2015). Marker genes were then aligned using MAFFT v7.475 (Katoh and Standley, 2013), and cleaned using trimAl v1.4 (Capella-Gutierrez *et al*, 2009). Next, marker gene alignments were concatenated. Finally, the phylogeny was inferred using IQ-TREE v2.0.3 (Nguyen *et al*, 2015) under the LG model. PhyloPhlAn was run with the *–diversity* high and *–fast* options, which together set a range of parameters for the reconstruction of high-ranked taxonomic level phylogenies (Asnicar *et al*, 2020). By default, at least 100 markers of the PhyloPhlAn database must be present in a genome for it to be included in the analysis (*--min_num_markers* 100) and each marker should be found in at least four genomes to be included in the analysis (*--min_num_entries* 4). Seventy of the reference genomes had less than 100 markers and were excluded from the analysis resulting in a total of 2810 genomes in the final reconstructed tree. Using the same PhyloPhlAn parameters, a phylogeny of the set of the 314 selected bins was also inferred (Fig S2).

6.4. Distribution of selected bins in gut metagenomes

Short-read metagenomes were mapped against the 314 selected bins using bowtie2 v2.3.5 with local sensitive alignment parameters. The distribution of bins across samples was explored using anvi'o v7. Anvi'o enables us to compute the percentage of mapped reads on each bin for each metagenome and computes statistics such as the coverage, percentage of recruited reads, abundance, and detection (horizontal coverage) of bins across metagenomes (samples). The distribution across samples was studied using the anvi'o interface by visualizing the detection of selected genome bins across samples. A detection threshold of 0.25 % of the reference covered by at least one read was chosen to consider a bin detected in a sample. The number of bins shared among the different samples and host species and bins specific to a sample or host species were calculated. Applying the same approach, the distribution of selected bins reconstructed from the aardvark, ground pangolin, and southern

aardwolf samples (n = 140) was explored in the eight soil samples collected in South Africa near feces sampling sites (Fig S4).

7. Identification of bacterial chitinase genes

Chitinase genes were searched in selected genome bins. More specifically, sequences of enzymes belonging to the Glycoside Hydrolase 18 (GH18) family (comprising chitinases and chitin-binding proteins) as determined by the classification of Carbohydrate Active enzymes (CAZymes) (Drula *et al*, 2022) were scanned using dbCAN2 (Zhang *et al*, 2018) with default parameters. dbCAN2 is designed to annotate CAZymes sequences in genomes using an HMMER search (Eddy, 2011) against the dbCAN HMM database containing HMM models for each CAZyme family (Buchfink *et al*, 2015).

The amino acid sequences of 420 identified GH18 genes were imported in Geneious Prime 2022.0.2. Sequences were aligned using MAFFT v7.450 with default parameters. The alignment was cleaned by removing sites not present in at least 50% of the sequences, resulting in an alignment of 390 amino acids. The chitinase gene tree was inferred with RAxML v8.2.11 (Stamatakis, 2014) under the LG+G model and the rapid hill-climbing algorithm for topology rearrangements. After removing 26 sequences that did not align correctly and had very long branches, the final alignment included 394 sequences. The conserved chitinolytic site (DXXDXDXE; Olland *et al*, 2009; Hamid *et al*, 2013), typical of chitin-degrading enzymes, was searched in the aligned sequences. BLAST searches against the NCBI non-redundant protein database were conducted to assess whether identified GH18 genes were similar to known microbial chitinases.

8. Data availability

Raw RNAseq short Illumina and long ONT reads have been submitted to the Short Read Archive (SRA) of the National Center for Biotechnology Information (NCBI) and are available under BioProject number PRJNA942254. Metagenome assemblies, genome bins (all and selected), phylogenetic datasets (alignments and corresponding trees), and other supplementary materials and results are available from zenodo.org (DOI: 10.5281/zenodo.7808597).

Acknowledgments

We are indebted to Lionel Hautier, Rémi Allio, Fabien Condamine, Sérgio Ferreira-Cardoso, Pierre-Henri Fabre, Quentin Martinez, Mathilde Barthe, Aude Caizergues, Nathalie Delsuc, Jessica Briner, Tshediso Putsane, Baptiste Chenet, Edith Guilloton, Solène Lefort, Margo Traimond, Michel Blanc, Bertrand Goguillon, Antoine Baglan for their help with sampling. Fieldwork sessions at Tswalu Kalahari Reserve (South Africa) were conducted under the auspices of the Tswalu Foundation thanks to Duncan MacFadyen (Oppenheimer Generations

Research and Conservation), Dylan Smith, and Gus van Dyk. Computational analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB) platform. This work has been supported by grants from the European Research Council (ConvergeAnt project: ERC-2015-CoG-683257) and Investissements d'Avenir of the Agence Nationale de la Recherche (CEBA: ANR-10-LABX-25-01; CEMEB: ANR-10-LABX-0004). The JAGUARS collection, managed by Kwata NGO is supported by the Collectivité Territoriale de Guyane and the Direction Générale des Territoires et de la Mer de Guyane. This is contribution ISEM 2023-XXX-SUD of the Institut des Sciences de l'Evolution de Montpellier.

References

Adrangi, S., and Faramarzi, M.A. (2013). From bacteria to human: A journey into the world of chitinases. Biotechnology Advances 31, 1786–1795.

Alessandri, G., Rizzo, S.M., Ossiprandi, M.C., van Sinderen, D., and Ventura, M. (2021). Creating an atlas to visualize the biodiversity of the mammalian gut microbiota. Current Opinion in Biotechnology 73, 28–33.

Allio, R. (2021). Phylogenomics and comparative genomics in ant-eating mammals. (Doctoral dissertation, Université de Montpellier).

Allio, R., Teullet, S., Lutgen, D., Magdeleine, A., Koual, R., Tilak, M.-K., Thoisy, B. de, Emerling, C.A., Lefébure, T., and Delsuc, F. (2023). Comparative transcriptomics reveals divergent paths of chitinase evolution underlying dietary convergence in ant-eating mammals. bioRxiv. 2022.11.29.518312.

Álvarez-Carretero, S., Tamuri, A.U., Battini, M., Nascimento, F.F., Carlisle, E., Asher, R.J., Yang, Z., Donoghue, P.C.J., and dos Reis, M. (2021). A Species-Level Timeline of Mammal Evolution Integrating Phylogenomic Data. Nature 1–8.

Anacleto T.C. da S. (2007). Food Habits of Four Armadillo Species in the Cerrado Area, Mato Grosso, Brazil. Zoological Studies 46, 529–537.

Arumugam, K., Bessarab, I., Haryono, M.A.S., Liu, X., Zuniga–Montanez, R.E., Roy, S., Qiu, G., Drautz–Moses, D.I., Law, Y.Y., Wuertz, S., *et al*. (2021). Recovery of complete genomes and non-chromosomal replicons from activated sludge enrichment microbial communities with long read metagenome sequencing. Npj Biofilms and Microbiomes 7, 1–13.

Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., *et al*. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nature Communications *11*, 2500.

Aya-Cuero, C., Chacón-Pacheco, J., and Anacleto, T.C.S. (2019). Dasypus kappleri (Cingulata: Dasypodidae). mmsp 51, 51–60.

Banerjee, S., and van der Heijden, M.G.A. (2023). Soil microbiomes and one health. Nat Rev Microbiol 21, 6–20.

Beier, S., and Bertilsson, S. (2013). Bacterial chitin degradation—mechanisms and ecophysiological strategies. Front. Microbiol. 4.

Bickhart, D.M., Kolmogorov, M., Tseng, E., Portik, D.M., Korobeynikov, A., Tolstoganov, I., Uritskiy, G., Liachko, I., Sullivan, S.T., Shin, S.B., *et al*. (2022). Generating lineage-resolved, complete metagenome-assembled genomes from complex microbial communities. Nat Biotechnol 1–9.

Blount, Z.D., Lenski, R.E., and Losos, J.B. (2018). Contingency and determinism in evolution: Replaying life's tape. Science 362, eaam5979.

Bøhle, L.A., Mathiesen, G., Vaaje-Kolstad, G., and Eijsink, V.G.H. (2011). An endo-β-N-acetylglucosaminidase from *Enterococcus faecalis* V583 responsible for the hydrolysis of high-mannose and hybrid-type N-linked glycans. FEMS Microbiology Letters 325, 123–129.

Bornbusch, S.L., Greene, L.K., Rahobilalaina, S., Calkins, S., Rothman, R.S., Clarke, T.A., LaFleur, M., and Drea, C.M. (2022). Gut microbiota of ring-tailed lemurs (Lemur catta) vary across natural and captive populations and correlate with environmental microbiota. Animal Microbiome 4, 29.

Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature Methods *12*, 59–60.

Buist, G., Steen, A., Kok, J., and Kuipers, O.P. (2008). LysM, a widely distributed protein motif for binding to (peptido)glycans. Mol Microbiol 68, 838–847.

Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Chaumeil, P.-A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

Chen, Y.-H., Chiang, P.-W., Rogozin, D.Y., Degermendzhy, A.G., Chiu, H.-H., and Tang, S.-L. (2021). Salvaging high-quality genomes of microbial species from a meromictic lake using a hybrid sequencing approach. Commun Biol 4, 996.

Cheng, S.-C., Liu, C.-B., Yao, X.-Q., Hu, J.-Y., Yin, T.-T., Lim, B.K., Chen, W., Wang, G.-D., Zhang, C.-L., Irwin, D.M., *et al*. (2022). Hologenomic insights into mammalian adaptations to myrmecophagy. Natl Sci Rev, nwac174.

Coleman, G.A., Davín, A.A., Mahendrarajah, T.A., Szánthó, L.L., Spang, A., Hugenholtz, P., Szöllősi, G.J., and Williams, T.A. (2021). A rooted phylogeny resolves early bacterial evolution. Science 372, eabe0511.

Cuscó, A., Pérez, D., Viñes, J., Fàbregas, N., and Francino, O. (2022). Novel canine high-quality metagenome-assembled genomes, prophages and host-associated plasmids provided by long-read metagenomics together with Hi-C proximity ligation. Microbial Genomics 8, 000802.

Delsuc, F., Scally, M., Madsen, O., Stanhope, M.J., de Jong, W.W., Catzeflis, F.M., Springer, M.S., and Douzery, E.J.P. (2002). Molecular Phylogeny of Living Xenarthrans and the Impact of Character and Taxon Sampling on the Placental Tree Rooting. Mol Biol Evol 19, 1656–1671.

Delsuc, F., Metcalf, J.L., Wegener Parfrey, L., Song, S.J., González, A., and Knight, R. (2014). Convergence of gut microbiomes in myrmecophagous mammals. Mol Ecol 23, 1301–1317.

Drula, E., Garron, M.-L., Dogan, S., Lombard, V., Henrissat, B., and Terrapon, N. (2022). The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Research *50*, D571–D577.

Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., *et al*. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science 356, 92–95.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLOS Computational Biology 7, e1002195.

Eizirik, E., Murphy, W.J., Koepfli, K.-P., Johnson, W.E., Dragoo, J.W., Wayne, R.K., and O'Brien, S.J. (2010). Pattern and timing of diversification of the mammalian order Carnivora inferred from multiple nuclear gene sequences. Molecular Phylogenetics and Evolution *56*, 49–63.

Emerling, C.A., Delsuc, F., and Nachman, M.W. (2018). Chitinase genes (CHIAs) provide genomic footprints of a post-Cretaceous dietary radiation in placental mammals. Science Advances 4, eaar6478.

Emerling, C.A., Gibb, G.C., Tilak, M.-K., Hughes, J.J., Kuch, M., Duggan, A.T., Poinar, H.N., Nachman, M.W., and Delsuc, F. (2022). Genomic data suggest parallel dental vestigialization within the xenarthran radiation. bioRxiv, 2022.12.09.519446.

Eren, A.M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S.E., Schechter, M.S., Fink, I., Pan, J.N., Yousef, M., Fogarty, E.C., *et al*. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6, 3–6.

Ezer, A., Matalon, E., Jindou, S., Borovok, I., Atamna, N., Yu, Z., Morrison, M., Bayer, E.A., and Lamed, R. (2008). Cell surface enzyme attachment is mediated by family 37 carbohydrate-binding modules, unique to *Ruminococcus albus*. J Bacteriol 190, 8220–8222.

Ferreira-Cardoso, S., Delsuc, F., and Hautier, L. (2019). Evolutionary Tinkering of the Mandibular Canal Linked to Convergent Regression of Teeth in Placental Mammals. Current Biology 29, 468-475.

Ferreira-Cardoso, S., Fabre, P.-H., Thoisy, B. de, Delsuc, F., and Hautier, L. (2020). Comparative masticatory myology in anteaters and its implications for interpreting morphological convergence in myrmecophagous placentals. PeerJ 8, e9690.

Ferreira-Cardoso, S., Claude, J., Goswami, A., Delsuc, F., and Hautier, L. (2022). Flexible conservatism in the skull modularity of convergently evolved myrmecophagous placental mammals. BMC Ecology and Evolution 22, 87.

Gooday, G.W. (1990). The Ecology of Chitin Degradation. In Advances in Microbial Ecology, K.C. Marshall, ed. (Boston, MA: Springer US), pp. 387–430.

Griffiths, M. (1968). Echidnas. Oxford; New York: Pergamon Press.

Groussin, M., Mazel, F., Sanders, J.G., Smillie, C.S., Lavergne, S., Thuiller, W., and Alm, E.J. (2017). Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. Nature Communications 8, 14319.

Gull, J.M., Stahl, M., Osmann, C., Ortmann, S., Kreuzer, M., Hatt, J.-M., and Clauss, M. (2015). Digestive physiology of captive giant anteaters (Myrmecophaga tridactyla): determinants of faecal dry matter content. Journal of Animal Physiology and Animal Nutrition 99, 565–576.

Hamid, R., Khan, M.A., Ahmad, M., Ahmad, M.M., Abdin, M.Z., Musarrat, J., and Javed, S. (2013). Chitinases: An update. J Pharm Bioallied Sci 5, 21–29.

Hehemann, J.-H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010). Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. Nature 464, 908–912.

Hyatt, D., Chen, G.-L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11, 119.

Irulan, A., Nathan, P. T., Priya, Y. S., Marimuthu, G., & Elangovan, V. (2011). Isolation and characterization of chitinase producing gut microflora of insectivorous bats. Trends in Biosciences 4, 8-11.

Jain, C., Rodriguez-R, L.M., Phillippy, A.M., Konstantinidis, K.T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9, 5114.

Janiak, M.C., Chaney, M.E., and Tosi, A.J. (2018). Evolution of Acidic Mammalian Chitinase Genes (*CHIA*) Is Related to Body Mass and Insectivory in Primates. Molecular Biology and Evolution 35, 607–622.

Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165.

Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30, 3059–3066.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30, 772–780.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., *et al*. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.

Kohl, K.D., Dieppa-Colón, E., Goyco-Blas, J., Peralta-Martínez, K., Scafidi, L., Shah, S., Zawacki, E., Barts, N., Ahn, Y., Hedayati, S., *et al*. (2022). Gut Microbial Ecology of Five Species of Sympatric Desert Rodents in Relation to Herbivorous and Insectivorous Feeding Strategies. Integrative and Comparative Biology icac045.

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol *37*, 540–546.

Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., *et al*. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17, 1103–1110.

La Reau, A.J., Meier-Kolthoff, J.P., and Suen, G. (2016). Sequence-based analysis of the genus *Ruminococcus* resolves its phylogeny and reveals strong host association. Microb Genom 2, e000099.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9, 357–359.

Lavrinienko, A., Tukalenko, E., Mousseau, T.A., Thompson, L.R., Knight, R., Mappes, T., and Watts, P.C. (2020). Two hundred and fifty-four metagenome-assembled bacterial genomes from the bank vole gut microbiota. Sci Data 7.

Leger, A., and Leonardi, T. (2019). pycoQC, interactive quality control for Oxford Nanopore Sequencing. JOSS 4, 1236.

Leisner, J. J., Larsen, M. H., Ingmer, H., Petersen, B. O., Duus, J. O., and Palcic, M. M. (2009). Cloning and comparison of phylogenetically related chitinases from *Listeria monocytogenes* EGD and *Enterococcus faecalis* V583. J. Appl. Microbiol. 107, 2080–2087.

Levin, D., Raab, N., Pinto, Y., Rothschild, D., Zanir, G., Godneva, A., Mellul, N., Futorian, D., Gal, D., Leviatan, S., *et al*. (2021). Diversity and functional landscapes in the microbiota of animals in the wild. Science 372, eabb5352.

Ley R.E, Lozupone C.A, Hamady M, Knight R, Gordon J.I. (2008a). Worlds within worlds: evolution of the vertebrate gut microbiota. Nature Reviews Microbiology, 6, 776–788.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R., *et al*. (2008b). Evolution of Mammals and Their Gut Microbes. Science 320, 1647–1651.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100.

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. Bioinformatics 37, 4572–4574.

Losos, J.B. (2011). Convergence, adaptation, and constraint. Evolution 65, 1827–1840.

Ma, J.-E, Li, L.-M., Jiang, H.-Y., Zhang, X.-J., Li, J., Li, G.-Y., Yuan, L.-H., Wu, J., and Chen, J.-P. (2017). Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*). PeerJ. 5:e4140.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, G.-Y., Li, H.-M., Jin, X.-J., and Chen, J.-P. (2018). The fecal metagenomics of Malayan pangolins identifies an extensive adaptation to myrmecophagy. Front. Microbiol 9, 2793.

Ma, J.-E., Jiang, H.-Y., Li, L.-M., Zhang, X.-J., Li, H.-M., Li, G.-Y., Mo, D.-Y., and Chen, J.-P. (2019). SMRT sequencing of the full-length transcriptome of the Sunda pangolin (*Manis javanic*a). Gene 692, 208-216.

McDonald, C., Barden, S., and Foley, S. (2014). Isolation and characterization of chitin-degrading micro-organisms from the faeces of Goeldi's monkey, *Callimico goeldii*. Journal of Applied Microbiology 116, 52–59.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., *et al*. (2013). Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 110, 3229–3236.

McNab, B.K. (1984). Physiological convergence amongst ant-eating and termite-eating mammals. Journal of Zoology 203, 485–510.

Magdeleine, A., Tilak, M.-K., Teullet, S., and Delsuc, F. (2021). High molecular weight bacterial DNA extraction from field-collected fecal samples preserved in ethanol for long-read sequencing. Protocol.Io.

Mallott, E.K., and Amato, K.R. (2021). Host specificity of the gut microbiome. Nat Rev Microbiol 19, 639-653.

Mazel, F., Davis, K., Loudon, A., Kwong, W., Groussin, M., and Wegener Parfrey, L. (2018). Is Host Filtering the Main Driver of Phylosymbiosis across the Tree of Life? mSystems 3, e00097-18.

Mazel, F., Guisan, A., and Parfrey, L.W. Transmission mode and dispersal traits correlate with host specificity in mammalian gut microbes. Molecular Ecology 0, 1–11.

Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eizirik, E., Simão, T.L.L., Stadler, T., *et al*. (2011). Impacts of the Cretaceous Terrestrial Revolution and KPg Extinction on Mammal Diversification. Science 334, 521–524.

Moeller, A.H., and Sanders, J.G. (2020). Roles of the gut microbiota in the adaptive evolution of mammalian species. Philosophical Transactions of the Royal Society B: Biological Sciences 375, 20190597.

Muegge, B.D., Kuczynski, J., Knights, D., Clemente, J.C., González, A., Fontana, L., Henrissat, B., Knight, R., and Gordon, J.I. (2011). Diet Drives Convergence in Gut Microbiome Functions Across Mammalian Phylogeny and Within Humans. Science 332, 970–974.

Musacchio, A., Wilmanns, M., and Saraster, M. (1994). Structure and function of the SH3 domain. Progress in Biophysics and Molecular Biology 61, 283–297.

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., Wu, D., Paez-Espino, D., Chen, I.-M., Huntemann, M., *et al*. (2020). A genomic catalog of Earth's microbiomes. Nature Biotechnology 1–11.

Nishida, A.H., and Ochman, H. (2018). Rates of gut microbiome divergence in mammals. Molecular Ecology 27, 1884–1897.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 32, 268–274.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834.

Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and Phillippy, A.M. (2016). Mash: fast genome and metagenome distance estimation using MinHash. Genome Biology 17, 132.

Olland, A.M., Strand, J., Presman, E., Czerwinski, R., Joseph-McCarthy, D., Krykbaev, R., Schlingmann, G., Chopra, R., Lin, L., Fleming, M., *et al*. (2009). Triad of polar residues implicated in pH specificity of acidic mammalian chitinase. Protein Sci. 18, 569–578.

Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017). dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. The ISME Journal 11, 2864–2868.

Olsen, M.A., Blix, A.S., Utsi, T., Sørmo, W., and Mathiesen, S.D. (1999). Chitinolytic bacteria in the minke whale forestomach. Can. J. Microbiol. 46, 85–94.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055.

Patterson, B. (1975). The Fossil Aardvarks (Mammalia: Tubulidentata). Bulletin Museum of Comparative Zoology 147, 185-237.

Pel, R., Hessels, G., Aalfs, H., and Gottschal, J.C. (1989). Chitin degradation by *Clostridium* sp. strain 9.1 in mixed cultures with saccharolytic and sulfate-reducing bacteria. FEMS Microbiology Letters 62, 191–200.

Perez-Lamarque, B., Sommeria-Klein, G., Duret, L., and Morlon, H. (2023). Phylogenetic comparative approach reveals evolutionary conservatism, ancestral composition, and integration of vertebrate gut microbiota. 2023.01.03.522549.

Perry, T., West, E., Eisenhofer, R., Stenhouse, A., Wilson, I., Laming, B., Rismiller, P., Shaw, M., and Grützner, F. (2022). Characterising the Gut Microbiomes in Wild and Captive Short-Beaked Echidnas Reveals Diet-Associated Changes. Frontiers in Microbiology 13, 687115.

Rathore, A.S., and Gupta, R.D. (2015). Chitinases from Bacteria to Human: Properties, Applications, and Future Perspectives. Enzyme Research, 791907.

Redford, K.H. (1985). Food habits of armadillos (Xenarthra, Dasypodidae) G.G. Montgomery (Ed.), The Evolution and Ecology of Armadillos, Sloths, and Vermilinguas, Smithsonian Institution Press, Washington, DC, pp. 429-438

Redford, K.H. (1986). Dietary specialization and variation in two mammalian myrmecophages (variation in mammalian myrmecophagy). Revista Chilena de Historia Natural 59, 201-208.

Redford, K.H. (1987). Ants and Termites As Food. In Current Mammalogy, H.H. Genoways, ed. (Boston, MA: Springer US), pp. 349–399.

Reiss, K.Z. (2001). Using Phylogenies to Study Convergence: The Case of the Ant-Eating Mammals. Integr Comp Biol 41, 507–525.

Revell, L.J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3, 217–223.

Sanders, J.G., Beichman, A.C., Roman, J., Scott, J.J., Emerson, D., McCarthy, J.J., and Girguis, P.R. (2015). Baleen whales host a unique gut microbiome with similarities to both carnivores and herbivores. Nature Communications 6, 8285.

Sangkhobol, V., and Skerman, V.B.D.Y. (1981). *Chitinophaga*, a New Genus of Chitinolytic Myxobacteria. International Journal of Systematic and Evolutionary Microbiology 31, 285–293.

Sebaihia, M., Peck, M. W., Minton, N. P., Thomson, N. R., Holden, M. T., Mitchell, W. J., *et al*. (2007). Genome sequence of a proteolytic (Group I) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. Genome Res. 17, 1082–1092

Segata, N., Börnigen, D., Morgan, X.C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. Nature Communications 4, 2304.

Šimůnek, J., Hodrová, B., Bartoňová, H., and Kopečný, J. (2001). Chitinolytic bacteria of the mammal digestive tract. Folia Microbiol 46, 76–78.

Simůnek, J., Kopecný, J., Hodrová, B., and Bartonová, H. (2002). Identification and characterization of *Clostridium paraputrificum*, a chitinolytic bacterium of human digestive tract. Folia Microbiol (Praha) 47, 559–564.

Singleton, C.M., Petriglieri, F., Kristensen, J.M., Kirkegaard, R.H., Michaelsen, T.Y., Andersen, M.H., Kondrotaite, Z., Karst, S.M., Dueholm, M.S., Nielsen, P.H., *et al*. (2021). Connecting structure to function with the recovery of over 1000 high-quality metagenome-assembled genomes from activated sludge using long-read sequencing. Nature Communications 12, 2009.

Shaffer, J.P., Nothias, L.-F., Thompson, L.R., Sanders, J.G., Salido, R.A., Couvillion, S.P., Brejnrod, A.D., Lejzerowicz, F., Haiminen, N., Huang, S., *et al*. (2022). Standardized multi-omics of Earth's microbiomes reveals microbial and metabolite diversity. Nat Microbiol 7, 2128–2150.

Shakir, H.A., Irfan, M., and Qazi, J.I. (2016). Chitinase production in organisms: A Review. Punjab Univ. K. Zool 31, 101-106.

Smith, S. (1996). Isolation and Characterization of Chitinases From *Dasypus novemcinctus*, the Nine-Banded Armadillo. MSU Graduate Theses.

Song, S.J., Sanders, J.G., Delsuc, F., Metcalf, J., Amato, K., Taylor, M.W., Mazel, F., Lutz, H.L., Winker, K., Graves, G.R., *et al*. (2020). Comparative Analyses of Vertebrate Gut Microbiomes Reveal Convergence between Birds and Bats. mBio 11.

Spaink, H.P. (2004). Specific recognition of bacteria by plant LysM domain receptor kinases. Trends in Microbiology 12, 201–204.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 131.

Stevens, C.E., and Hume, I.D. (1998). Contributions of Microbes in Vertebrate Gastrointestinal Tract to Production and Conservation of Nutrients. Physiological Reviews 78, 393–427.

Suzuki, T.A. (2017). Links between Natural Variation in the Microbiome and Host Fitness in Wild Mammals. Integrative and Comparative Biology 57, 756–769.

Swiontek Brzezinska, M., Jankiewicz, U., Burkowska, A., and Walczak, M. (2014). Chitinolytic Microorganisms and Their Possible Application in Environmental Protection. Curr Microbiol 68, 71–81.

Thomas, C.M., Desmond-Le Quéméner, E., Gribaldo, S., and Borrel, G. (2022). Factors shaping the abundance and diversity of the gut archaeome across the animal kingdom. Nat Commun *13*, 3358.

Vaaje-Kolstad, G., Bunaes, A.C., Mathiesen, G., and Eijsink, V.G.H. (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 J 276, 2402–2415.

Vaaje-Kolstad, G., Bøhle, L. A., Gåseidnes, S., Dalhus, B., Bjørås, M., Mathiesen, G., *et al*. (2012). Characterization of the chitinolytic machinery of *Enterococcus faecalis* V583 and high-resolution structure of its oxidative CBM33 enzyme. J. Mol. Biol. 416, 239–254.

Vacca, M., Celano, G., Calabrese, F.M., Portincasa, P., Gobbetti, M., and De Angelis, M. (2020). The Controversial Role of Human Gut Lachnospiraceae. Microorganisms 8, 573.

Vaz, V.C., Santori, R.T., Jansen, A.M., Delciellos, A.C., and D'Andrea, P.S. (2012). Notes on Food Habits of Armadillos (Cingulata, Dasypodidae) and Anteaters (Pilosa, Myrmecophagidae) at Serra Da Capivara National Park (PiauÍ State, Brazil). eden 13, 84–89.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., *et al*. (2014). Pilon: An Integrated Tool for Comprehensive Microbial Variant Detection and Genome Assembly Improvement. PLOS ONE *9*, e112963.

Wardman, J.F., Bains, R.K., Rahfeld, P., and Withers, S.G. (2022). Carbohydrate-active enzymes (CAZymes) in the gut microbiome. Nat Rev Microbiol 20, 542–556.

Weisenfeld, N.I., Yin, S., Sharpe, T., Lau, B., Hegarty, R., Holmes, L., Sogoloff, B., Tabbaa, D., Williams, L., Russ, C., *et al*. (2014). Comprehensive variation discovery in single human genomes. Nat Genet 46, 1350–1355.

Whitaker, J.O., Jr., Dannelly, H.K., and Prentice, D.A. (2004). Chitinase in Insectivorous Bats. Journal of Mammalogy 85, 15–18.

Wick, R.R., Judd, L.M., Gorrie, C.L., and Holt, K.E.Y. 2017 Completing bacterial genome assemblies with multiplex MinION sequencing. Microbial Genomics 3, e000132.

Wick. 2021. Filtlong v0.2.1. https://github.com/rrwick/Filtlong. Retrieved February 2022.

Xu, Q., Morrison, M., Nelson, K.E., Bayer, E.A., Atamna, N., and Lamed, R. (2004). A novel family of carbohydrate-binding modules identified with *Ruminococcus albus* proteins. FEBS Lett 566, 11–16.

Youngblut, N.D., Reischer, G.H., Walters, W., Schuster, N., Walzer, C., Stalder, G., Ley, R.E., and Farnleitner, A.H. (2019). Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. Nature Communications 10, 2200.

Youngblut, N.D., de la Cuesta-Zuluaga, J., Reischer, G.H., Dauser, S., Schuster, N., Walzer, C., Stalder, G., Farnleitner, A.H., and Ley, R.E. (2020). Large-Scale Metagenome Assembly Reveals Novel Animal-Associated Microbial Genomes, Biosynthetic Gene Clusters, and Other Genetic Diversity. mSystems 5, e01045-20.

Zepeda Mendoza, M.L., Xiong, Z., Escalera-Zamudio, M., Runge, A.K., Thézé, J., Streicker, D., Frank, H.K., Loza-Rubio, E., Liu, S., Ryder, O.A., *et al.* (2018). Hologenomic adaptations underlying the evolution of sanguivory in the common vampire bat. Nat Ecol Evol 2, 659–668.

Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., Busk, P.K., Xu, Y., and Yin, Y. (2018). dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 46, W95–W101.

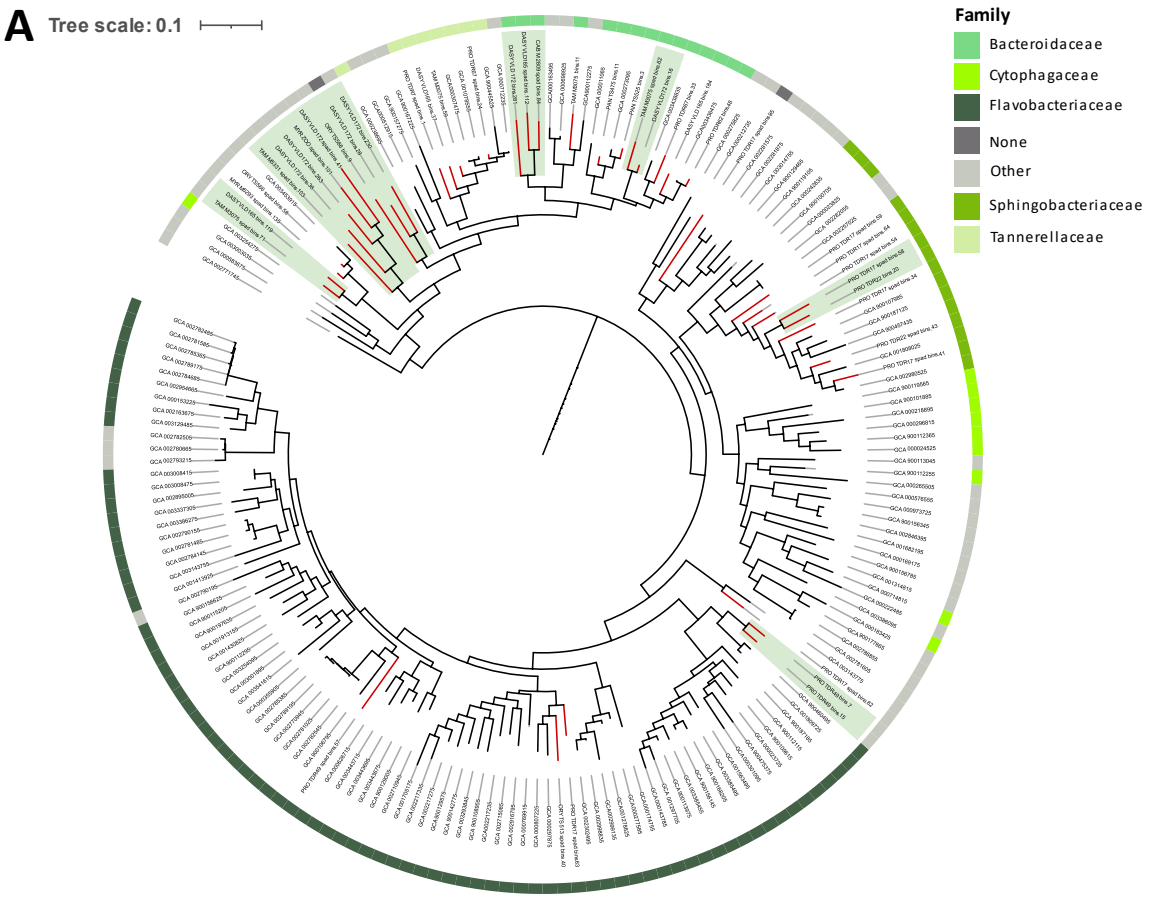
Zilber-Rosenberg, I., and Rosenberg, E. (2008). Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiology Reviews 32, 723–735.

Zimin, A.V., Puiu, D., Luo, M.-C., Zhu, T., Koren, S., Marçais, G., Yorke, J.A., Dvořák, J., and Salzberg, S.L. (2017). Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. Genome Res 27, 787–792.

Supplementary material

Table S1. Raw results of the different analyses conducted on each gut metagenome to reconstruct high-quality genome bins from raw metagenomic data for each dataset (long- and short-reads).

Table available via Zenodo.



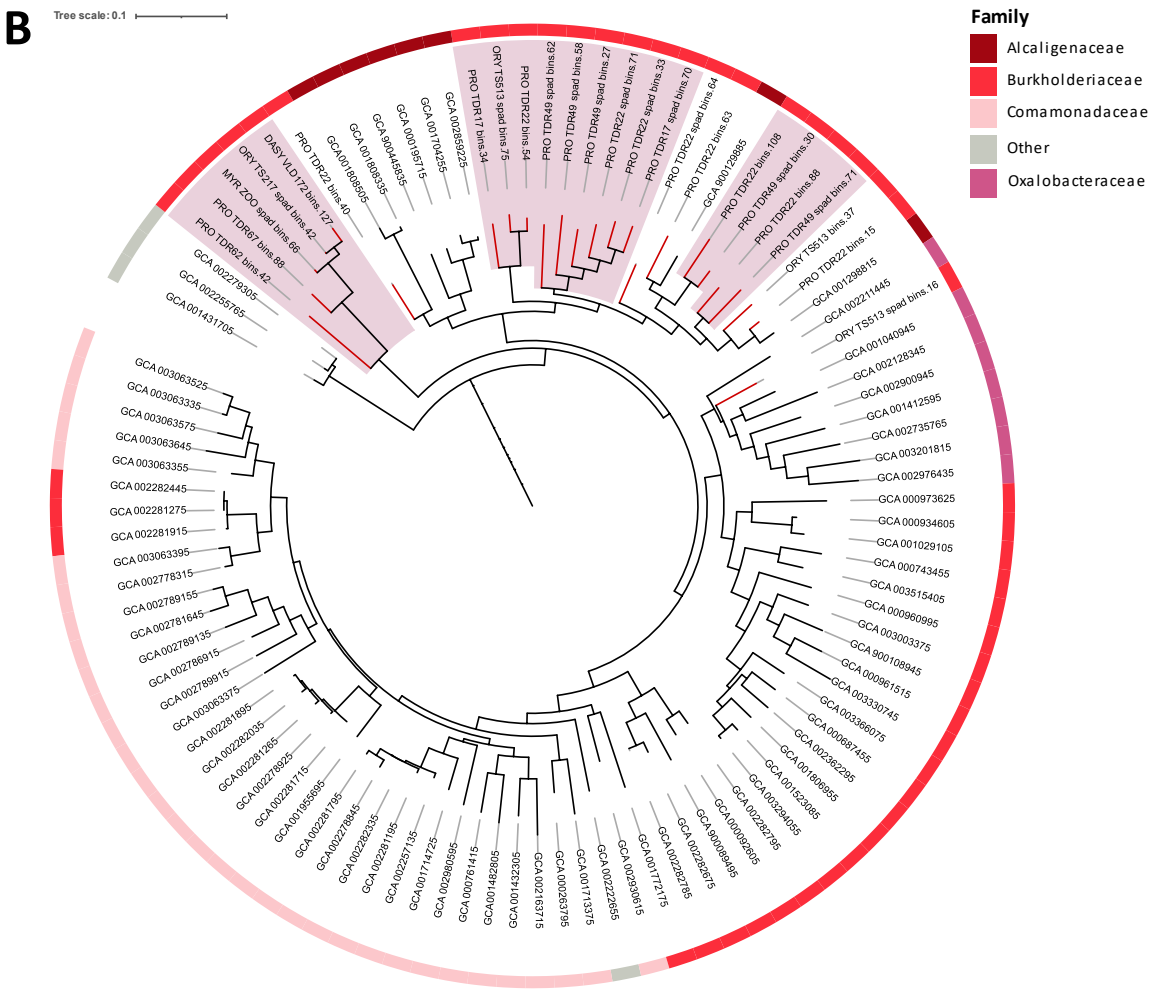


Figure. S1 Myrmecophagous-specific clades within Bacteroidetes (A; green highlights) and **Proteobacteria** (B; pink highlights). The two trees are subtrees of Figure 1. Outer circles indicate the bacterial family to which these genome bins were assigned based on the Genome Taxonomy Database (Chaumeil *et al*, 2020).

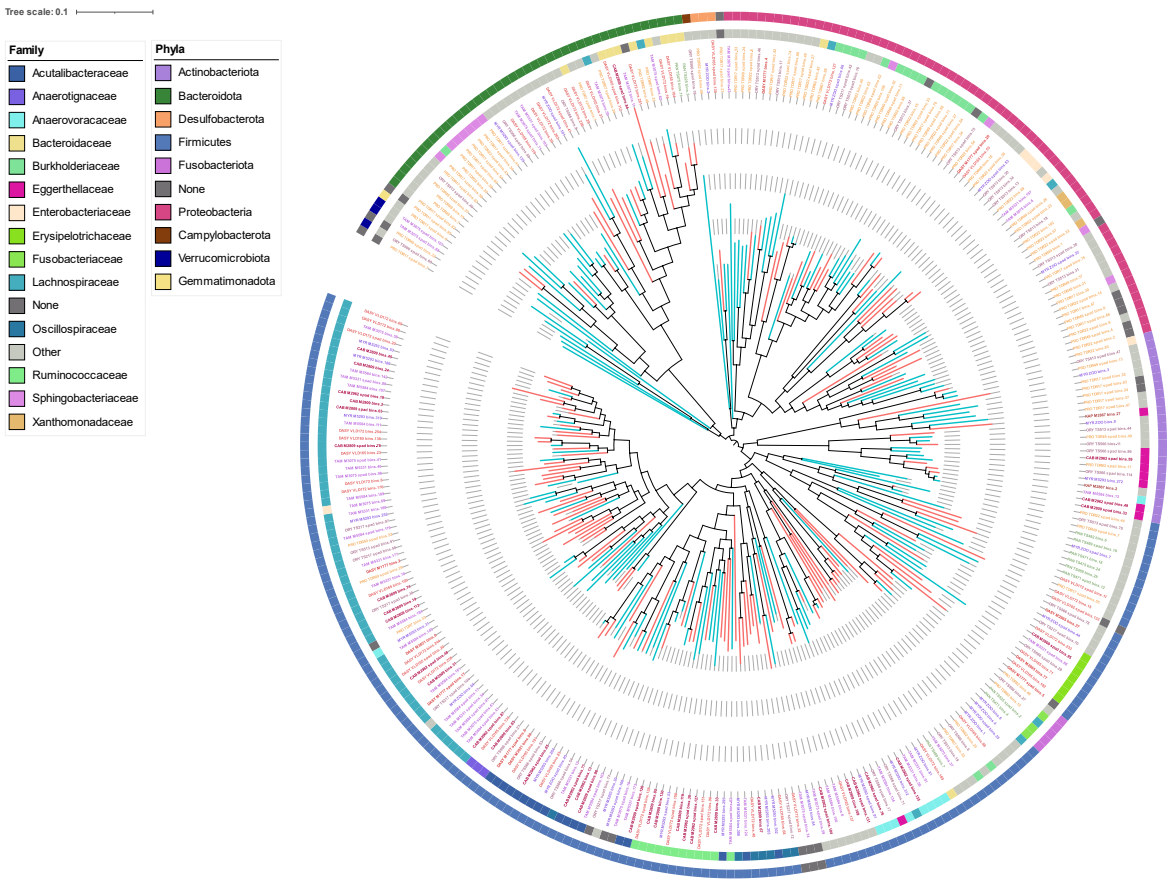


Figure. S2. Phylogeny of the 314 high-quality selected bins reconstructed from long-read assemblies (n = 156; red branches) **and short-read assemblies** (n = 158; blue branches)**.** Circles respectively indicate (from inner to outer circles): the bacterial family and phyla the bin was assigned to based on the Genome Taxonomy Database (Chaumeil *et al*, 2020). Colored sequence names indicate the host species.

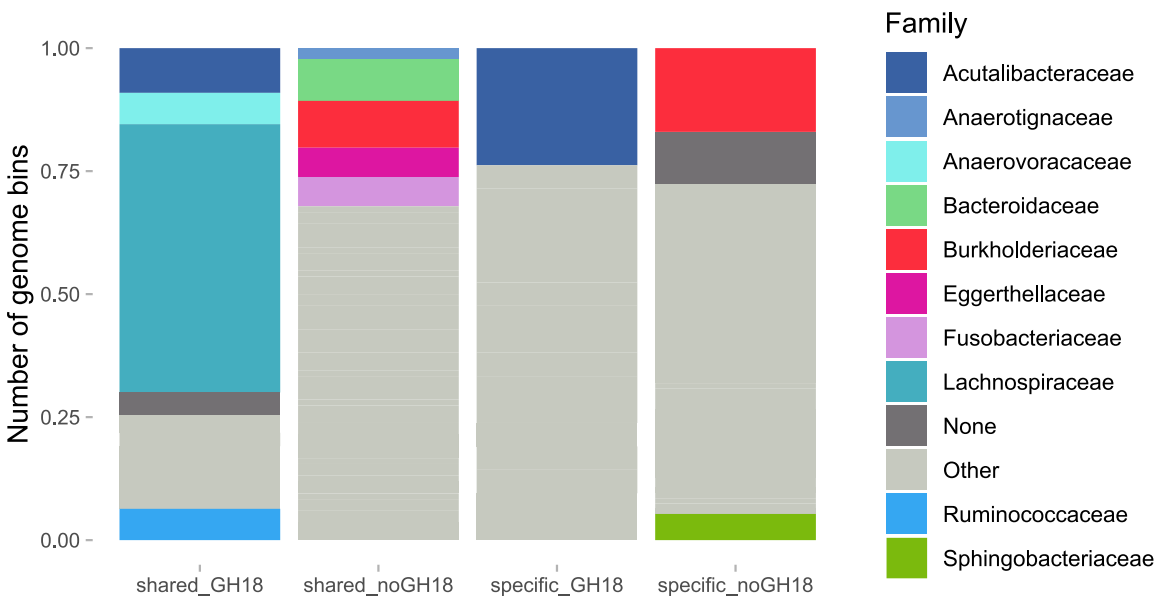


Figure. S3. Bacterial families of host-species specific and shared genome bins carrying or not carrying GH18 genes.

Table S2. Proportion of chitinolytic genome bins (genomes having at least one GH18 with an active chitinolytic site) **detected in the nine focal myrmecophagous species.**

Species	# chitinolytic bins detected	total # detected bins	# chitinolytic bins / total # bins detected
Cabassous unicinctus	46	113	0.407079646
Dasypus sp. nov. FG	48	105	0.457142857
Dasypus novemcinctus	50	129	0.387596899
Dasypus kappleri	2	25	0.08
Myrmecophaga tridactyla	48	141	0.340425532
Orycteropus afer	15	97	0.154639175
Smutsia temminckii	5	18	0.277777778
Proteles cristatus	12	104	0.115384615
Tamandua tetradactyla	37	121	0.305785124

Table S3. Detailed sample information for the eight soil samples collected in South Africa.

Sample	Sample type	Species	Common name	Class	Order	Family	Location
TDR012	Soil from Midden #1	*Proteles cristatus*	Southern ardwolf	Mammalia	Carnivora	Hyaenidae	Tussen Die Riviere Reserve
TDR014	Nest fragments and soil	*Trinervitermes sp.*	Snouted termites	Insecta	Blattodea	Termitidae	Tussen Die Riviere Reserve
TDR019	Soil from Midden #2	*Proteles cristatus*	Southern ardwolf	Mammalia	Carnivora	Hyaenidae	Tussen Die Riviere Reserve
TDR023	Soil from Midden #3	*Proteles cristatus*	Southern ardwolf	Mammalia	Carnivora	Hyaenidae	Tussen Die Riviere Reserve
TS218	Soil from midden	*Orycteropus afer*	Aardvark	Mammalia	Tubulidentata	Orycteropodidae	Tswalu Kalahari Reserve
TS274	Nest fragments and soil	*Trinervitermes sp.*	Snouted termites	Insecta	Blattodea	Termitidae	Tswalu Kalahari Reserve
TS281	Soil from midden	*Orycteropus afer*	Aardvark	Mammalia	Tubulidentata	Orycteropodidae	Tswalu Kalahari Reserve
TS298	Soil from midden	*Orycteropus afer*	Aardvark	Mammalia	Tubulidentata	Orycteropodidae	Tswalu Kalahari Reserve

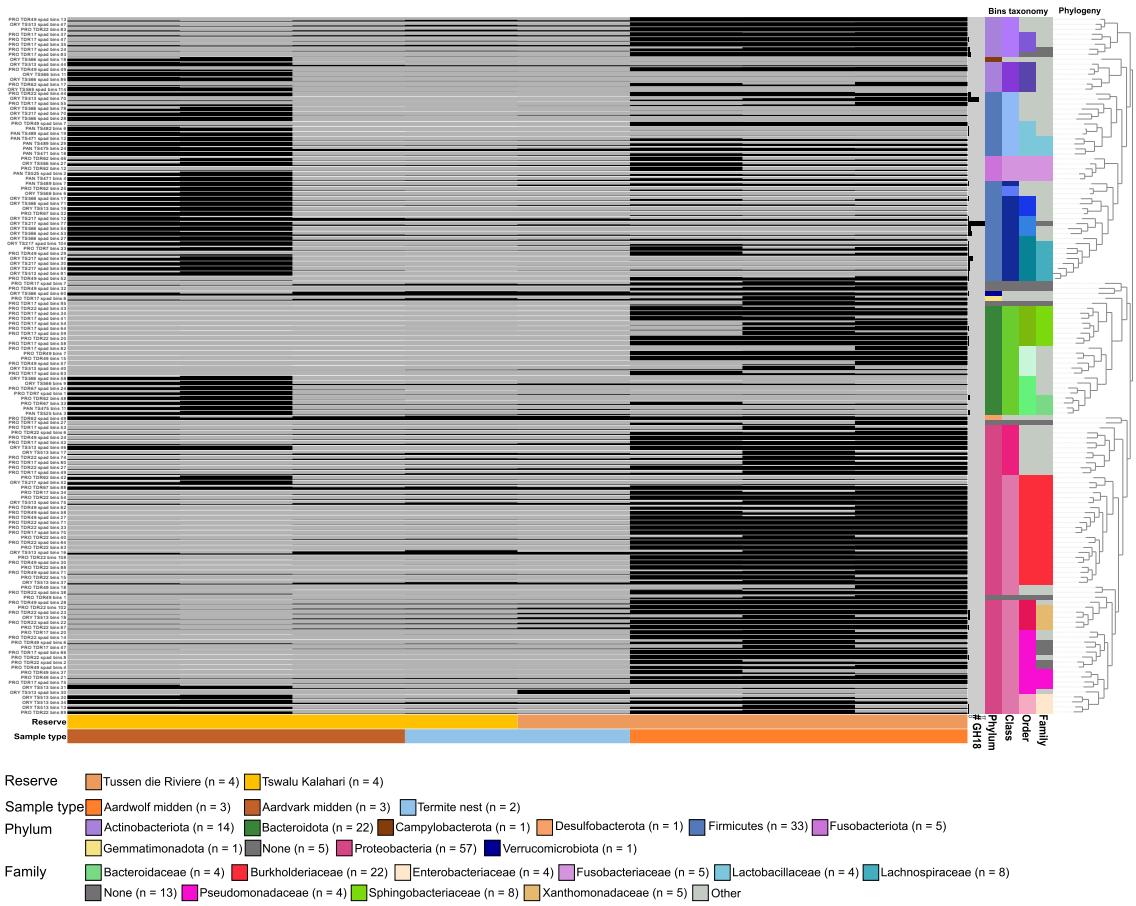


Figure S4. Detection of the 140 high-quality selected bins reconstructed from the aardvark, ground pangolin, and southern aardwolf samples (lines) **in the eight soil samples** (columns) **collected in South Africa near feces sampling sites.** Each square indicates the detection of a bin in a sample as estimated by anvi'o v7 (Eren *et al*, 2021). Bin names are indicated on the left. The provenance of each sample is indicated by different colors at the bottom of the graph. Columns on the right indicate (from left to right): the number of GH18 sequences identified in each bin (from 0 to 17), the bin's taxonomic phylum, class, order, and family. The phylogeny of the 140 selected bins inferred with PhyloPhlAn v3.0.58 (Asnicar *et al*, 2020) is also represented on the right of the graph. See detection table available via Zenodo for detailed values of detection.

Supplementary results (also available on Zenodo)

1. Long- and short-read sequencing enabled the reconstruction of diverse high-quality bacterial genome bins

We reconstructed 314 high-quality genome bins (Table S1F) of completeness > 90% and redundancy < 5% (Fig sup result 1A and B). A similar number of total and high-quality selected genome bins were reconstructed using the long- and short-read assemblies (respectively 156 and 158, Table S1F). These bins had characteristics of bacterial genomes with a total length on average of 2 653 051 and 2 350 576 Mbp respectively for the long-read and short-read datasets (Fig sup result 1C; Table S1F) and a number of predicted genes on average of 2 466 and 2 185 for the long-read and short-read datasets respectively (Fig sup result 1D; Table S1F). The similar number of predicted genes found between bins reconstructed from long- or short-read assemblies (Fig sup result 1D) is likely due to the polishing step. This step ensures better open reading frame predictions as frameshifts were corrected, further confirming the usefulness of polishing long-read assemblies with short reads (Latorre-Pérez *et al*, 2020). metaFlye has been shown to be more efficient compared to other long-read assemblers (Hu *et al*, 2020; Kolmogorov *et al*, 2020; Latorre-Pérez *et al*, 2020; Buttler and Brown, 2022) and was used here to reconstruct contiguous assemblies. As expected, genome bins reconstructed from long-read polished assemblies were more contiguous than those reconstructed from short-read assemblies with on average fewer contigs (Fig sup result 1E), a higher N50 (Fig sup result 1F), and a higher size of the longest contig (Fig sup result 1G) (Table S1F).

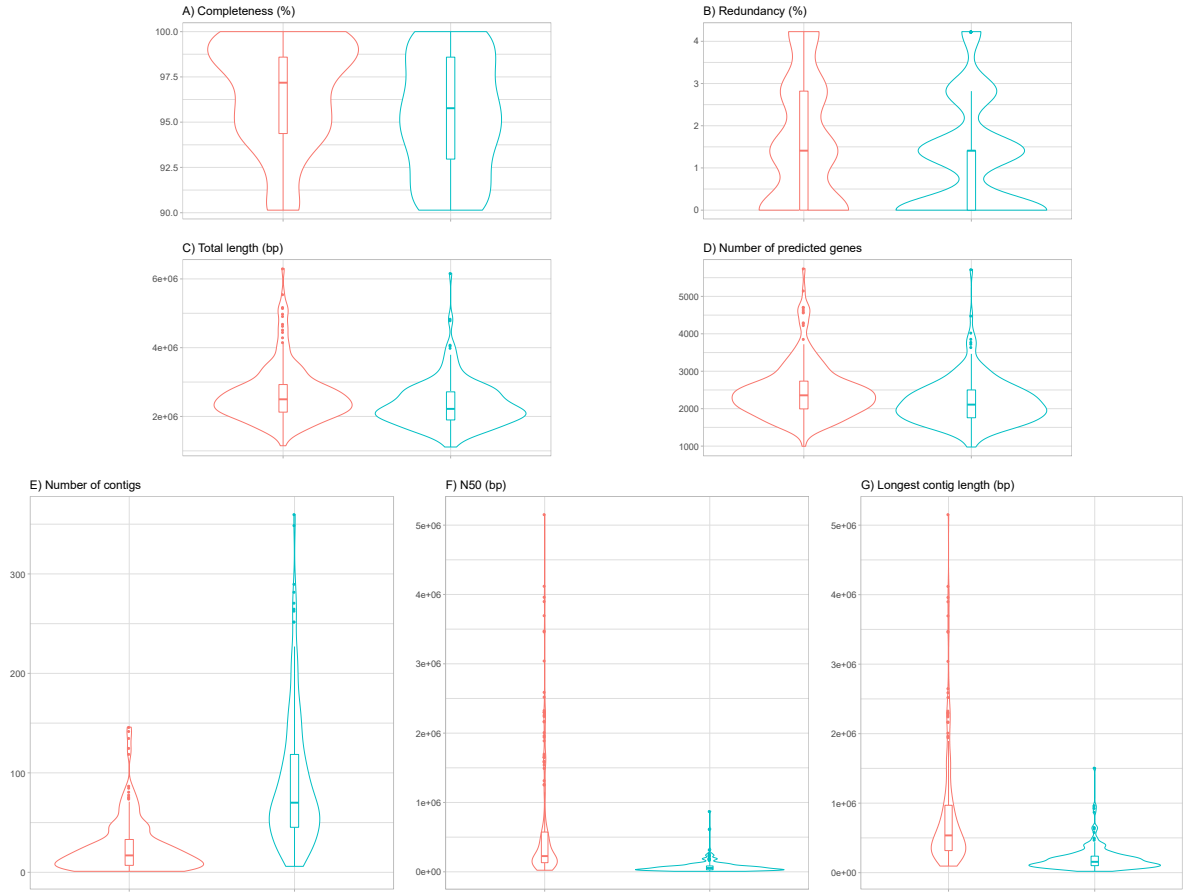


Figure sup result 1. Comparison of high-quality selected genome bin statistics between bins (> 90% completeness, < 5 % redundancy) **reconstructed from long-read polished assemblies** (red, n = 201) **and short-read assemblies** (blue, n = 206).

A phylogeny of the 407 selected bins was inferred using PhyloPhlAn v3.0.58 (Asnicar *et al*, 2020) (Fig sup result 2) as described in the Materials and Methods. The phylogeny was rooted according to Coleman *et al* (2021), placing the root of the bacterial tree separating Terrabacteria (e.g., Firmicutes, Actinobacteriota) and Gracilicutes (e.g., Proteobacteria, Desulfobacterota, Campylobacterota, Fibrobacteres-Chlorobi-Bacteroidetes group). These analyses revealed the presence of similar genomes in the same samples reconstructed using the two different datasets. Indeed, they were clustered together during dereplication (see outputs of dRep analyses available via Zenodo DOI 10.5281/zenodo.7808597) and were closely related (with short branches) in the phylogeny of the selected bins (Fig sup result 2) Some bacterial genomes were also reconstructed in only one of the two datasets (Fig sup result 2). This is probably due to the higher sequencing depth of the Illumina strategy which might have enabled the sequencing of bacterial taxa of low abundance. Differences in GC content might also have biased the sequencing, particularly when using high-throughput short-read sequencing (Dohm *et al*, 2008). To capture the majority of the bacterial taxa

present in our samples, we decided to combine high-quality selected bins reconstructed from both long- and short-read assemblies dereplicated with an ANI of 98% in our final dataset for downstream analyses (n = 314; Fig S2 supplementary material).

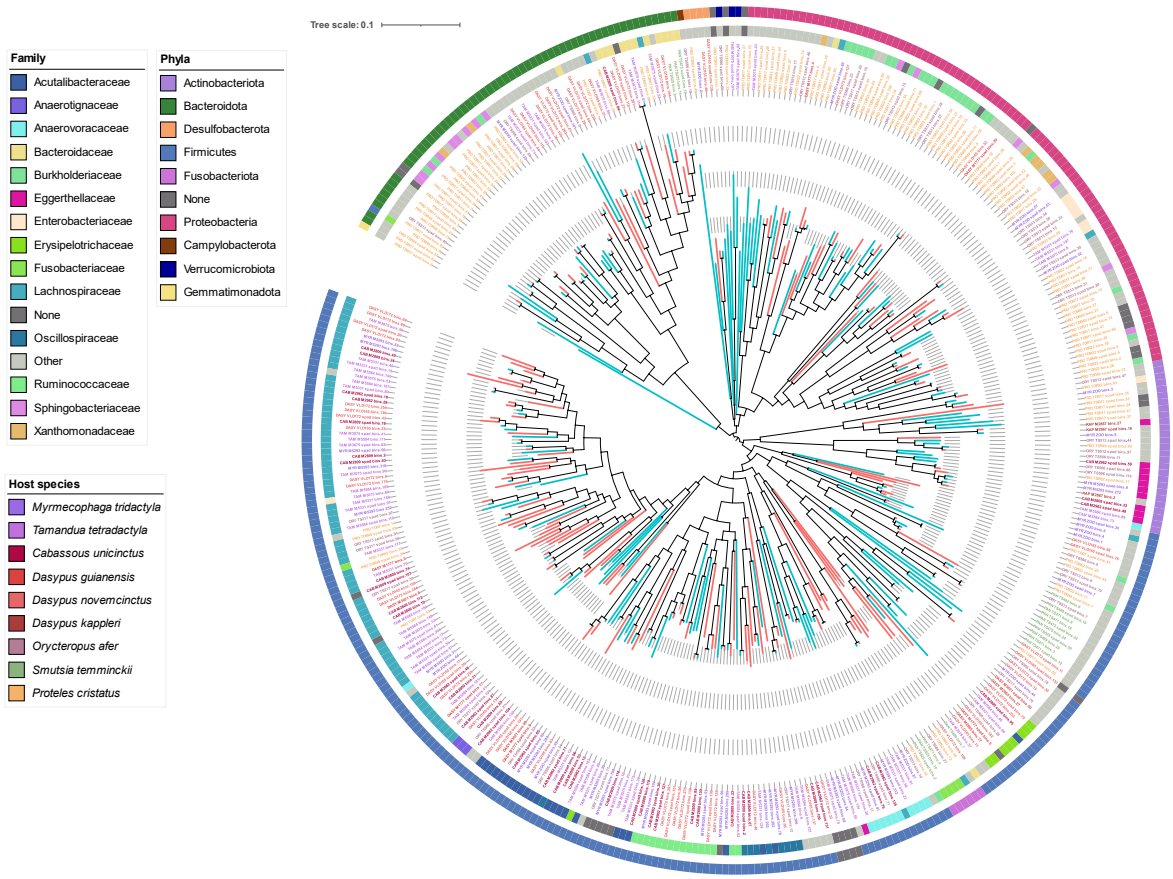


Figure sup result 2. Phylogeny of the 407 high-quality selected genome bins not dereplicated and reconstructed from long-read assemblies (n = 201; red branches) and short-read assemblies (n = 206; blue branches). Circles respectively indicate (from inner to outer circles): the bacterial family and phyla the bin was assigned to based on the Genome Taxonomy Database (Chaumeil *et al*, 2020). Colored sequence names indicate the host species.

2. Distribution of the 314 high-quality selected genome bins across the 29 fecal samples of the nine focal myrmecophagous species

There were more shared selected bins between host species carrying GH18 genes than shared bins with no GH18 gene. At the host order level, 54% of selected bins shared among host orders have at least one GH18 gene and belong to 15 bacterial families (mainly the Lachnospiraceae, n = 59) whereas shared bins not carrying chitinase genes belong to 34 different bacterial families, mainly the Burkholderiaceae (n = 8) and Bacteroidaceae (n = 7) (Fig sup result 3). 26% of bins found in only one host order carry at least one chitinase gene and belong to 15 bacterial families (mainly the Acutalibacteraceae, n = 10), and those not having GH18 genes are found in 39 families, mainly the Burkholderiaceae (n = 16) (Fig sup result 3).

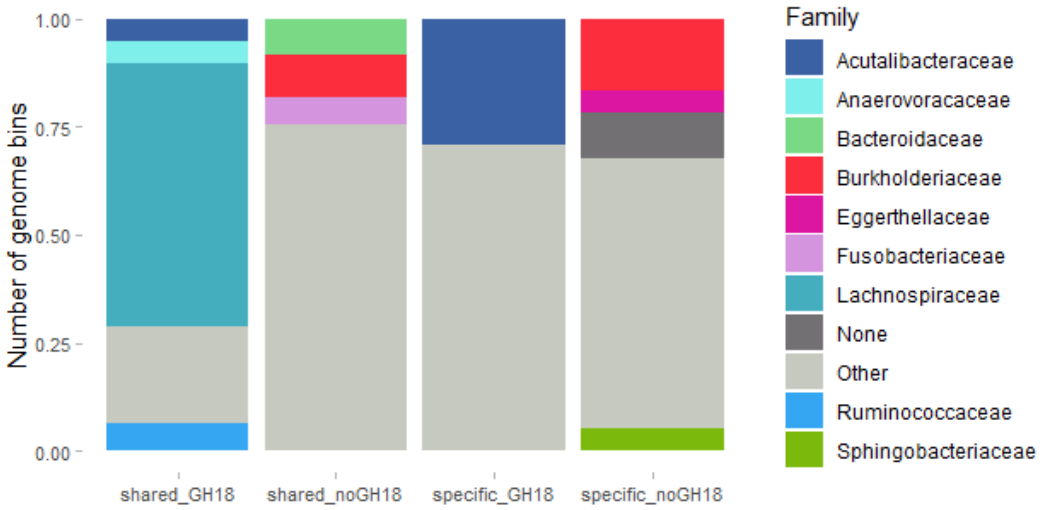


Figure sup result 3. Bacterial families of host-order specific and shared genome bins carrying or not carrying GH18 genes.

Supplementary results references

Asnicar, F., Thomas, A.M., Beghini, F., Mengoni, C., Manara, S., Manghi, P., Zhu, Q., Bolzan, M., Cumbo, F., May, U., *et al*. (2020). Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. Nature Communications *11*, 2500.

Buttler, J., and Drown, D. (2022). Accuracy and completeness of long read metagenomic assemblies. 2022.11.23.517703.

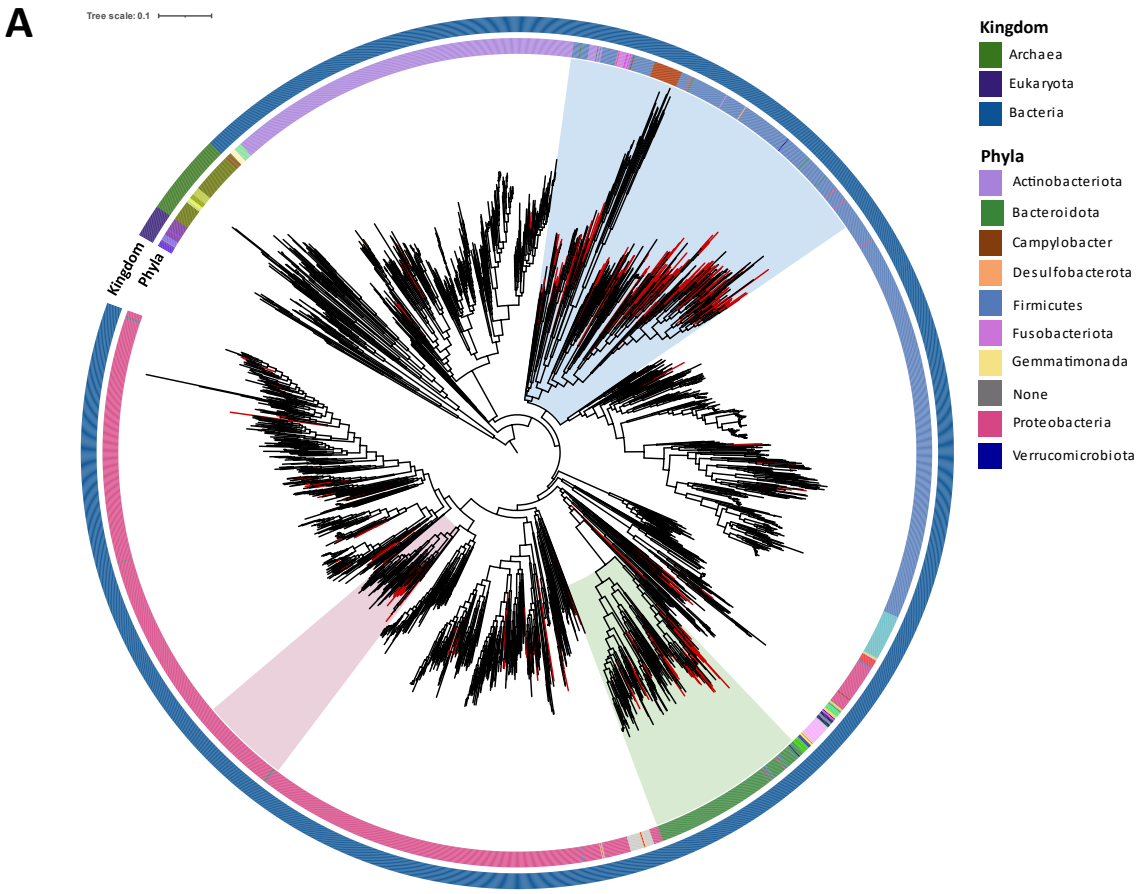
Coleman, G.A., Davín, A.A., Mahendrarajah, T.A., Szánthó, L.L., Spang, A., Hugenholtz, P., Szöllősi, G.J., and Williams, T.A. (2021). A rooted phylogeny resolves early bacterial evolution. Science 372, eabe0511.

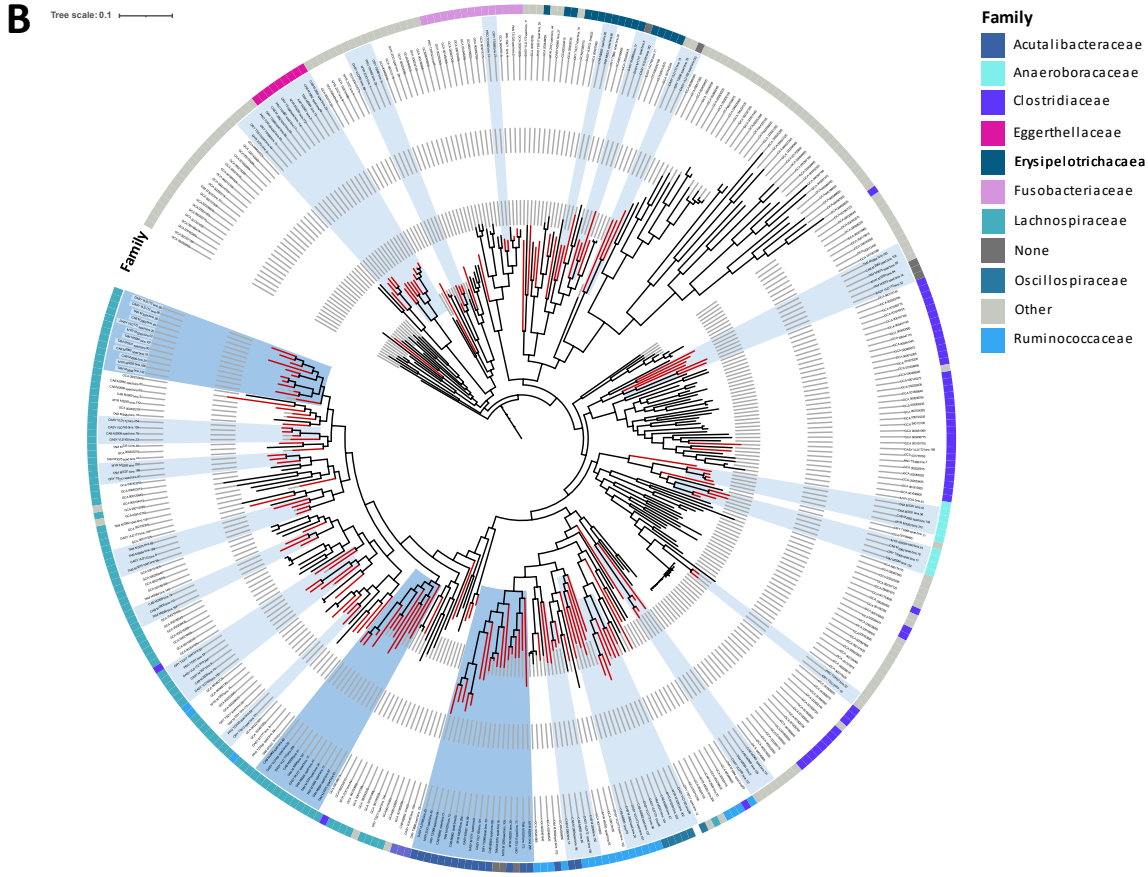
Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res 36, e105.

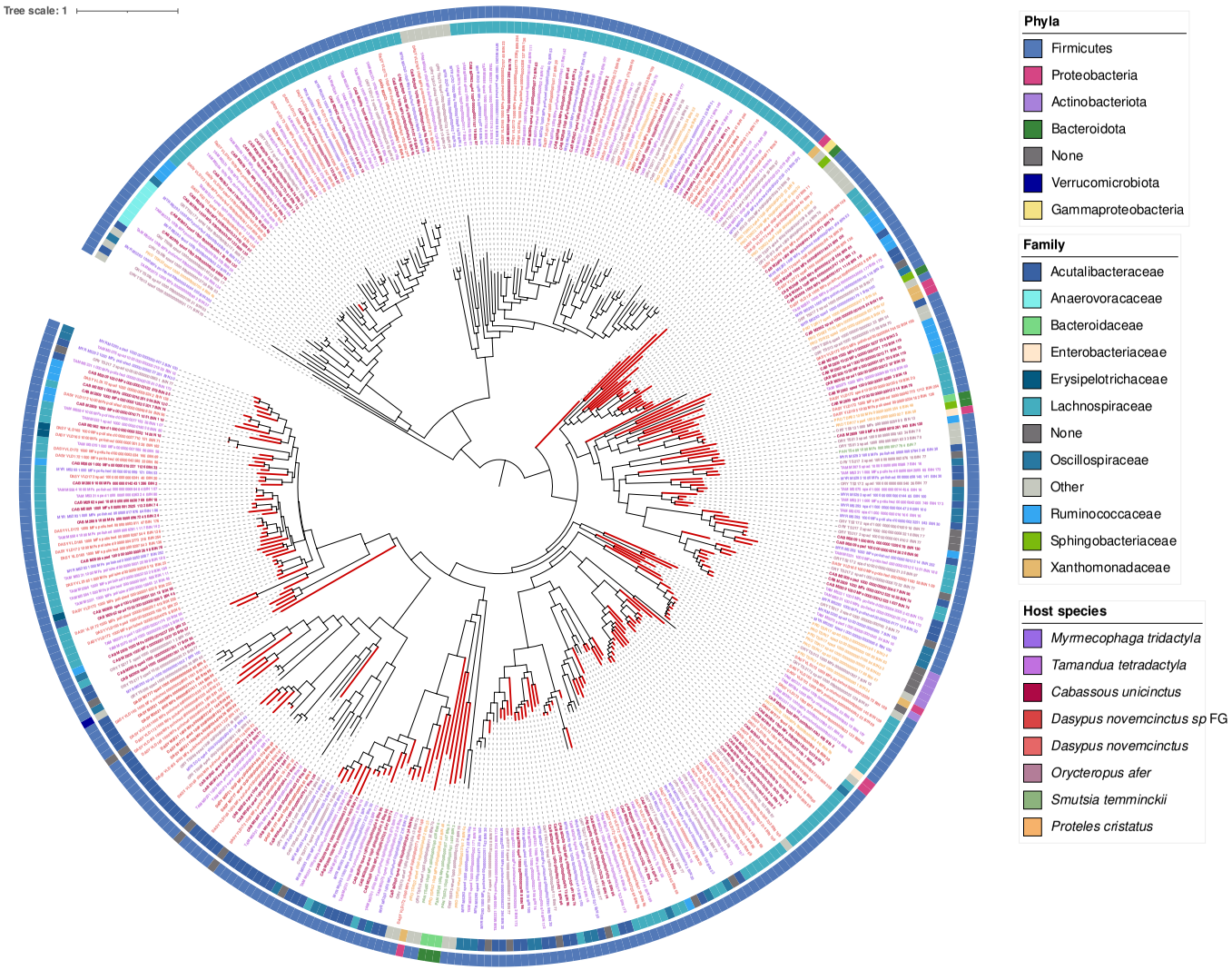
Hu, Y., Fang, L., Nicholson, C., and Wang, K. (2020). Implications of Error-Prone Long-Read Whole-Genome Shotgun Sequencing on Characterizing Reference Microbiomes. IScience 23, 101223.

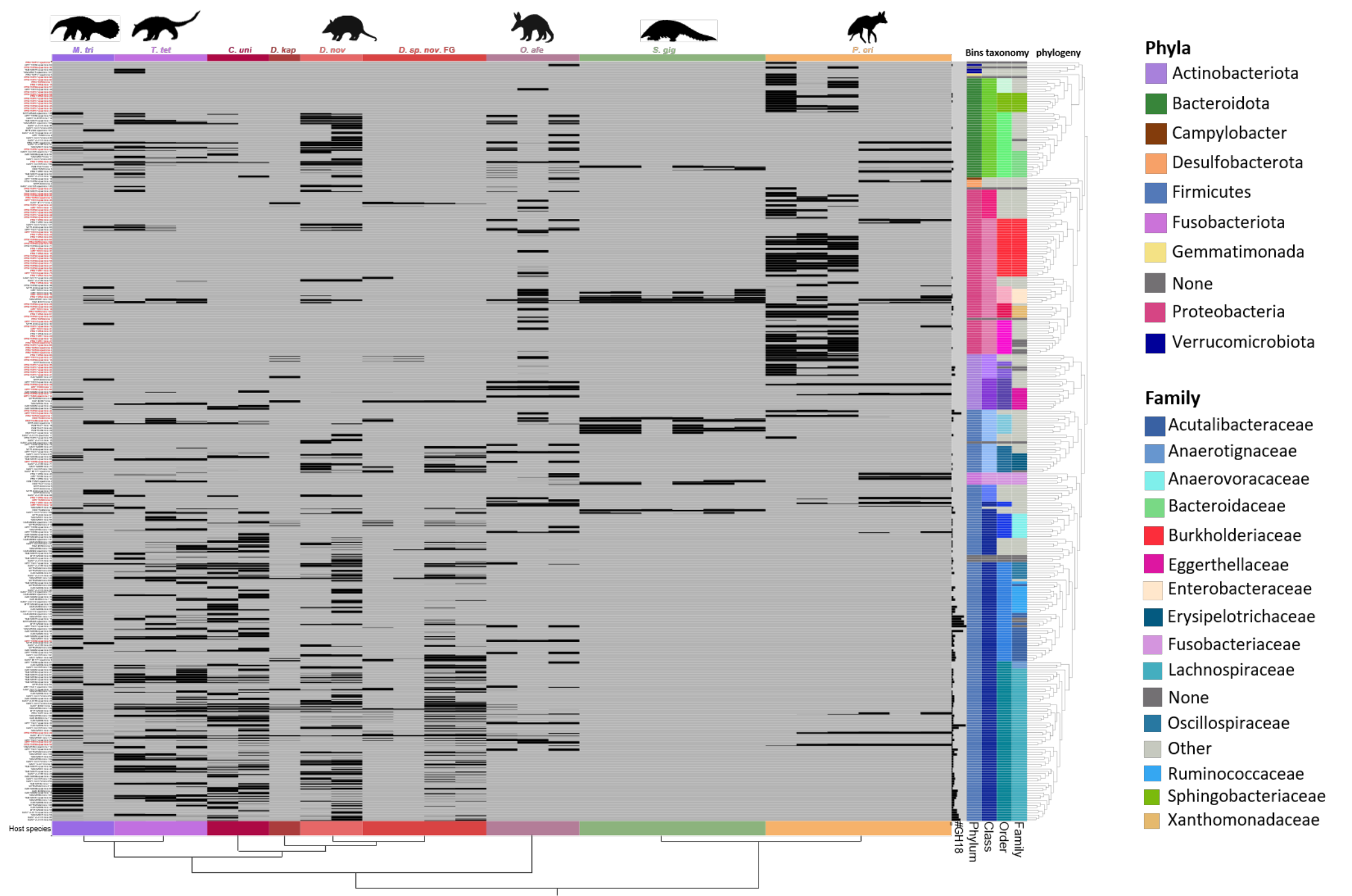
Kolmogorov, M., Bickhart, D.M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S.B., Kuhn, K., Yuan, J., Polevikov, E., Smith, T.P.L., *et al*. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods 17, 1103–1110.

Latorre-Pérez, A., Villalba-Bermell, P., Pascual, J., and Vilanova, C. (2020). Assembly methods for nanopore-based metagenomic sequencing: a comparative study. Sci Rep 10, 13588.


A
Tree scale: 0.1
Kingdom
Phyla
Kingdom
Archaea
Eukaryota
Bacteria
Phyla
Actinobacteriota
Bacteroidota
Campylobacter
Desulfobacterota
Firmicutes
Fusobacteriota
Gemmatimonada
None
Proteobacteria
Verrucomicrobiota


B
Tree scale: 0.1
Family
Family
Acutalibacteraceae
Anaeroboracaceae
Clostridiaceae
Eggerthellaceae
Erysipelotrichacaea
Fusobacteriaceae
Lachnospiraceae
None
Oscillospiraceae
Other
Ruminococcaceae


Tree scale: 1
Phyla
Firmicutes
Proteobacteria
Actinobacteriota
Bacteroidota
None
Verrucomicrobiota
Gammaproteobacteria
Family
Acutalibacteraceae
Anaerovoracaceae
Bacteroidaceae
Enterobacteriaceae
Erysipelotrichaceae
Lachnospiraceae
None
Oscillospiraceae
Other
Ruminococcaceae
Sphingobacteriaceae
Xanthomonadaceae
Host species
Myrmecophaga tridactyla
Tamandua tetradactyla
Cabassous unicinctus
Dasypus novemcinctus sp FG
Dasypus novemcinctus
Orycteropus afer
Smutsia temminckii
Proteles cristatus


M. tri
T. tet
C. uni
D. kap
D. nov
D. sp. nov. FG
O. afe
S. gig
P. ori
Bins taxonomy
phylogeny
Host species
#GH18
Phylum
Class
Order
Family
Phyla
Actinobacteriota
Bacteroidota
Campylobacter
Desulfobacterota
Firmicutes
Fusobacteriota
Gemmatimonada
None
Proteobacteria
Verrucomicrobiota
Family
Acutalibacteraceae
Anaerotignaceae
Anaerovoracaceae
Bacteroidaceae
Burkholderiaceae
Eggerthellaceae
Enterobacteraceae
Erysipelotrichaceae
Fusobacteriaceae
Lachnospiraceae
None
Oscillospiraceae
Other
Ruminococcaceae
Sphingobacteriaceae
Xanthomonadaceae

Dasypus kappleri
Dasypus sp. nov. FG
Dasypus novemcinctus
Cabassous unicinctus
Myrmecophaga tridactyla
Tamandua tetradactyla
Orycteropus afer
Proteles cristatus
Smutsia temminckii
Firmicutes
Actinobacteriota
Bacteroidota
Proteobacteria

Sample name	Species name	Common name	Order	Diet	Sex	Age	Wild / Captive	Country of origin	Location	Accession number	
										Illumina	ONT
CAB M2809	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925023	SRR23925022
CAB M2962	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925011	SRR23925000
CAB M3141	*Cabassous unicinctus*	Southern naked-tailed armadillo	Cingulata	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924989	SRR23924978
DASY M1746	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	NA	SRR23924967
DASY M1777	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Juvenile	Wild	French Guiana	coll. JAGUARS	SRR23924962	SRR23924961
DASY M2255	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	NA	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924960	SRR23925021
DASY M2865	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925020	SRR23925019
DASY M3021	*Dasypus novemcinctus sp FG*	Guianan long-nosed armadillo	Cingulata	Omnivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925018	SRR23925017
DASY VLD165	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Female	Adult	Wild	USA	Valdosta (GA)	SRR23925016	SRR23925015
DASY VLD168	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Male	Adult	Wild	USA	Valdosta (GA)	NA	SRR23925014
DASY VLD172	*Dasypus novemcinctus*	Nine-banded armadillo	Cingulata	Omnivorous	Female	Adult	Wild	USA	Valdosta (GA)	SRR23925013	SRR23925012
KAP M2867	*Dasypus kappleri*	Greater long-nosed armadillo	Cingulata	Insectivorous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925010	SRR23925009
MYR M5293	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23925008	SRR23925007
MYR M5295	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	NA	Juvenile	Wild	French Guiana	coll. JAGUARS	SRR23925006	SRR23925005
MYR ZOO	*Myrmecophaga tridactyla*	Giant anteater	Pilosa	Myrmecophagous	Female	Adulte	Captive	France	Montpellier zoo	SRR23925004	SRR23925003
ORY TS217	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23925002	SRR23925001
ORY TS513	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924999	SRR23924998
ORY TS566	*Orycteropus afer*	Aardvark	Tubulidentata	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924997	SRR23924996
PAN TS471	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924995	SRR23924994
PAN TS475	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924993	SRR23924992
PAN TS482	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924991	SRR23924990
PAN TS488	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924988	SRR23924987
PAN TS489	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924986	SRR23924985
PAN TS525	*Smutsia temminckii*	Ground pangolin	Pholidota	Myrmecophagous	NA	NA	Wild	South Africa	Tswalu Kalahari reserve	SRR23924984	SRR23924983
PRO TDR7	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924982	SRR23924981
PRO TDR17	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924980	SRR23924979
PRO TDR22	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924977	SRR23924976
PRO TDR49	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924975	SRR23924974
PRO TDR62	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924973	SRR23924972
PRO TDR67	*Proteles cristatus*	Southern aardwolf	Carnivora	Myrmecophagous	NA	NA	Wild	South Africa	Tussen-die-Riviere reserve	SRR23924971	SRR23924970
TAM M3075	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924969	SRR23924968
TAM M5331	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Male	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924966	SRR23924965
TAM M5584	*Tamandua tetradactyla*	Southern tamandua	Pilosa	Myrmecophagous	Female	Adult	Wild	French Guiana	coll. JAGUARS	SRR23924964	SRR23924963