Host phylogeny and diet shape gut microbiome and virome in wild small mammals of Gongga Mountain, China

Running title: Gut Microbiome and Virome in Small Mammals

Le-wei He[1,2,#], Rui-xiang Tang[1,#], Shao-ying Liu[3,#], Zhen-jie Zhang[1], Yi Li[1], Xu-ming Wang[3], Bi-song Yue[1], Zhen-xin Fan[1,2,*]

[1]Key Laboratory of Bioresources and Eco-Environment (Ministry of Education), College of Life Sciences, Sichuan University, Chengdu, Sichuan 610065, China

[2]The conservation of endangered wildlife key laboratory of Sichuan Province, College of Life Sciences, Sichuan University, Chengdu, Sichuan 610064, China

[3]Sichuan Academy of Forestry, Chengdu, Sichuan 610068, China

# These authors contributed equally to this work.

*Correspondence to:

Zhen-xin Fan, zxfan@scu.edu.cn

Emails for other authors:

Le-wei He: 2986500143@qq.com; Rui-xiang Tang: rbio_tang@foxmail.com; Shao-ying Liu: shaoyliu@163.com; Zhen-jie Zhang: zhangzhenjie_118@163.com; Yi Li: leeyee_art@163.com; Xu-ming Wang: wangxuming1994@163.com; Bi-song Yue: bsyue@scu.edu.cn

Science Press ZR Zoological Research

ABSTRACT

Gut microbiotas play pivotal roles in host adaptation, yet their composition and function in high-altitude small mammals remain poorly characterized. This study investigated how host phylogeny (order-level) and dietary habits shape the gut microbiome and virome of three mammalian orders (Eulipotyphla, Rodentia, Lagomorpha) in Gongga Mountain, a biodiversity hotspot on the Qinghai-Xizang Plateau. Metagenomic sequencing of 219 samples from 22 species revealed order-specific microbial signatures: Eulipotyphla (carnivorous) harbored higher abundances of potential pathogens (e.g., *Helicobacter*, *Hafnia*) and Retroviridae; Lagomorpha (herbivorous) was enriched in cellulolytic bacteria (e.g., Lachnospiraceae, *Prevotella*) and carbohydrate-active enzymes (CAZymes); Rodentia (omnivorous) showed intermediate traits. We reconstructed 1,385 high-quality metagenome-assembled genomes (MAGs), 1,328 representing novel species, and identified 749 viral operational taxonomic units (vOTUs), >80% being Caudoviricetes. Crucially, Retroviridae abundance in Eulipotyphla suggests zoonotic risk. Phage-host network analysis indicated Caudoviricetes regulates cellulolytic bacteria in Lagomorpha. Host phylogeny and diet jointly drive gut microbiome divergence in small mammals. We establish the first gut microbiome and virome resource of small mammals in the high-altitude area of Gongga Mountain, highlighting Eulipotyphla as a potential vector for zoonotic pathogens.

Keywords: Small mammals; Gut microbiota; MAGs; Virome

INTRODUCTION

Gongga Mountain is located between the southeastern of the Qinghai-Xizang Plateau and the Sichuan Basin, with altitudes ranging from 1,000m to 7,556m. It is one of the highest mountain ranges in southwestern China(Zu et al., 2019). The relative altitude difference exceeding 6,000m has endowed the Gongga Mountain with a vertical natural zone spectrum from subtropical to frigid zones and a complex biota(Zu et al., 2019). Previous studies have shown that the area where Gongga Mountain is located is widely recognized as one of the global biodiversity hotspots(Myers et al., 2000; Wu et al., 2013). Small mammals are considered to play important roles in terrestrial ecosystems worldwide(Searing et al., 2023). Their special ecological niches and their impacts on energy flow and material cycling in ecosystems are among their fundamental characteristics(Searing et al., 2023). In ecosystems, small mammals can act as predators of invertebrates and prey for medium-sized carnivores and raptors, greatly enhancing the complexity of food webs(Dorigo et al., 2021; Paniccia et al., 2022). In addition, their activities can also assist in vegetation regeneration and soil aeration(Dorigo et al., 2021; Paniccia et al., 2022).

Studies have shown that there are significant differences in morphology and feeding habits among small mammals of different orders. Eulipotyphla usually feed on invertebrates, and some species also supplement their diet with small vertebrates, so Eulipotyphla animals can be called carnivores(Symonds, 2005). Rodentia is the largest order of mammals worldwide, with more than 2,000 species classified into Rodentia, accounting for approximately 40% of all mammalian species(Kay & Hoekstra, 2008).

Studies on the feeding habits of Rodentia have shown that the adaptive evolution of the mandibular dental organs in rodents is closely related to their omnivorous diet(Landry, 1970). There is a distinct gap between the incisors and molars in Rodentia, which enables rodents to gnaw on almost anything, including grasses, seeds, and chitin-rich insects(Hunter & Jernvall, 1995; Kay & Hoekstra, 2008). Lagomorpha such as domestic rabbits, hares and pikas are herbivorous mammals(Iraçabal et al., 2024). Lagomorpha have highly specialized jaw morphology. They lack canines but possess sharp incisors specifically adapted for gnawing, and these incisors continuously grow(Zachos, 2016). Furthermore, the gut morphology among small mammals with different feeding habits is diversified. Herbivorous mammals require large fermentation chambers to digest plant fibers. Therefore, it is generally believed that the gut of herbivorous mammals are longer than those of carnivorous mammals(Duque-Correa et al., 2021). Most carnivorous mammals lack a cecum, while herbivorous mammals have a relatively well-developed cecum(Mcgrosky et al., 2019).

The gut microbiota is a diverse community composed of bacteria, fungi, protozoa, and viruses, which is crucial for maintaining healthy gut physiology in mammals(Puschhof et al., 2021). Unlike the genome of mammalian hosts, the gut microbiome is more plastic and can easily adapt to various external environments and host-derived stimuli(Zmora et al., 2019). Among various endogenous and exogenous factors, diet is a key determinant that influences the structure and function of the host gut microbial community(Groussin et al., 2017; Youngblut et al., 2019; Zmora et al., 2019). From studies on herbivores and carnivores, researchers have strongly inferred that there are huge differences in the gut

microbiome of mammals with different feeding habits(Zmora et al., 2019). Various mammalian lineages co-evolve with their gut microbiome, which distinguish different hosts through the hosts' dietary preferences rather than phylogeny. This phenomenon is specifically manifested as follows: from herbivores to omnivores and then to carnivores, the diversity of gut bacterial communities decreases, and typical microbial composition patterns emerge(Degregori et al., 2024b; Ley et al., 2008; Muegge et al., 2011; Youngblut et al., 2019). Metagenomic sequencing can randomly sequence the entire metagenome of a sample without specific primers, such as 16S rRNA sequencing(Fischbach & Segre, 2016; Wang et al., 2022). This approach provides information on the genetic composition and functional potential of the gut microbiota, thereby enabling a more detailed taxonomic characterization of microbial communities than marker gene-based methods by offering resolution at the species or even strain level(Gao et al., 2021; Schloissnig et al., 2013). Therefore, metagenomic sequencing has broad application prospects in studying the community structure and functions of gut microbiota in small mammals.

In this study, metagenomic sequencing was performed on the gut microbiota of 22 species of small mammals belonging to three different orders (Eulipotyphla, Rodentia, and Lagomorpha). Based on the metagenomic sequencing data, we explored the differences in species composition and functions among the three orders of small mammals in the Gongga Mountain. Additionally, the reconstruction of metagenome-assembled genomes (MAGs) and prediction of viromes were completed, with the aim of revealing the impact of different ordinal taxonomic categories of small mammals on the gut microbiota and providing the first gut microbiome and virome resource for high-altitude small mammals.

MATERIALS AND METHODS

Sample collection

All small mammal samples in this study were collected from Gongga Mountain, Sichuan Province, China. From September to October 2022, we captured a total of 219 small mammals by randomly placing cage-style small animal traps at the sampling sites and used ether to anesthetize all captured small mammals on the spot. Information such as altitude and habitat where the samples were located was recorded during sampling (Supplementary Figure S1; Supplementary Table S1). All anesthetized small mammal samples were transported to the laboratory for processing, and euthanasia was performed on them via cervical dislocation. The killing of all small mammals has obtained the permission from Sichuan Academy of Forestry Sciences. During the sampling process, we wore sterile gloves to cut out the middle section from the midgut of all small mammals and collect the contents. One sample was taken from the gut of each small mammal, with a total of 219 samples obtained. Total DNA was extracted from the intestinal contents of each small mammal using the Intestinal Genomic DNA Extraction Kit (Tiangen Biochemical Technology (China) Co., Ltd.), and the extraction procedure was performed in accordance with the kit's instruction manual. After extraction, the samples were immediately refrigerated with ice packs and stored in -80°C refrigerator until DNA extraction.

In this study, the collected small mammals included three orders (Eulipotyphla, Rodentia, and Lagomorpha). Eulipotyphla comprised 2 families, 7 genera, 9 species, and a total of 58 individuals. Rodentia included 3 families, 7 genera, 11 species, and a total of

125 individuals. Lagomorpha consisted of 1 family, 1 genus, 2 species, and a total of 36 individuals (Supplementary Table S1). The samples of Lagomorpha include only two species, which may be one of the limitations of this study.

Metagenomic sequencing and quality control of raw data

The total amount of DNA per metagenomic sample was 0.2 μg, serving as the input material for DNA library preparation. Sequencing was performed using the NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA, Catalog #: E7370L) according to the manufacturer's instructions. Metagenomic DNA samples were fragmented to a size of 350 bp by sonication. The ends of DNA fragments were repaired and ligated with full-length adapters for Illumina sequencing, followed by further PCR amplification. PCR products were purified using the AMPure XP system (Beverly, USA). Subsequently, the quality of the DNA libraries was evaluated on an Agilent 5400 system and quantified by QPCR (1.5 nM). According to the effective DNA library concentration and required data volume, qualified DNA libraries were pooled and sequenced using the PE150 strategy on the Illumina NovaSeq 6000 platform. Sequencing information for all metagenomic samples is presented in Table S2.

Raw fluorescent image files obtained from the Illumina platform were converted into Raw Reads through base calling and recorded in FASTQ file format. We used Fastp v.0.23.4(Chen et al., 2018) to perform quality control on the Raw Reads of all metagenomic samples. Due to the lack of assembled genome resources for the small mammal species involved in this study, we identified the close relatives of each small mammal species based on existing phylogenetic research on small mammals. We

downloaded the reference genomes of the close relatives of small mammal species from the National Center for Biotechnology Information (NCBI), covering a total of 13 small mammal species. The reference genomes used to remove host sequences for each small mammal species were shown in Supplementary Table S3. Finally, we aligned the Clean Reads to the reference genomes of different small mammal close relatives based on Bowtie2 v.2.4.5(Langmead & Salzberg, 2012), so as to remove the host genome sequences contained in the metagenomic sequences as much as possible.

Identification of taxa in metagenomes

Based on a standard microbial database containing "bacteria", "archaea", "viruses", "fungi", and "protozoa", we used kraken2 v.2.1.3(Wood et al., 2019) to identify and quantify the microorganisms in the intestinal metagenomic samples of all small mammals. The number of times Clean Reads were aligned to the microbial database was recorded as counts.

Reconstruction and quality assessment of MAGs

We used all contigs assembled from individual samples and those co-assembled within each species to generate Bins. First, the Clean Reads corresponding to the contigs assembled from each single sample were mapped back using Bowtie2 v.2.5.4(Langmead & Salzberg, 2012). The resulting sam files were converted into bam file format and sorted using SAMtools v.1.22.1(Danecek et al., 2021). For the co-assembly results, the mapping results of each sample need to be merged using "samtools merge", and an index is built for the bam file corresponding to each contig for subsequent binning processes. To obtain as many high-quality gut microbial genomes of small mammals as possible, we used

MetaBAT2(Kang et al., 2019), CONCOCT(Alneberg et al., 2014), and MaxBin(Wu et al., 2016) to perform binning on the mapped bam files based on contigs' coverage information and clustering information, further dividing contigs into potential genomes. Next, DAS Tool v.1.1.7(Sieber et al., 2018) was used to evaluate and integrate the assembly results of the above software with the parameter set as "--score_threshold 0", thereby obtaining a cluster-based set of Bins. In addition, we also used the deep learning-based MetaBinner v.1.4.4(Wang et al., 2023) to perform binning on contigs. This tool relies on a large number of known bacterial genome markers and uses a neural network model to predict the membership of each bin, improving the accuracy of final genome separation through large-scale simulation training. Finally, we obtained a cluster-based set of Bins and a prediction-based set of Bins, which were used as the initial Bins sets for further evaluation and filtering.

To obtain high-quality MAGs, we dereplicated the initial Bins sets using dRep v.3.4.2(Olm et al., 2017) with the parameters "dereplicate_wf -comp 70 -con 10 -sa 0.99 -nc 0.1 --S_algorithm fastANI". fastANI was used to evaluate the average nucleotide identity (ANI) between genomes(Jain et al., 2018). We used an ANI<99% as the threshold for identifying strains, and genomes with differences within this range were considered by us to belong to different strains. In addition, we evaluated the completeness and contamination of Bins by invoking CheckM v.1.2.4(Parks et al., 2015) through dRep. Finally, a non-redundant set of Bins with completeness greater than 70% and contamination less than 10% was obtained, which was used as the assembly result at the strain level.

198 **Taxonomic identification and abundance calculation of MAGs**

199 GTDB-Tk v.2.5.2 (Genome Taxonomy Database Toolkit)(Chaumeil et al., 2022a) is a
200 software package for the Taxonomic identification of bacterial and archaeal genomes
201 based on the Genome Taxonomy Database (GTDB)(Parks et al., 2020). In this study, we
202 used the "classify_wf" function of GTDB-Tk to perform microbial species annotation on
203 all MAGs, with the GTDB version being Release 08-RS214. The final taxonomic result
204 for each MAG is represented by the lowest taxonomic level obtained from species
205 annotation. We used the "de_novo_wf" of GTDB-Tk to construct a phylogenetic tree of
206 MAGs, and this process invoked FastTree(Price et al., 2009) to rapidly build the
207 phylogenetic tree.

208 **Mobile genetic element (MGE) prediction of MAGs**

209 We identified insertion sequences (IS) in MAGs using the Python-based tool ISEScan
210 v.1.7.3(Xie & Tang, 2017). Bacterial plasmid sequences from the PLSDB database were
211 downloaded, and MAGs were mapped to the bacterial plasmid sequence database using
212 diamond v.2.1.8 to predict plasmids in MAGs(Buchfink et al., 2015; Schmartz et al.,
213 2022). The Perl-based tool TransposonPSI v.1.0.0 (http://transposonpsi.sourceforge.net/)
214 was used to predict transposon sequences in MAGs.

215 **Identifying gut viruses of small mammals from metagenomes**

216 The contigs used for viral genome prediction were derived from the results of individual
217 assembly of Clean Reads from each metagenomic sample using MEGAHIT v.1.2.9(Li et
218 al., 2015), yielding contigs with lengths exceeding 300 bp. We used three tools to predict
219 viral genomes in the metagenome, namely DeepVirFinder v.1.0(Ren et al., 2020),

VirSorter2 v.2.2.4(Guo et al., 2021), and VIBRANT v.1.2.1(Kieft et al., 2020). We merged all viral sequences predicted by the three tools according to the assembly source of contigs, and used CD-HIT v.4.8.1(Fu et al., 2012) to dereplicate all viral sequence sets by source. Next, CheckV v.1.0.3(Nayfach et al., 2021a) was used to evaluate the genomic quality of non-redundant viral sequences in each sample. According to the quality assessment results of CheckV v.1.0.3(Nayfach et al., 2021a), viral sequences from individual samples were filtered, retaining only those with a length exceeding 5000 bp, a completeness of at least 50%, a contamination level of less than 5%, kmer_freq≤1.0, and a number of viral genes greater than that of host genes. After completing quality control, all high-quality viral genome sequences were merged, and then CD-HIT v.4.8.1(Fu et al., 2012) was used again to dereplicate all viral sequences, resulting in a non-redundant viral sequence set across all samples, designated as viral operational taxonomic units (vOTUs).

Taxonomic identification and abundance calculation of vOTUs

The taxonomic identification of vOTUs was primarily based on the ictv-mmseqs2-protein-database (https://github.com/apcamargo/ictv-mmseqs2-protein-database). This database provides instructions and scripts for constructing a viral protein MMseqs2 database with taxonomic annotations, which can be used to assign sequences to viral taxonomic groups defined by the International Committee on Taxonomy of Viruses (ICTV)(Siddell et al., 2023). CoverM (https://github.com/wwood/CoverM) is an easy-to-use tool capable of rapidly calculating DNA reads coverage and relative abundance, currently used in metagenomics research. We mapped the Clean Reads obtained from metagenomic sequencing and quality

Science Press ZR Zoological Research

control to vOTUs to calculate the abundance of each viral genome and obtain an abundance table for all vOTUs.

Comparison with viral sequences in other databases

We downloaded three known gut virome databases: (1) The Gut Virome Database (GVD)(Gregory et al., 2020), constructed from 2,697 viral particles or microbial metagenomes of 1,986 individuals from 16 countries; (2) The Gut Phage Database (GPD)(Camarillo-Guerrero et al., 2021), a collection of approximately 142,000 non-redundant viral genomes obtained by mining datasets of 28,060 human gut metagenomes and 2,898 cultured gut bacterial reference genomes; (3) The Metagenomic Gut Virus (MGV)(Nayfach et al., 2021b), which includes 189,680 viral genomes from 11,810 publicly available human fecal metagenomes. The vOTUs were pairwise compared with all vOTU sequences in the three gut virome databases using blastn. A vOTU sequence was considered a newly discovered viral genome if the average nucleotide identity (ANI) was <95% or the alignment fraction (AF) was <85%.

Host prediction and lifestyle prediction of Caudoviricetes vOTUs

We used CHERRY(Shang & Sun, 2022) to predict the microbial hosts of all vOTUs identified as phages and the MAG hosts of Caudoviricetes vOTUs. PhaTYP(Shang et al., 2023) was used to predict the lifestyle of Caudoviricetes vOTUs.

Gene assembly, functional prediction, and quantification

We individually assembled the Clean Reads of each metagenomic sample using MEGAHIT v.1.2.9(Li et al., 2015), with the parameter set as "--min-contig-len 300" to obtain contigs with lengths exceeding 300 bp. Prodigal v.2.6.3(Hyatt et al., 2010) was

Science Press ZR Zoological Research

used for gene prediction on the contigs of each sample. Redundant genes were removed using CD-HIT v.4.8.1(Fu et al., 2012) based on 95% similarity and 90% coverage. Finally, the nucleotide sequences were translated into amino acid sequences. We used DIAMOND v.2.1.8(Buchfink et al., 2015) to align non-redundant amino acid sequences against the Carbohydrate-Active enZYmes (CAZYmes) database and the virulence factor database (VFDB), to identify carbohydrate-active enzyme-related genes and virulence genes in the metagenome. Salmon v.1.10.1(Patro et al., 2017) was used to quantify the abundance of non-redundant genes in the metagenome, and the gene abundance results were normalized using Transcripts Per Million (TPM).

For all MAGs, we performed whole-genome annotation using Prokka v.1.14.5(Seemann, 2014) with the parameters "--metagenome --force --gcode 11". Combined with the taxonomic results of MAGs, the smallest taxonomic unit was specified for more accurate genome annotation. To investigate the abundance differences in gene functions of MAGs among small mammals of different orders, we performed gene function prediction on the gene prediction results of MAGs based on the CAZYmes database and VFDB. Before functional prediction, we first used CD-HIT v.4.8.1(Fu et al., 2012) to remove redundancy from the gene sequences of all MAGs, and then used Prodigal v.2.6.3(Hyatt et al., 2010) to predict the protein sequences of non-redundant gene sequences, thereby obtaining a set of non-redundant amino acid sequences for gene function prediction.

Data statistical analysis

We performed β-diversity analysis and permutational multivariate analysis of variance

Science Press ZR Zoological Research

using the vegan package in R, with the Bray-Curtis distance employed as the β-diversity metrics; the α-diversity of gut microbiota at the genus and species levels was calculated using the R package microeco v.1.7.1, with Shannon index specifically employed as the measurement metric(Liu et al., 2021). Linear discriminant analysis (LDA) effect size (LEfSe)(Segata et al., 2011) was used to identify gut microbial and functional biomarkers in small mammals of different orders. In this process, the Benjamini and Hochberg (BH) method was used for False Discovery Rate (FDR) correction, and the significance was evaluated based on the criteria of FDR<0.05 and LDA>2. The gut microbial co-occurrence networks of small mammals across different orders were constructed using the R packages phyloseq v.1.42.0(Mcmurdie & Holmes, 2013) and ggClusterNet v.0.1.0(Wen et al., 2022). Spearman's rank correlation test was used to calculate the correlation coefficients. Procrustes(Peres-Neto & Jackson, 2001) correlation was used to analyze the relationship between the composition (relative abundance, %) of Caudoviricetes vOTUs and their predicted MAGs hosts. The significance test method in this study is Wilcox.test.

Data visualization

Data visualization was completed using the R packages ggplot2 v.3.5.1, ggpubr v.0.6.0, ggvenn v.0.1.10, pheatmap v.1.0.12, amplicon v.1.19.0(Liu et al., 2023), and UpSetR v.1.4.0(Conway et al., 2017).

Science Press ZR Zoological Research

306 **Results**

307 **Composition and functions of gut microbiota in small mammals of different orders**

308 We first identified the factors with greater impacts on the β-diversity of gut microbiota in
309 small mammals through permutational multivariate analysis of variance (PERMANOVA).
310 The results showed that taxonomic units at various levels of small mammals had
311 significant effects on the β-diversity of gut microbiota, with the order-level taxonomic
312 units having the greatest impact (Supplementary Figure S2A). In addition, neither altitude
313 nor habitat type at the sampling sites emerged as a significant predictor of gut microbiota
314 variation in the small mammals, after accounting for other factors (Supplementary Figure
315 S2A). However, we acknowledge that the potential influence of altitude, particularly
316 through its correlation with shifts in diet, couldn't be fully disentangled in this study, as
317 noted in previous research(Degregori et al., 2024a).

318 We further analyzed the composition and relative abundance differences of gut
319 microbiota among small mammals of different orders. At the phylum level,
320 Pseudomonadota, Actinomycetota, Campylobacterota, and Bacillota were dominant in
321 Eulipotyphla, while Bacillota, Pseudomonadota, Bacteroidota, and Actinomycetota were
322 dominant in Rodentia and Lagomorpha (Figure 1A). LEfSe analysis at the phylum level
323 (LDA>3, FDR<0.05) showed that the relative abundances of Pseudomonadota,
324 Campylobacterota, Actinomycetota, etc., in Eulipotyphla were significantly higher than
325 those in Rodentia and Lagomorpha, while the relative abundances of Bacillota,
326 Bacteroidota, etc., in Lagomorpha were significantly higher than those in Eulipotyphla
327 and Rodentia (Supplementary Figure S2B; Supplementary Table S4).

Science Press ZR Zoological Research

328 At the genus level, dominant genera in Eulipotyphla included *Helicobacter*,
329 *Pseudomonas*, *Streptomyces*, *Ralstonia*, etc. Rodentia was primarily dominated by
330 *Pasteurella*, *Paenibacillus*, *Blautia*, *Clostridium*, while Lagomorpha was mainly
331 characterized by *Paenibacillus*, *Blautia*, *Alistipes*, *Prevotella*, etc. (Figure 1B). Alpha
332 diversity analysis at the genus level showed that both Shannon and Chao1 indices of
333 Lagomorpha were significantly higher than those of Eulipotyphla and Rodentia, and the
334 Shannon and Chao1 indices of Rodentia were significantly higher than those of
335 Eulipotyphla (Figure 1C, D). LEfSe analysis at the genus level (LDA>3, FDR<0.05)
336 identified *Pseudomonas*, *Hafnia*, *Ralstonia*, *Streptomyces*, *Campylobacter*, *Serratia*, etc.,
337 as marker microbial genera in Eulipotyphla; *Pasteurella* as a marker microbial genus in
338 Rodentia; and *Alistipes*, *Prevotella*, *Flavonifractor*, *Paenibacillus*, *Faecalibacterium*, etc.,
339 as marker microbial genera in Lagomorpha (Figure 1E; Supplementary Table S5).

340 At the species level, the dominant species in Eulipotyphla were *Ralstonia insidiosa*,
341 *Hafnia alvei*, *Helicobacter enhydrae*, *Helicobacter cholecystus*, etc. The main species in
342 Rodentia were *Pasteurella multocida*, *Paenibacillus larvae*, *Flavonifractor plautii*,
343 *Dysosmobacter welbionis*, etc. In Lagomorpha, the dominant species were *Paenibacillus*
344 *larvae*, *Flavonifractor plautii*, *D. welbionis*, *Lawsonibacter asaccharolyticus*, etc
345 (Supplementary Figure S2C). Alpha diversity analysis showed that the Shannon index of
346 Lagomorpha was significantly higher than those of Eulipotyphla and Rodentia, while there
347 was no significant difference in the Shannon index between Eulipotyphla and Rodentia
348 (Figure 1F). The differences in Chao1 index among the three orders were consistent with
349 the results at the genus level (Figure 1G). LEfSe analysis (LDA>3, FDR<0.05) showed

that *Hafnia alvei*, *Helicobacter enhydrae*, *R. insidiosa*, *Helicobacter cholecystus*, *Serratia liquetaciens*, and other species were marker microorganisms of Eulipotyphla. *Pasteurella multocida*, *Natranaerofaba carboxydovora*, *Acutalibacter muris*, and others were marker microorganisms of Rodentia. *Flavonifractor plautii*, *Paenibacillus larvae*, *L. asaccharolyticus*, *D. welbionis*, *Faecalibacterium prausnitzii*, and others were marker microorganisms of Lagomorpha (Figure 1H; Supplementary Table S6).

Analyzing the gut microbial co-occurrence networks of small mammals from different orders through metagenomic data could help us identify the core gut microbiota and infer the interactions between species. We selected the top 150 gut microbiota with the highest relative abundance from each of the Eulipotyphla, Rodentia, and Lagomorpha, and analyzed the co-occurrence networks of gut microbiota in small mammals from different orders. After normalizing the microbial data using the TMM (Trimmed Mean of M-values) method, we calculated the correlations between microbes. Microbial correlation data were filtered based on a correlation coefficient>0.8 and significance<0.05. The co-occurrence networks showed that only positive correlations existed among microbes in the co-occurrence networks of each of the three orders (Supplementary Figure S2D-F). We found that the gut microbiota co-occurrence network of Eulipotyphla contained the largest number of microbial phyla, with a total of 14 phyla, followed by Rodentia, and Lagomorpha had the fewest, with only 6 phyla (Supplementary Figure S2D-F). In the co-occurrence networks of gut microbiota in Lagomorpha and Rodentia, Bacillota and Bacteroidota played dominant roles, while Pseudomonadota and Campylobacterota were the main contributors in Eulipotyphla (Supplementary Figure S2D-F). Furthermore, the

Science Press ZR Zoological Research

372 species-level gut microbial biomarkers identified for each order in the LEfSe analysis all
373 occupied core positions in the microbial co-occurrence networks (Supplementary Figure
374 S2D-F).

375 Carbohydrate-active enzyme (CAZYmes) genes in mammalian gut microbiota play
376 important roles in host health and disease, we performed functional annotation and
377 abundance calculation of CAZYmes for all gut metagenomic samples of small mammals.
378 We compared the total abundance of CAZYmes among these three orders. The total
379 abundance of CAZYmes in Lagomorpha was significantly higher than that in Rodentia
380 and Eulipotyphla. (Figure S2G). We further compared the abundances of six major
381 categories of CAZYmes across different orders. The abundances of all six CAZYme
382 categories in Lagomorpha were significantly higher than those in Rodentia and
383 Eulipotyphla. We identified CAZYmes biomarkers for small mammals of different orders
384 through LEfSe analysis (LDA>3, FDR<0.05). Multiple GT family members, along with
385 GH56 and CE11, were identified as CAZYmes biomarkers for Eulipotyphla. Among them,
386 GT9 is primarily active in several heptosyltransferase families, GT51 in peptidoglycan
387 glycosyltransferase, and CE11 in UDP-3-O-acyl N-acetylglucosamine deacetylase. GT2,
388 GH13, and GH25 were identified as CAZYmes biomarkers for Rodentia, with GH13
389 mainly active in families such as α-glucosidase. The CAZYmes biomarkers for
390 Lagomorpha were predominantly GH family members, including GH3, GH43, and GH2,
391 which are mainly active in families such as xylosidase, arabinofuranosidase, and
392 glucosidase (Figure 2B; Supplementary Table S7).

393 The Virulence Factor Database (VFDB) provides major virulence genes from various

well-characterized bacterial pathogens. We performed functional prediction and abundance calculation of virulence genes for all gut metagenomic samples of small mammals. Using LEfSe analysis, we analyzed differences in the relative abundances of virulence genes among small mammals of different orders and identified virulence gene biomarkers for each order (LDA>3, FDR<0.05). Multiple virulence gene biomarkers in Eulipotyphla (VFG019550, VFG006522, VFG019549, VFG019551) are involved in encoding flagellin B (FlaB) (Figure S2H; Supplementary Table S8).

Analyzing the relative abundance correlation between gut microbial markers at the species level and functional markers using metagenomic data helps predict the microbial species executors of key functions. We performed correlation analyses between the species-level gut microbial biomarkers (LDA>3, FDR<0.05) of Eulipotyphla and its CAZYmes biomarkers (LDA>3.5, FDR<0.05), as well as between these microbial biomarkers and virulence gene biomarkers (LDA>3.5, FDR<0.05). In Eulipotyphla, the relative abundances of five *Helicobacter* species (*Helicobacter mustelae*, *H. canis*, *H. pylori*, *H. cholecystus*, *H. enhydrae*) were significantly positively correlated with the relative abundances of GT9, CE11, GT30, and GT19 (Figure 2C). GT30 is primarily active in α-3-deoxy-D-manno-octulosonic-acid (KDO) transferase, while GT19 is mainly active in Lipid-A-disaccharide synthase. In addition, the relative abundances of the above five species were significantly positively correlated with the relative abundances of four virulence genes (VFG019550, VFG006522, VFG019549, VFG019551) encoding flagellin B (Figure 2D). At the species level, gut microbial biomarkers (LDA>3, FDR<0.05) in Rodentia showed overall significant positive correlations with CAZYmes biomarkers

(LDA>3.5, FDR<0.05). β-glucosidase is one of the active families in GH1, and GH19 contains classes I, II, and IV chitinases (Figure 2E). Correlation analysis between species-level gut microbial biomarkers (LDA>3, FDR<0.05) and CAZYmes biomarkers (LDA>3.5, FDR<0.05) in Lagomorpha showed that the relative abundances of species such as *Flavonifractor plautii* and *Faecalibacterium prausnitzii* were significantly positively correlated with the relative abundances of GH2 and GH3 (Figure 2F).

Reconstruction of MAGs in gut metagenomes of small mammals

During the genome assembly of gut microbiota from small mammals, a total of 13,259 Bins were assembled from 219 samples. To obtain strain-level MAGs, we set thresholds of ANI<99%, completeness≥70%, and contamination<10% to screen out 1,385 high-quality non-redundant MAGs. Among them, 229 were derived from co-assembled bins of small mammal species. The standard for high-quality draft genomes(Bowers et al., 2017) is defined as genome completeness≥90% and contamination≤5%. Among all 1,385 MAGs, 634 MAGs met the criteria for high-quality draft genomes. In addition, 408 MAGs had a completeness of over 95%, 1,147 MAGs had a contamination level below 5%, and 3 MAGs achieved 100% completeness with 0% contamination (Figure S3a). We used GTDB-Tk(Chaumeil et al., 2022b) to perform microbial taxonomic identification on the 1,385 MAGs. The taxonomic identification results of MAGs adopted the standard taxonomic labels of GTDB(Parks et al., 2020). Only 1 MAG was identified as archaeal genome, while 1,384 MAGs were identified as bacterial genomes. The MAGs identified as bacterial genomes were assigned to 12 phyla and 16 classes. Among them, 1,383 MAGs were classified into 32 orders, 1,374 MAGs into 52 families, 1,325 MAGs into 163 genera,

438 and only 56 MAGs into 34 species (Supplementary Table S9). We constructed a
439 phylogenetic tree of bacterial MAGs using the approximate maximum likelihood method
440 via FastTree called by GTDB-Tk. The results showed that the phylogenetic relationships
441 among MAGs were consistent with the annotation results at the bacterial phylum level
442 (Figure 3A). Among the MAGs identified as bacterial genomes, a total of 828 MAGs were
443 annotated as Bacillota, accounting for approximately 59.83% of all MAGs
444 (Supplementary Table S9).

445 We compared the composition and abundance of gut microbiota across different
446 orders of small mammals at the bacterial genome level. In Eulipotyphla, the dominant
447 phyla were Pseudomonadota and Campylobacterota; in Rodentia, the dominant phyla were
448 Bacillota_A and Bacteroidota; and in Lagomorpha, the dominant phyla were Bacillota_A,
449 Bacteroidota, and Spirochaetota (Figure 3B). LEfSe analysis at the phylum level (LDA>2,
450 FDR<0.05) showed that the relative abundances of Pseudomonadota and
451 Campylobacterota in Eulipotyphla were significantly higher than those in Rodentia and
452 Lagomorpha, while the relative abundances of Bacteroidota, Spirochaetota, etc., in
453 Lagomorpha were significantly higher than those in Eulipotyphla and Rodentia (Figure
454 S3b). At the genus level, the dominant genera in Eulipotyphla include *Helicobacter_G*,
455 *Hafnia*, *Serratia_J*, and *Ralstonia*. In Rodentia, the main genera were *UBA3282*, *COE1*,
456 and *1XD8-76* (genera within Lachnospiraceae). In Lagomorpha, the predominant genera
457 were *UBA3282* (a genus within Lachnospiraceae), *Acetatifactor*, *Treponema_D*,
458 *Enterenecus*, and *Prevotella* (Figure 3C). The results of α-diversity analysis were close to
459 those based on metagenomics (Supplementary Figure S3C, D). LEfSe analysis at the

genus level (LDA>3, FDR<0.05) showed that *Helicobacter_G*, *Hafnia*, *Serratia_J*, *Ralstonia*, and others were marker genera for Eulipotyphla; *UBA3282*, *1XD8-76*, *COE1*, *Pelethomonas*, and others were marker genera for Rodentia; and *Treponema_D*, *Acetatifactor*, *Prevotella*, *Enterenecus*, and others were marker genera for Lagomorpha (Figure 3D; Supplementary Table S10). LEfSe analysis of MAGs (LDA>3, FDR<0.05) showed that MAG1225 (a member of the *Hafnia*), *Serratia_J liquefaciens_A*, *Ralstonia sp001078575*, MAG1147 (a member of the genus *Helicobacter_G*), and others were marker MAGs for Eulipotyphla. Members of the Lachnospiraceae (MAG499, MAG121, MAGA479) and MAG1079 (a member of the genus *Alistipes*) were marker MAGs for Rodentia. MAG677 (a member of the *Enterenecus*), MAG145 (a member of *Acetatifactor*), and others were marker MAGs for Lagomorpha (Supplementary Figure S3E; Supplementary Table S11).

We performed CAZYmes functional prediction on the MAGs of gut microbiota from all small mammals. The total abundance of CAZYmes and the total abundance of each major category among different orders were relatively close to the results of metagenomics-based analysis (Figure 3E; Supplementary S3F). LEfSe was used to analyze the differences in the relative abundances of CAZYmes-related genes among MAGs of small mammals from different orders (LDA>3, FDR<0.05). GT9, GT51, CE11, and others were identified as CAZYmes markers with significant differences in Eulipotyphla, consistent with the results of metagenomics-based analysis; GT4, GH13, GT8, GT2, and others were identified as CAZYmes markers in Rodentia; the CAZYmes markers with significant differences in Lagomorpha were mainly members of the GH

Science Press ZR Zoological Research

family, including GH77, GH3, GH2, GH25, and GH5 (Figure 3F; Supplementary Table S12).

We performed virulence gene prediction for all MAGs. PCoA analysis showed that the distribution of virulence genes in Lagomorpha was significantly different from that in Rodentia and Eulipotyphla, while the distribution of virulence genes in Rodentia and Eulipotyphla partially overlapped (Supplementary Figure S3G). Further LEfSe analysis was conducted to examine the differences in the relative abundances of virulence genes among small mammals from different orders. The virulence gene markers (VFG006522, VFG006524, VFG019549, VFG019550, VFG019551) in Eulipotyphla MAGs were involved in encoding flagellin B (FlaB) (Supplementary Figure S3h; Supplementary Table S13).

To infer the potential functions of specific microbial genomes and validate the results of metagenomics-based association analysis, we performed correlation analysis between the relative abundance of high-quality MAGs and the relative abundance of CAZyme and VFG families. We performed correlation analyses between the gut microbiota MAG markers (LDA>3, FDR<0.05) of Eulipotyphla and its CAZYmes markers (LDA>3.5, FDR<0.05) as well as virulence gene markers (LDA>3.5, FDR<0.05), respectively. In Eulipotyphla, the relative abundances of multiple members of the *Helicobacter_G* (MAG1159, MAG1156, MAG1144, MAG1143, MAG1145, MAG1148, MAG1146, MAG1154, MAG1153) were significantly positively correlated with the relative abundances of GT51 and CE11 (Figure 3G). Furthermore, the relative abundances of multiple members of the genus *Helicobacter_G* in Eulipotyphla (MAG1149, MAG1159,

Science Press ZR Zoological Research

MAG1156, MAG1144, MAG1143, MAG1145, MAG1148, MAG1146, MAG1154, MAG1153) were significantly positively correlated with the relative abundances of one gene related to chemotaxis response regulator (VFG043386) and four genes encoding flagellin B (VFG006522, VFG006524, VFG019549, VFG019550) (Figure 3H). Correlation analysis between the MAG markers (LDA>3, FDR<0.05) of Rodentia and its CAZYmes markers (LDA>3.5, FDR<0.05) showed significant positive correlations between members of the Lachnospiraceae family (such as MAG395, MAG302, MAG551) in the MAG markers and GH13 and GH94 in the CAZYmes markers (Supplementary Figure S3I). GH94 is mainly active in families such as cellobiose phosphorylase, cellodextrin phosphorylase, and cellobionic acid phosphorylase. Correlation analysis between the MAG markers (LDA>3, FDR<0.05) of Lagomorpha and its CAZYmes markers (LDA>3.5, FDR<0.05) revealed significant positive correlations between members of the Lachnospiraceae family (such as MAG504, MAG536, MAG228) in the MAG markers and one or more members of GH2, GH3, and GH5 in the CAZYmes markers (Supplementary Figure S3J).

Comparison of *Helicobacter* MAGs and cellulolytic bacteria MAGs from different sources

In the taxonomic identification results of MAGs, MAGs classified as *Helicobacter* and derived from Eulipotyphla were annotated as *Helicobacter_G*, while those classified as *Helicobacter* and derived from Rodentia and Lagomorpha were annotated as *Helicobacter_C* and *Helicobacter_D*, respectively (Supplementary Table S9). To compare the differences in *Helicobacter* MAGs from different sources at the genomic level, we

526 performed comparative genomic analysis on 62 *Helicobacter* MAGs assembled in this
527 study and 8 *Helicobacter* MAGs downloaded from the integrated mouse gut metagenome
528 catalog (iMGMC). First, we constructed a phylogenetic tree of the 70 *Helicobacter* MAGs
529 described above. All 20 MAGs from Eulipotyphla were annotated as *Helicobacter_G* and
530 clustered into an independent branch (Figure 4A). This indicated that they have a
531 relatively distant phylogenetic relationship with *Helicobacter* MAGs from Rodentia,
532 Lagomorpha, and laboratory mice. In addition, the *Helicobacter* MAGs from Rodentia
533 and Lagomorpha all belong to two genera, namely *Helicobacter_C* and *Helicobacter_D*
534 respectively (Figure 4A). They also clustered into independent branches almost entirely
535 based on their sources (Figure 4A). To compare the gene functional differences of
536 *Helicobacter* MAGs from different sources, we further performed predictions of
537 CAZYmes and virulence genes for *Helicobacter* MAGs. In the gene function prediction
538 results of CAZYmes, the CAZYmes markers (GT9, GT51, and CE11) in Eulipotyphla
539 identified based on metagenomics and MAGs were annotated in all *Helicobacter* MAGs
540 (Figure 4B). However, in terms of the overall functional prediction results of CAZYmes,
541 *Helicobacter* MAGs from Eulipotyphla predicted fewer CAZYmes than those from
542 Rodentia, Lagomorpha, and laboratory mice. For example, GH73 was only minimally
543 detected in *Helicobacter* MAGs of Eulipotyphla, while it was abundantly detected in other
544 small mammals (Figure 4B). In addition, β-diversity analysis revealed that the
545 *Helicobacter* MAGs from Eulipotyphla exhibit significant differences in the function of
546 CAZymes compared to *Helicobacter* MAGs from other sources (Supplementary Figure
547 S4A). In the virulence gene prediction results, flagellin-related genes such as filE, fliS,

flgG_2, fliM, flgC, and fliP were detected in *Helicobacter* MAGs of most small mammals (Figure 4C). However, some flagellin-related genes could not be detected in all *Helicobacter* MAGs of small mammals. For example, fliN, fliG, and fliQ were only minimally detected in Eulipotyphla, while flaB and fliA were only minimally detected in Rodentia. In terms of the overall prediction of virulence genes, fewer virulence genes were detected in *Helicobacter* MAGs from Eulipotyphla compared to those from Rodentia, Lagomorpha, and laboratory mice. β-diversity analysis also revealed distinct differences in the function of virulence genes between *Helicobacter* MAGs from Eulipotyphla and those from other sources (Supplementary Figure S4B). Therefore, the MAGs of *Helicobacter* in Eulipotyphla show obvious differences in gene function from those of *Helicobacter* in other small mammals. Comparative analysis of the relative abundance differences of *Helicobacter* MAGs (including *Helicobacter_G*, *Helicobacter_C*, and *Helicobacter_D*) among the three orders revealed that the relative abundance of *Helicobacter* genomes in the gut of Eulipotyphla was the highest. This result suggests that invertebrate diet drives the greater colonization of *Helicobacter* in the gut of Eulipotyphla and the exertion of its virulence functions (Figure 4D). We further predicted three types of mobile genetic elements (insertion sequences, plasmids, transposons) in *Helicobacter* MAGs. The association between mobile genetic elements and virulence genes in *Helicobacter* MAGs was assessed by calculating the ratio of virulence genes that were co-located with mobile elements to the total number of virulence genes in each genome. Analysis revealed distinct patterns of association between MGE types and virulence genes across different host orders. In *Helicobacter* MAGs from Rodentia and Lagomorpha, virulence genes showed

570 stronger association with plasmids than with other MGEs. In contrast, virulence genes in
571 *Helicobacter* MAGs from Eulipotyphla were more frequently associated with transposons.
572 In *Helicobacter* MAGs from laboratory mice, virulence genes were primarily linked to
573 both plasmids and transposons. IS showed relatively weak associations with virulence
574 genes across *Helicobacter* MAGs from the gut of small mammals (Supplementary Figure
575 S4C).

576 A large number of cellulolytic bacteria genomes from Rodentia and Lagomorpha were
577 present in our MAGs, including those from the families Lachnospiraceae, and the genera
578 Treponema and Prevotella. As a research hotspot among ruminants, the genomic resources
579 of goat gut microbiota have been extensively explored, and numerous genomes of
580 cellulolytic bacteria have been reconstructed. To compare the differences in genomes of
581 cellulolytic bacteria from different sources, we downloaded MAGs of cellulolytic bacteria
582 from the goat gut microbial genomes assembled by Xuefeng Peng et al.(Peng et al., 2021)
583 and compared them with MAGs of cellulolytic bacteria from Rodentia and Lagomorpha. A
584 total of 523 Lachnospiraceae MAGs were assembled from the metagenomic data, of
585 which 419 were from Rodentia and 104 were from Lagomorpha. We constructed a
586 phylogenetic tree of Lachnospiraceae MAGs from the intestines of small mammals and 51
587 Lachnospiraceae MAGs from goat intestines. The results showed that the Lachnospiraceae
588 MAGs did not completely cluster according to their sources (Figure 4E). To compare the
589 gene functions of Lachnospiraceae MAGs from different sources, we detected the
590 CAZYmes of the 574 Lachnospiraceae MAGs described above and performed gene
591 function prediction based on the Uniref50 database. In the CAZYmes detection results, the

Science Press ZR Zoological Research

heatmap showed that the detection results of Lachnospiraceae MAGs from different sources were relatively close in CAZYmes (Supplementary Figure S4D). In the gene function prediction results based on the Uniref50 database, the NMDS plot showed that the differences in gene function levels among these Lachnospiraceae MAGs were not obvious (Figure 4F). In the analysis of gut microbial relative abundance differences based on MAGs, we found that *Treponema* is a marker genus in Lagomorpha. Among the 121 *Treponema* MAGs we reconstructed, 63 were from Lagomorpha and 58 were from Rodentia. We constructed a phylogenetic tree of the above *Treponema* MAGs and three *Treponema* MAGs from goat gut. The results showed that the *Treponema* MAGs did not completely cluster according to their sources (Supplementary Figure S4E). We further compared the CAZYmes detection results of *Treponema* MAGs from different sources. The heatmap showed that the CAZYmes detection results of *Treponema* MAGs from different sources were relatively close (Supplementary Figure S4F). We reconstructed 16 and 19 *Prevotella* MAGs from Rodentia and Lagomorpha, respectively. We constructed a phylogenetic tree of the above 35 *Prevotella* MAGs and 5 *Prevotella* MAGs from goat gut. The results showed that the *Prevotella* MAGs did not completely cluster according to their sources (Supplementary Figure S4G). We further compared the CAZYmes detection results of *Prevotella* MAGs from different sources, and the CAZYmes detection results of *Prevotella* MAGs from different sources were relatively close (Supplementary Figure S4H).

Prediction and Characteristics of Gut Virome in Small Mammals

We integrated three viral sequence prediction tools (DeepVirFinder(Ren et al., 2020),

VirSorter2(Guo et al., 2021), VIBRANT(Kieft et al., 2020)) and predicted a total of 52,768 potential non-redundant viral sequences from all samples. We evaluated the quality of all non-redundant viral sequences using CheckV(Nayfach et al., 2021a), retaining those with more viral genes than host genes, contig length exceeding 5,000 bp, completeness greater than 50%, contamination less than 5%, and kmer_freq≤1.0. Finally, we obtained 749 high-quality gut viral operational taxonomic units (vOTUs) from small mammals. The average length of vOTUs was 37,021.88 bp, and the median length was 26,237 bp. Approximately 36% of vOTUs were complete or high-quality (Supplementary Figure S5A, B). When compared with three existing gut virome databases (GVD, GPD, MGV), a total of 150 vOTUs could be identified by alignment with any one of the three existing human gut virome databases. Among them, 50 vOTUs could be simultaneously aligned with all three existing human gut virome databases, and 599 vOTUs did not cluster with any viral genomes from other databases, indicating that they might be potential novel gut vOTUs (Figure 5A).

We performed taxonomic identification on the 749 vOTUs using the ictv-mmseqs2-protein-database in combination with multiple tools. A total of 722 vOTUs were taxonomically identified at least at the domain level, while 27 vOTUs could not be assigned to any viral domain (Supplementary Table S14). The 722 vOTUs identified at the domain level included 711 kingdoms, 711 phyla, 710 classes, 111 orders, 119 families, 24 genera and 4 species (Supplementary Table S14). In the taxonomic classification of vOTUs at the domain level, the vast majority of vOTUs were classified as Duplodnaviria, accounting for approximately 80.91% of all vOTUs (Figure 5B; Supplementary Table

S14). In our classification of vOTUs, the Duplodnaviria included only one phylum (Heunggongvirae), one class (Uroviricota), and one order (Caudoviricetes). At the order level classification of vOTUs, the vast majority of vOTUs could not be identified to the order level (85.18%). The two most identified orders were Tubulavirales and Ortervirales, containing 49 and 44 vOTUs, respectively (Supplementary Figure S5C; Supplementary Table S14). At the family level classification of vOTUs, the vast majority of vOTUs could not be identified to the family level (84.11%). The two most identified families were Inoviridae and Retroviridae, containing 49 and 44 vOTUs, respectively (Figure S5D; Supplementary Table S14). Notably, all vOTUs classified into the Retroviridae did not form clusters with any viral sequences in existing virome databases, suggesting that all Retroviridae vOTUs identified from the gut metagenomes of small mammals are potentially novel viruses (Supplementary Table S14).

We compared the composition and abundance of gut viromes among small mammals of different orders. At the class level, Caudoviricetes played an important role in the gut virome of Lagomorpha, with significantly higher relative abundances than those in Rodentia and Eulipotyphla (Figure 5C, D). Revtraviricetes were dominant in the gut viromes of Eulipotyphla and Rodentia, with significantly higher relative abundances than those in Lagomorpha (Figure 5C, D). At the order level, Ortervirales played a dominant role in the gut virome of Eulipotyphla; in Rodentia, the dominant orders were Ortervirales and Tubulavirales (Figure 5E, F), while in Lagomorpha, the dominant orders were Tubulavirales and Petitvirales (Figure 5E, F). Ortervirales had the highest relative abundance in Eulipotyphla, and its relative abundance in Rodentia was significantly

658 higher than that in Lagomorpha. Petitvirales and Tubulavirales had the highest relative
659 abundances in Lagomorpha, and their relative abundances in Rodentia were significantly
660 higher than those in Eulipotyphla (Figure 5E, F). At the family level, Retroviridae
661 dominated the gut virome of Eulipotyphla (Figure 5G). In Rodentia, the dominant families
662 were Retroviridae, Inoviridae, and Adenoviridae (Figure 5G), while in Lagomorpha, the
663 dominant families were Inoviridae, Herelleviridae, and Microviridae (Figure 5G).
664 Retroviridae had the highest relative abundance in Eulipotyphla, and its relative
665 abundance in Rodentia was significantly higher than that in Lagomorpha. The relative
666 abundances of Inoviridae in both Lagomorpha and Rodentia were significantly higher than
667 those in Eulipotyphla, with no significant difference between Lagomorpha and Rodentia.
668 Herelleviridae and Microviridae had the highest relative abundances in Lagomorpha
669 (Figure 5H).

670 We performed GO functional prediction for all gut vOTUs of small mammals. PCoA
671 analysis showed that there was partial overlap in the distribution of GO functions among
672 the vOTUs of the three orders (Supplementary Figure S5E). We further compared the
673 abundance differences of GO function-related genes among small mammals of different
674 orders through LEfSe analysis. In Eulipotyphla, the relative abundances of genes
675 associated with GO functions related to symbiosis and interaction (GO:0044419,
676 GO:0044403, GO:0051701) were significantly higher than those in Rodentia and
677 Lagomorpha (Figure S5f). In Lagomorpha, the relative abundances of genes associated
678 with GO functions related to icosahedral capsid (GO:0019030, GO:0039617, GO:0046729)
679 were significantly higher than those in Eulipotyphla and Rodentia (Supplementary Figure

Science Press ZR Zoological Research

S5F).

Interactions between gut Caudoviricetes and microbial hosts in small mammals

Given the important role of the class Caudoviricetes in the gut virome of small mammals, we analyzed the differences in Caudoviricetes vOTUs among small mammals of different orders. We first predicted the lifestyle of all vOTUs classified as Caudoviricetes using PhaTYP(Shang et al., 2023). The results showed that among 606 Caudoviricetes vOTUs, 239 were virulent phages, 366 were temperate phages, and only 1 vOTU's lifestyle could not be predicted (Supplementary Table S15). We further analyzed the differences in the relative abundances of Caudoviricetes vOTUs by lifestyle among small mammals of different orders. In Eulipotyphla, the relative abundance of virulent Caudoviricetes was significantly higher than that of temperate Caudoviricetes. In Rodentia, the pattern was the opposite of that in Eulipotyphla. In Lagomorpha, there was no significant difference in the relative abundances between the two lifestyles of Caudoviricetes (Figure 6A).

The prediction of microbial hosts for Caudoviricetes vOTUs was achieved by CHERRY(Shang & Sun, 2022) (Supplementary Table S16). We further analyzed the host composition of Caudoviricetes and the corresponding abundance differences of Caudoviricetes vOTUs among small mammals of different orders. We first analyzed the phylum-level hosts of Caudoviricetes vOTUs, and the most dominant hosts were Bacillota, Pseudomonadota, and Bacteroidota (Figure 6B; Supplementary Table S16). There were also significant differences in the phylum-level host prediction results of Caudoviricetes vOTUs among small mammals of different orders. The phylum-level hosts of Caudoviricetes in Eulipotyphla were mainly Bacillota, Pseudomonadota, and

Science Press ZR Zoological Research

Cyanobacteriota (Figure 6B); those in Rodentia were primarily Bacillota, Pseudomonadota, Bacteroidota, and Cyanobacteriota (Figure 6B); while in Lagomorpha, the main hosts were Bacillota, Pseudomonadota, Bacteroidota, Actinomycetota, and Spirochaetota (Figure 6B). We further analyzed the differences in the genus-level host composition of Caudoviricetes and the corresponding abundance differences of Caudoviricetes vOTUs among small mammals of different orders. At the genus level, the hosts of Caudoviricetes vOTUs in Eulipotyphla were primarily *Streptococcus*, *Cylindrospermopsis*, *Paenibacillus*, and *Veillonella* (Figure 6C). In Rodentia, the main hosts were *Enterocloster*, *Clostridium*, *Fischerella*, and *Selenomonas* (Figure 6C). In Lagomorpha, the dominant hosts were *Paenibacillus*, *Streptococcus*, *Acinetobacter*, and *Treponema* (Figure 6C). LEfSe analysis (LDA top 20, FDR<0.05) showed that in Eulipotyphla, the relative abundances of Caudoviricetes vOTUs with genus-level host predictions of *Veillonella*, *Rhodoferax*, and *Aeromonas* were significantly higher than those in Rodentia and Lagomorpha (Figure 6D). In Rodentia, the relative abundances of Caudoviricetes vOTUs with genus-level host predictions of *Enterocloster*, *Fischerella*, *Selenomonas*, *Lawsonibacter*, *Alicyclobacillus*, and *Microcystis* were significantly higher than those in Eulipotyphla and Lagomorpha (Figure 6D). In Lagomorpha, the relative abundances of Caudoviricetes vOTUs with genus-level host predictions of *Paenibacillus*, *Prevotella*, *Treponema*, and *Streptomyces* were significantly higher than those in Eulipotyphla and Rodentia (Figure 6D).

Given the potential impact of gut Caudoviricetes diversity in mammals on the gut microbiota, we investigated the correlation between the diversity of gut Caudoviricetes

Science Press ZR Zoological Research

and the diversity of gut microbial MAGs in small mammals. At the species level, the α-diversity of Caudoviricetes vOTUs and MAGs in Rodentia and Lagomorpha was significantly positively correlated, while there was no significant correlation between the α-diversity of Caudoviricetes vOTUs and MAGs in Eulipotyphla (Figure 6E). We compared the distributions of gut Caudoviricetes vOTUs and MAGs based on Bray-Curtis distances through Procrustes analysis. Although the distributions of gut Caudoviricetes vOTUs and MAGs showed significant correlations across all three orders, the correlation was relatively low in Eulipotyphla (r=0.3722419; Supplementary Figure S5G), while higher correlations were observed in Rodentia (r=0.8117904; Supplementary Figure S5H) and Lagomorpha (r=0.7251562; Supplementary Figure S5I). We used the CHERRY(Shang & Sun, 2022) to predict the MAGs hosts of Caudoviricetes vOTUs, thereby investigating the interactions between gut Caudoviricetes and MAGs in small mammals. A total of 423 vOTUs were found to interact with MAGs (Supplementary Table S17). We further investigated the correlation between the relative abundances of Caudoviricetes vOTUs and their MAGs hosts in small mammals. In Eulipotyphla, the correlation between temperate Caudoviricetes vOTUs and their MAGs hosts was stronger than that between virulent Caudoviricetes vOTUs and their MAGs hosts (Figure 6F). In Rodentia and Lagomorpha, the correlations between the two types of Caudoviricetes vOTUs and their MAGs hosts were comparable (Figure 6G, H). Additionally, Rodentia had the largest number and variety of MAGs hosts significantly correlated with Caudoviricetes vOTUs (n=324, 19 families, 65 genera; Figure 6I; Supplementary Table S18), while Eulipotyphla had the lowest number and variety of such MAGs hosts (n=45,

Science Press ZR Zoological Research

11 families, 23 genera; Supplementary Figure S5J; Supplementary Table S19). In Rodentia and Lagomorpha, Caudoviricetes showed significant positive correlations with MAGs of cellulose-decomposing bacteria such as Lachnospiraceae and *Treponema_D* (Figure 6I, J; Supplementary Table S18, 20).

Discussion

The structural characteristics of gut microbial communities in mammals of different orders can reflect the shaping effects of host phylogeny, physiological traits, ecological niche, feeding habits, etc., on the gut microbiota. Studies based on 16S rRNA sequencing have shown that both diet and phylogeny of mammals can influence the diversity of gut microbiota. Multiple studies have shown that as mammals transition from carnivory to omnivory and then to herbivory, the diversity of their gut microbiota gradually increases, and gut microbial communities and mammalian hosts diversify jointly(De Jonge et al., 2022; Ley et al., 2008; Xie et al., 2024). This phenomenon could be related to the more complex carbon source of plant-based diets, which requires a more diverse range of microorganisms to participate in the hydrolysis of carbon sources(Bayané & Guiot, 2011). Although omnivores have a mixed diet, the herbivore digestive system is specifically adapted (via gut morphology, residence time, fermentation chambers) for microbial breakdown of complex plant matter, thereby fostering a more diverse microbiome(Huang et al., 2025). By contrast, omnivores may rely more on host digestive enzymes and less on extensive microbial fermentation, thereby constraining microbial niche diversity.

We investigated the gut microbial community characteristics of small mammals of different orders in the Gongga Mountain based on metagenomic sequencing, MAGs reconstruction, and virome prediction(Chen et al., 2021; Lou et al., 2024; Qin et al., 2010; Wang et al., 2019; Zhou et al., 2022). We found that as the proportion of plants in the diet increases, the species diversity of gut microbiota in small mammals gradually increases. This phenomenon may be consistent with previous research results on the gut microbial

773 diversity of mammals with different feeding habits(Ley et al., 2008). The past "Glires"
774 hypothesis and phylogenetic studies have suggested that Rodentia and Lagomorpha share
775 a close evolutionary relationship(Nishihara et al., 2006). The convergence of gut microbial
776 composition between the two orders reflected the influence of phylogenetic relationships
777 of small mammals on their gut microbiota.

778 The gut of Eulipotyphla showed higher relative abundances of bacteria carrying
779 virulence-associated genes, as well as Retroviridae. This may be associated with the
780 feeding behavior of Eulipotyphla, which primarily feeds on invertebrates and occasionally
781 consumes carrion(He et al., 2023; Lee et al., 2023; Silby et al., 2011). In the gut
782 metagenome of Eulipotyphla, CAZYmes with higher abundances are primarily associated
783 with the synthesis of lipopolysaccharides (LPS) and peptidoglycan(Cote & Taylor, 2017;
784 Dahmane et al., 2018; De Falco et al., 2015; Fujita et al., 2022). Members of the
785 *Helicobacter* in the gut of Eulipotyphla showed a significant positive correlation with
786 genes involved in regulating the synthesis of LPS and flagellin B. LPS is a major
787 component of the cell surface of most Gram-negative bacteria and has been shown to play
788 important roles in cell motility, gut colonization, biofilm formation, and the exertion of
789 virulence by pathogenic bacteria(Kaito et al., 2020; Raetz & Whitfield, 2002). LPS has
790 been demonstrated to be one of the virulence factors produced by *Helicobacter pylori*,
791 playing a critical role in host-microbe interactions and the establishment of chronic
792 infections(Rubin & Trent, 2013). Flagellin B is one of the components of the flagellar
793 filament, and strains lacking the gene (*flaB*) encoding flagellin B exhibit insufficient
794 motility(Eletto et al., 2021; Josenhans et al., 1995). Certain *Helicobacter sp.* require

Science Press ZR Zoological Research

795 flagella for host cell colonization and infection(Ottemann & Lowenthal, 2002). Therefore,
796 *Helicobacter sp.* detected in the gut of Eulipotyphla may possess the genetic potential to
797 support virulence mechanisms, such as the synthesis of LPS and flagellin B.

798 Through comparative genomic analysis, we further revealed the high specificity of
799 *Helicobacter* in the gut of Eulipotyphla. A diet dominated by invertebrates such as insects
800 leads to the colonization of more *Helicobacter sp.* in their gut and the exertion of stronger
801 pathogenic effects. *Hafnia*(Zhang et al., 2019) and *Helicobacter*(Mladenova-Hristova et
802 al., 2017) have been demonstrated to be important genera of zoonotic pathogenic bacteria.
803 Emerging infectious diseases (EIDs) of wildlife origin have gained increasing attention in
804 public health discussions(Jones et al., 2008). Salvatore Lucio Cutuli et al. have found that
805 *Hafnia alvei* can cause lung infections and dysregulation of the lung microbiota(Cutuli et
806 al., 2021). Kazuya Okushin et al. have demonstrated that *Helicobacter pylori* infection is
807 closely associated with the occurrence and transmission of liver diseases(Okushin et al.,
808 2018). Numerous studies have demonstrated that the Retroviridae is an important group of
809 human and veterinary pathogens(Balvay et al., 2007; Coffin et al., 2021). They are present
810 in the genomes of many mammals, including rodents, cats, non-human primates, and
811 humans(Balvay et al., 2007; Coffin et al., 2021). Moreover, retroviral infections readily
812 lead to various pathologies such as cancer and immunodeficiency(Balvay et al., 2007;
813 Coffin et al., 2021). Therefore, the high-abundance characteristics of pathogenic bacteria
814 and Retroviridae in the gut of Eulipotyphla suggest that they may serve as potential
815 transmission vectors for zoonotic diseases, highlighting their potential value in public
816 health surveillance. Establishing a long-term monitoring system for the gut microbiome

817 and virome of small mammals (especially species of Eulipotyphla) in Gongga Mountain,
818 and regularly assessing the dynamic changes of pathogens (such as *Helicobacter*, *Hafnia*,
819 and Retroviridae), may contribute to biodiversity conservation and the development of
820 early prevention and control plans for pathogens.

821 Resident gut microbiota widely participate in the decomposition of cellulose in the
822 large intestine of herbivorous mammals(Froidurot & Julliand, 2022). In the gut of
823 Lagomorpha, more colonized microorganisms include bacteria with strong
824 cellulose-decomposing abilities and the production of SCFAs(Fischer et al., 2020; Li et al.,
825 2017; Mukherjee et al., 2023; Rosés et al., 2021). High abundances of Lachnospiraceae
826 were detected in both Rodentia and Lagomorpha. Lachnospiraceae is also an important
827 group of SCFA-producing and cellulose-decomposing bacteria(Jalanka-Tuovinen et al.,
828 2011; Xu et al., 2017). Gut microbiota primarily degrade cellulose in the gut of
829 mammalian hosts through their own CAZYmes(Flint et al., 2012). Compared with
830 carnivorous animals, herbivorous animals have a higher proportion of cellulose in their
831 diet. Therefore, the gut of herbivorous animals may contain more cellulases(Alcaide et al.,
832 2012; Froidurot & Julliand, 2022; Zhou et al., 2020). We found that from Eulipotyphla to
833 Rodentia and then to Lagomorpha, as the proportion of plants in the diet increases, the
834 total abundance of CAZYmes genes in the gut metagenome also increases, this
835 phenomenon consistent with the pattern of gut microbial diversity changes. In the gut
836 metagenome of Lagomorpha, we detected GHs with higher abundances. High abundant
837 CAZYmes play important roles in the decomposition of cellulose(Guo et al., 2024;
838 Mendonça et al., 2023; Zayulina et al., 2019). The composition and function of the gut

839 microbiota in Lagomorpha reflected its feeding characteristics as a typical herbivorous
840 mammal. There were significant positive correlations between gut microbial markers and
841 CAZYmes markers in Rodentia. β-glucosidase in GH1 and chitinases of classes I, II, and
842 IV in GH19, which are more abundant in the gut microbiota of Rodentia, can catalyze the
843 metabolism of cellulose and chitin, respectively(Hemati et al., 2022; Udaya Prakash et al.,
844 2010; Yano et al., 2005). The association between gut microbial species and CAZYmes in
845 Rodentia reflects its characteristics as an omnivorous mammal. During the phylogenetic
846 evolution of mammals, the composition and functions of gut microbiota among closely
847 related herbivorous mammals have extensively converged(Ley et al., 2008; Muegge et al.,
848 2011). Phylogenetic and functional convergence of cellulose-decomposing MAGs from
849 the gut metagenomes of Rodentia and Lagomorpha with those from goat fecal
850 metagenomes further reveals that a cellulose-rich diet shapes the gut microbiota of
851 distantly related mammals. Given that Lagomorpha rely on fiber-rich plants and harbor
852 specialized cellulose-degrading microbial communities (such as Lachnospiraceae and
853 *Prevotella*), the native vegetation in their habitats should be protected to prevent
854 destructive activities like overgrazing and reclamation.

855 In this study, more than 80% of vOTUs predicted from the gut metagenomic data of
856 small mammals were classified into Caudoviricetes, suggesting the important role of
857 Caudoviricetes in the gut virome of small mammals. This phenomenon has also been
858 observed in the gut viromes of yaks on the Qinghai-Xizang Plateau and domestic pigs(Lu
859 et al., 2024). Most phages are ideal candidates for regulating bacterial and archaeal
860 communities in the mammalian gut(Cao et al., 2023). In the gut of Lagomorpha,

Caudoviricetes vOTUs with hosts of *Paenibacillus*, *Prevotella*, and *Treponema* have higher abundances. Based on the results of this study on the characteristic gut microbial genera of Lagomorpha using metagenomics and MAGs, we hypothesize that Caudoviricetes may play a significant role in regulating the structure and function of cellulose-decomposing bacteria in the gut of Lagomorpha(Jiang et al., 2022; Mukherjee et al., 2023; Su et al., 2022). We further predicted the MAGs hosts of all Caudoviricetes vOTUs and found that the association between Caudoviricetes vOTUs and bacterial MAGs in Eulipotyphla suggests weaker regulatory effects compared to Rodentia and Lagomorpha. Additionally, we found that among Rodentia and Lagomorpha, there is a significant positive correlation between the vOTUs of Caudoviricetes and the MAGs of Lachnospiraceae and *Treponema*. While this positive correlation seems to contradict the predator-prey relationship, it may reflect a stable, long-term coexistence state between viruses and bacteria, in which phages play a regulatory rather than an eliminating role. This dynamic process indicates that the predation behavior of phages can regulate but not eliminate dominant bacterial populations, thereby promoting the stability and functional redundancy of microbial communities.

Compared with the widely studied gut microbiota of humans and other medium-to-large mammals, the gut microbiota of small mammals currently receives less attention and lacks MAGs resources(Chen et al., 2021; Deng et al., 2023; Hess et al., 2011; Peng et al., 2021; Svartström et al., 2017). High-quality assembly resources of gut microbial MAGs in small mammals of Rodentia mainly originate from mouse models under experimental conditions(De Nies et al., 2022; Lesker et al., 2020). For wild rodent

883 species, relevant research on gut microbial MAGs has only been published for the bank
884 vole (*Myodes glareolus*). A total of 254 MAGs were ultimately identified, with
885 Bacteroidota being the most abundant phylum in the gut microbiota, followed by Bacillota,
886 Spirochaetes, and Pseudomonadota(Lavrinienko et al., 2020). Multiple studies have
887 characterized the gut microbial communities of Ochotonidae and Leporidae in
888 Lagomorpha, but most of them have not reached the genomic level. Only the wildlife gut
889 MAGs database constructed by Doron Levin et al. includes European hare samples(Kohl
890 et al., 2018; Levin et al., 2021; Paës et al., 2022). In Eulipotyphla, gut microbiota studies
891 have only been conducted on *Suncus murinus* and *Suncus ater*, but neither has reached the
892 microbial genomic level(Koziol et al., 2022; Shinohara et al., 2019). Given the current
893 limited research investment in the gut microbiota of small mammals, we can hypothesize
894 that MAGs hold great potential for discovering previously undescribed gut microbial
895 communities. In this study, microbial genome assembly and screening were conducted on
896 219 gut metagenomic samples of small mammals collected from Gongga Mountain, and
897 the first gut microbiota resource library for small mammals in high-altitude regions was
898 established. A total of 1,385 MAGs were ultimately constructed, among which 1,384
899 MAGs were classified as bacterial genomes. When compared with microbial genomes in
900 the GTDB database, a total of 1,328 MAGs in this study were not classified to the species
901 level, which can be considered as obtaining 1,328 potential novel microbial species
902 genomes. The discovery of novel microbial species genomes can significantly increase the
903 number of available microbial reference genomes and the gut microbial richness of small
904 mammals.

As an important group of mammals, only a few scientific studies have investigated the characteristics of gut viromes and bacteriophage genomes in small mammals under the condition of using mice as models for human disease treatment(Mao et al., 2024; Ritz et al., 2024). In our study, viromes were predicted and filtered from the gut metagenomic sequencing data of 219 small mammals, and a total of 749 vOTUs were ultimately constructed. These vOTUs could effectively supplement the gut virome resources of small mammals. Compared with three established gut virome databases (GPD, GVD, MGV), a total of 599 vOTUs failed to cluster with viral genomes in the databases. This suggests that most of the vOTUs obtained in this study are potential novel gut viruses. Nathaniel L Ritz et al. found a high abundance of Caudoviricetes in the gut virome of mouse models, but did not detect Retroviridae(Ritz et al., 2024). Xiaotian Mao et al. also did not detect Retroviridae in the fecal virome of mice(Mao et al., 2024). In contrast, we predicted 44 vOTUs belonging to Retroviridae from the gut metagenomic samples of small mammals in Gongga Mountain. Notably, all these Retroviridae vOTUs have not been reported in existing gut virus databases (GPD, GVD, MGV). The above-mentioned Retroviridae vOTUs demonstrated the unique characteristics of the gut virome of small mammals in the high-altitude area of Gongga Mountain, which can provide valuable references for future research on retroviruses.

In conclusion, we conducted an in-depth investigation of the gut microbiota of 219 small mammals based on metagenomic sequencing technology. A compositional and functional landscape of gut microbiota of small mammals belonging to three orders in the high-altitude area of Gongga Mountain was established. Our study compared the

compositional and functional differences of gut microbiota among three orders of small mammals, and identified gut microbial species and functional markers specific to different orders of small mammals. The gut of Eulipotyphla harbored more potential pathogens and Retroviridae, suggesting their potential role as transmission vectors for zoonotic diseases and potential public health early-warning value. Through metagenome-based microbial genome assembly and viral prediction, the first gut microbiota and virome resource library for small mammals in high-altitude area was established, providing important data support for future research. It should be noted that the present study relies on metagenomic inference and lacks direct functional validation. Future investigations integrating experimental verification, such as metabolomic analysis, will be instrumental in confirming these predictive outcomes derived from metagenomic sequencing.

Data Availability

The data that support the findings of this study have been deposited into the CNGB Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0007487 and NCBI BioProjectID PRJNA1369446. The data were also deposited in GSA under accession No. CRA034394 and in Science Data Bank under DOI:10.57760/sciencedb.32321.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

L.W.H., R.X.T., Z.J.Z., and Y.L. performed the bioinformatics analyses; S.Y.L. and X.M.W. collected the samples; L.W.H. wrote the manuscript; Z.X.F., B.S.Y. and S.Y.L. revised the manuscript; Z.X.F. designed and supervised the study. All authors read and approved the final version of the manuscript.

Acknowledgements

This work was supported by Sichuan Science and Technology Program (2023NSFSC1935; Zhenxin Fan). We thank Jiao Wang, Xue-na Kang, Yang Yun and Xin-qi Wang at Sichuan University for their help.

Ethics approval and consent to participate

All experimental protocols and Sampling were approved by the Ethics Committee of College of Life Sciences, Sichuan University (No. 20210309009) and Gongga Mountain National Nature Reserve Administration Bureau (No. 513320202100081).

Consent for publication

Not applicable.

Reference

Alcaide M, Messina E, Richter M, et al. 2012. Gene sets for utilization of primary and secondary nutrition supplies in the distal gut of endangered Iberian lynx. *PLOS ONE*, **7**(12): e51521.

Alneberg J, Bjarnason BS, De Bruijn I, et al. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods*, **11**(11): 1144–1146.

Balvay L, Lopez Lastra M, Sargueil B, et al. 2007. Translational control of retroviruses. *Nat Rev Microbiol*, **5**(2): 128–140.

Bayané A & Guiot SR. 2011. Animal digestive strategies versus anaerobic digestion bioprocesses for biogas production from lignocellulosic biomass. *Reviews in Environmental Science and Bio/Technology*, **10**(1): 43–62.

Bowers RM, Kyrpides NC, Stepanauskas R, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol*, **35**(8): 725–731.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**(1): 59–60.

Camarillo-Guerrero LF, Almeida A, Rangel-Pineros G, et al. 2021. Massive expansion of human gut bacteriophage diversity. *Cell*, **184**(4): 1098–1109.e1099.

Cao Y, Feng T, Wu Y, et al. 2023. The multi-kingdom microbiome of the goat gastrointestinal tract. *Microbiome*, **11**(1): 219.

Chaumeil P-A, Mussig AJ, Hugenholtz P, et al. 2022a. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*, **38**(23): 5315–5316.

Chen C, Zhou Y, Fu H, et al. 2021. Expanded catalog of microbial genes and metagenome-assembled genomes from the pig gut microbiome. *Nat Commun*, **12**(1): 1106.

Chen S, Zhou Y, Chen Y, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884-i890.

Coffin J, Blomberg J, Fan H, et al. 2021. ICTV Virus Taxonomy Profile: Retroviridae 2021. *J Gen Virol*, **102**(12).

Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, **33**(18): 2938–2940.

Cote JM & Taylor EA. 2017. The Glycosyltransferases of LPS Core: A Review of Four Heptosyltransferase Enzymes in Context. *Int J Mol Sci*, **18**(11).

Cutuli SL, De Maio F, De Pascale G, et al. 2021. COVID-19 influences lung microbiota dynamics and favors the emergence of rare infectious diseases: A case report of Hafnia Alvei pneumonia. *J Crit Care*, **64**: 173–175.

Dahmane I, Montagner C, Matagne A, et al. 2018. Peptidoglycan glycosyltransferase-ligand binding assay based on tryptophan fluorescence quenching. *Biochimie*, **152**: 1–5.

Danecek P, Bonfield JK, Liddle J, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2): giab008.

De Falco M, Lucariello A, Iaquinto S, et al. 2015. Molecular Mechanisms of Helicobacter pylori Pathogenesis. *J Cell Physiol*, **230**(8): 1702–1707.

De Jonge N, Carlsen B, Christensen MH, et al. 2022. The Gut Microbiome of 54 Mammalian Species. *Front Microbiol*, **13**: 886252.

De Nies L, Busi SB, Tsenkova M, et al. 2022. Evolution of the murine gut resistome following broad-spectrum antibiotic treatment. *Nat Commun*, **13**(1): 2296.

Science Press ZR Zoological Research

Degregori S, Johnson GC, Barber PH, et al. 2024a. Firmicutes and Bacteroidetes contribute to mass gain variation in female obligate hibernators. *Journal of Mammalogy*, **105**(1): 2–12.
Degregori S, Schiettekatte NMD, Casey JM, et al. 2024b. Host diet drives gut microbiome convergence between coral reef fishes and mammals. *Mol Ecol*, **33**(19): e17520.
Deng F, Wang C, Li D, et al. 2023. The unique gut microbiome of giant pandas involved in protein metabolism contributes to the host's dietary adaption to bamboo. *Microbiome*, **11**(1): 180.
Dorigo L, Boscutti F, Sigura M. 2021. Landscape and microhabitat features determine small mammal abundance in forest patches in agricultural landscapes. *PeerJ*, **9**: e12306.
Duque-Correa MJ, Codron D, Meloro C, et al. 2021. Mammalian intestinal allometry, phylogeny, trophic level and climate. *Proc Biol Sci*, **288**(1944): 20202888.
Eletto D, Vllahu M, Mentucci F, et al. 2021. TFF1 Induces Aggregation and Reduces Motility of *Helicobacter pylori*. *Int J Mol Sci*, **22**(4).
Fischbach MA & Segre JA. 2016. Signaling in Host-Associated Microbial Communities. *Cell*, **164**(6): 1288–1300.
Fischer F, Romero R, Hellhund A, et al. 2020. Dietary cellulose induces anti-inflammatory immunity and transcriptional programs via maturation of the intestinal microbiota. *Gut Microbes*, **12**(1): 1–17.
Flint HJ, Scott KP, Duncan SH, et al. 2012. Microbial degradation of complex carbohydrates in the gut. *Gut Microbes*, **3**(4): 289–306.
Froidurot A & Julliand V. 2022. Cellulolytic bacteria in the large intestine of mammals. *Gut Microbes*, **14**(1): 2031694.
Fu L, Niu B, Zhu Z, et al. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, **28**(23): 3150–3152.
Fujita K, Takata I, Yoshida I, et al. 2022. TP0586532, a non-hydroxamate LpxC inhibitor, reduces LPS release and IL-6 production both in vitro and in vivo. *J Antibiot (Tokyo)*, **75**(3): 136–145.
Gao B, Chi L, Zhu Y, et al. 2021. An Introduction to Next Generation Sequencing Bioinformatic Analysis in Gut Microbiome Studies. *Biomolecules*, **11**(4).
Gregory AC, Zablocki O, Zayed AA, et al. 2020. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. *Cell Host Microbe*, **28**(5): 724–740.e728.
Groussin M, Mazel F, Sanders JG, et al. 2017. Unraveling the processes shaping mammalian gut microbiomes over evolutionary time. *Nat Commun*, **8**: 14319.
Guo J, Bolduc B, Zayed AA, et al. 2021. VirSorter2: a multi-classifier, expert-guided approach to detect diverse DNA and RNA viruses. *Microbiome*, **9**(1): 37.
Guo W, Zhou M, Li F, et al. 2024. Seasonal stability of the rumen microbiome contributes to the adaptation patterns to extreme environmental conditions in grazing yak and cattle. *BMC Biol*, **22**(1): 240.
He J, Tu J, Yu J, et al. 2023. A global assessment of Bergmann's rule in mammals and birds. *Glob Chang Biol*, **29**(18): 5199-5210.
Hemati A, Nazari M, Asgari Lajayer B, et al. 2022. Lignocellulosics in plant cell wall and their potential biological degradation. *Folia Microbiol (Praha)*, **67**(5): 671–681.
Hess M, Sczyrba A, Egan R, et al. 2011. Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. *Science*, **331**(6016): 463–467.
Huang F, Shi X, Chen P, et al. 2025. Dietary drivers of gut microbiota diversity and function in wildlife of Wolong Nature Reserve: a metagenomic study. *Current Zoology*: zoaf018.
Hunter JP & Jernvall J. 1995. The hypocone as a key innovation in mammalian evolution. *Proc Natl Acad Sci U S A*, **92**(23): 10718–10722.

Hyatt D, Chen GL, Locascio PF, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**: 119.
Iraçabal L, Barbosa MR, Selvatti AP, et al. 2024. Molecular time estimates for the Lagomorpha diversification. *PLOS ONE*, **19**(9): e0307380.
Jain C, Rodriguez RL, Phillippy AM, et al. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*, **9**(1): 5114.
Jalanka-Tuovinen J, Salonen A, Nikkilä J, et al. 2011. Intestinal microbiota in healthy adults: temporal analysis reveals individual and common core and relation to intestinal symptoms. *PLOS ONE*, **6**(7): e23035.
Jiang L, Shang M, Yu S, et al. 2022. A high-fiber diet synergizes with *Prevotella copri* and exacerbates rheumatoid arthritis. *Cell Mol Immunol*, **19**(12): 1414–1424.
Jones KE, Patel NG, Levy MA, et al. 2008. Global trends in emerging infectious diseases. *Nature*, **451**(7181): 990-993.
Josenhans C, Labigne A, Suerbaum S. 1995. Comparative ultrastructural and functional studies of *Helicobacter pylori* and *Helicobacter mustelae* flagellin mutants: both flagellin subunits, FlaA and FlaB, are necessary for full motility in Helicobacter species. *J Bacteriol*, **177**(11): 3010–3020.
Kaito C, Yoshikai H, Wakamatsu A, et al. 2020. Non-pathogenic Escherichia coli acquires virulence by mutating a growth-essential LPS transporter. *PLoS Pathog*, **16**(4): e1008469.
Kang DD, Li F, Kirton E, et al. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, **7**: e7359.
Kay EH & Hoekstra HE. 2008. Rodents. *Curr Biol*, **18**(10): R406–r410.
Kieft K, Zhou Z, Anantharaman K. 2020. VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome*, **8**(1): 90.
Kohl KD, Varner J, Wilkening JL, et al. 2018. Gut microbial communities of American pikas (Ochotona princeps): Evidence for phylosymbiosis and adaptations to novel diets. *J Anim Ecol*, **87**(2): 323–330.
Koziol A, Odriozola I, Nyholm L, et al. 2022. Enriching captivity conditions with natural elements does not prevent the loss of wild-like gut microbiota but shapes its compositional variation in two small mammals. *Microbiologyopen*, **11**(5): e1318.
Landry SO, Jr. 1970. The rodentia as omnivores. *Q Rev Biol*, **45**(4): 351–372.
Langmead B & Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357–359.
Lavrinienko A, Tukalenko E, Mousseau TA, et al. 2020. Two hundred and fifty-four metagenome-assembled bacterial genomes from the bank vole gut microbiota. *Sci Data*, **7**(1): 312.
Lee SJ, Ahn SY, Oh HB, et al. 2023. Structural and Biochemical Analysis of the Recombination Mediator Protein RecR from *Campylobacter jejuni*. *Int J Mol Sci*, **24**(16).
Lesker TR, Durairaj AC, Gálvez EJC, et al. 2020. An Integrated Metagenome Catalog Reveals New Insights into the Murine Gut Microbiome. *Cell Rep*, **30**(9): 2909–2922.e2906.
Levin D, Raab N, Pinto Y, et al. 2021. Diversity and functional landscapes in the microbiota of animals in the wild. *Science*, **372**(6539).
Ley RE, Hamady M, Lozupone C, et al. 2008. Evolution of mammals and their gut microbes. *Science*, **320**(5883): 1647–1651.
Li D, Liu CM, Luo R, et al. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, **31**(10): 1674–1676.

Li W, Wu X, Hu X, et al. 2017. Structural changes of gut microbiota in Parkinson's disease and its correlation with clinical features. *Sci China Life Sci*, **60**(11): 1223–1233.
Liu C, Cui Y, Li X, et al. 2021. microeco: an R package for data mining in microbial community ecology. *FEMS Microbiology Ecology*, **97**(2): fiaa255.
Liu YX, Chen L, Ma T, et al. 2023. EasyAmplicon: An easy-to-use, open-source, reproducible, and community-based pipeline for amplicon data analysis in microbiome research. *Imeta*, **2**(1): e83.
Lou YC, Chen L, Borges AL, et al. 2024. Infant gut DNA bacteriophage strain persistence during the first 3 years of life. *Cell Host Microbe*, **32**(1): 35–47.e36.
Lu X, Gong G, Zhang Q, et al. 2024. Metagenomic analysis reveals high diversity of gut viromes in yaks (*Bos grunniens*) from the Qinghai-Tibet Plateau. *Commun Biol*, **7**(1): 1097.
Mao X, Larsen SB, Zachariassen LSF, et al. 2024. Transfer of modified gut viromes improves symptoms associated with metabolic syndrome in obese male mice. *Nat Commun*, **15**(1): 4704.
Mcgrosky A, Meloro C, Navarrete A, et al. 2019. Gross intestinal morphometry and allometry in primates. *Am J Primatol*, **81**(8): e23035.
Mcmurdie PJ & Holmes S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, **8**(4): e61217.
Mendonça M, Barroca M, Collins T. 2023. Endo-1,4-β-xylanase-containing glycoside hydrolase families: characteristics, singularities and similarities. *Biotechnology Advances*, **65**: 108148.
Mladenova-Hristova I, Grekova O, Patel A. 2017. Zoonotic potential of Helicobacter spp. *J Microbiol Immunol Infect*, **50**(3): 265–269.
Muegge BD, Kuczynski J, Knights D, et al. 2011. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science*, **332**(6032): 970–974.
Mukherjee S, Lodha TD, Madhuprakash J. 2023. Comprehensive Genome Analysis of Cellulose and Xylan-Active CAZymes from the Genus *Paenibacillus*: Special Emphasis on the Novel Xylanolytic *Paenibacillus sp. LS1*. *Microbiol Spectr*, **11**(3): e0502822.
Myers N, Mittermeier RA, Mittermeier CG, et al. 2000. Biodiversity hotspots for conservation priorities. *Nature*, **403**(6772): 853–858.
Nayfach S, Camargo AP, Schulz F, et al. 2021a. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nature Biotechnology*, **39**(5): 578–585.
Nayfach S, Páez-Espino D, Call L, et al. 2021b. Metagenomic compendium of 189,680 DNA viruses from the human gut microbiome. *Nat Microbiol*, **6**(7): 960–970.
Nishihara H, Hasegawa M, Okada N. 2006. Pegasoferae, an unexpected mammalian clade revealed by tracking ancient retroposon insertions. *Proc Natl Acad Sci U S A*, **103**(26): 9929–9934.
Okushin K, Tsutsumi T, Ikeuchi K, et al. 2018. *Helicobacter pylori* infection and liver diseases: Epidemiology and insights into pathogenesis. *World J Gastroenterol*, **24**(32): 3617–3625.
Olm MR, Brown CT, Brooks B, et al. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, **11**(12): 2864–2868.
Ottemann KM & Lowenthal AC. 2002. *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infect Immun*, **70**(4): 1984–1990.
Paës C, Gidenne T, Bébin K, et al. 2022. Early Introduction of Plant Polysaccharides Drives the Establishment of Rabbit Gut Bacterial Ecosystems and the Acquisition of Microbial Functions. *mSystems*, **7**(3): e0024322.

Paniccia C, Laura Carranza M, Frate L, et al. 2022. Distribution and functional traits of small mammals across the Mediterranean area: landscape composition and structure definitively matter. *Ecological Indicators*, **135**: 108550.
Parks DH, Chuvochina M, Chaumeil P-A, et al. 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nature Biotechnology*, **38**(9): 1079–1086.
Parks DH, Imelfort M, Skennerton CT, et al. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*, **25**(7): 1043–1055.
Patro R, Duggal G, Love MI, et al. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, **14**(4): 417–419.
Peng X, Wilken SE, Lankiewicz TS, et al. 2021. Genomic and functional analyses of fungal and bacterial consortia that enable lignocellulose breakdown in goat gut microbiomes. *Nature Microbiology*, **6**(4): 499–511.
Peres-Neto PR & Jackson DA. 2001. How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, **129**(2): 169–178.
Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol*, **26**(7): 1641–1650.
Puschhof J, Pleguezuelos-Manzano C, Clevers H. 2021. Organoids and organs-on-chips: Insights into human gut-microbe interactions. *Cell Host Microbe*, **29**(6): 867–878.
Qin J, Li R, Raes J, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**(7285): 59–65.
Raetz CR & Whitfield C. 2002. Lipopolysaccharide endotoxins. *Annu Rev Biochem*, **71**: 635–700.
Ren J, Song K, Deng C, et al. 2020. Identifying viruses from metagenomic data using deep learning. *Quant Biol*, **8**(1): 64–77.
Ritz NL, Draper LA, Bastiaanssen TFS, et al. 2024. The gut virome is associated with stress-induced changes in behaviour and immune responses in mice. *Nat Microbiol*, **9**(2): 359–376.
Rosés C, Cuevas-Sierra A, Quintana S, et al. 2021. Gut Microbiota Bacterial Species Associated with Mediterranean Diet-Related Food Groups in a Northern Spanish Population. *Nutrients*, **13**(2).
Rubin EJ & Trent MS. 2013. Colonize, evade, flourish: how glyco-conjugates promote virulence of Helicobacter pylori. *Gut Microbes*, **4**(6): 439–453.
Schloissnig S, Arumugam M, Sunagawa S, et al. 2013. Genomic variation landscape of the human gut microbiome. *Nature*, **493**(7430): 45–50.
Schmartz GP, Hartung A, Hirsch P, et al. 2022. PLSDB: advancing a comprehensive database of bacterial plasmids. *Nucleic Acids Res*, **50**(D1): D273–d278.
Searing KB, Lomolino MV, Rozzi R. 2023. Melting climates shrink North American small mammals. *Proc Natl Acad Sci U S A*, **120**(50): e2310855120.
Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, **30**(14): 2068–2069.
Segata N, Izard J, Waldron L, et al. 2011. Metagenomic biomarker discovery and explanation. *Genome Biology*, **12**(6): R60.
Shang J & Sun Y. 2022. CHERRY: a Computational metHod for accuratE pRediction of virus-pRokarYotic interactions using a graph encoder-decoder model. *Brief Bioinform*, **23**(5).
Shang J, Tang X, Sun Y. 2023. PhaTYP: predicting the lifestyle for bacteriophages using BERT. *Brief Bioinform*, **24**(1): bbac487.
Shinohara A, Nohara M, Kondo Y, et al. 2019. Comparison of the gut microbiotas of laboratory and wild Asian house shrews (*Suncus murinus*) based on cloned 16S rRNA sequences. *Exp Anim*, **68**(4): 531–539.

Science Press
ZR Zoological Research

Siddell SG, Smith DB, Adriaenssens E, et al. 2023. Virus taxonomy and the role of the International Committee on Taxonomy of Viruses (ICTV). *J Gen Virol*, **104**(5).
Sieber CMK, Probst AJ, Sharrar A, et al. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nature Microbiology*, **3**(7): 836–843.
Silby MW, Winstanley C, Godfrey SA, et al. 2011. Pseudomonas genomes: diverse and adaptable. *FEMS Microbiol Rev*, **35**(4): 652–680.
Su M, Hao Z, Shi H, et al. 2022. Metagenomic Analysis Revealed Differences in Composition and Function Between Liquid-Associated and Solid-Associated Microorganisms of Sheep Rumen. *Front Microbiol*, **13**: 851567.
Svartström O, Alneberg J, Terrapon N, et al. 2017. Ninety-nine de novo assembled genomes from the moose (*Alces alces*) rumen microbiome provide new insights into microbial plant biomass degradation. *The ISME Journal*, **11**(11): 2538–2551.
Symonds MR. 2005. Phylogeny and life histories of the 'Insectivora': controversies and consequences. *Biol Rev Camb Philos Soc*, **80**(1): 93–128.
Udaya Prakash NA, Jayanthi M, Sabarinathan R, et al. 2010. Evolution, homology conservation, and identification of unique sequence signatures in GH19 family chitinases. *J Mol Evol*, **70**(5): 466–478.
Wang W, Hu H, Zijlstra RT, et al. 2019. Metagenomic reconstructions of gut microbial metabolism in weanling pigs. *Microbiome*, **7**(1): 48.
Wang Y, Zhang Y, Lane NE, et al. 2022. Population-based metagenomics analysis reveals altered gut microbiome in sarcopenia: data from the Xiangya Sarcopenia Study. *J Cachexia Sarcopenia Muscle*, **13**(5): 2340–2351.
Wang Z, Huang P, You R, et al. 2023. MetaBinner: a high-performance and stand-alone ensemble binning method to recover individual genomes from complex microbial communities. *Genome Biology*, **24**(1): 1.
Wen T, Xie P, Yang S, et al. 2022. ggClusterNet: An R package for microbiome network analysis and modularity-based multiple network layouts. *Imeta*, **1**(3): e32.
Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology*, **20**(1): 257.
Wu Y, Colwell RK, Rahbek C, et al. 2013. Explaining the species richness of birds along a subtropical elevational gradient in the Hengduan Mountains. *Journal of Biogeography*, **40**(12): 2310–2323.
Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*, **32**(4): 605–607.
Xie Y, Xu S, Xi Y, et al. 2024. Global meta-analysis reveals the drivers of gut microbiome variation across vertebrates. *iMetaOmics*, **1**(2): e35.
Xie Z & Tang H. 2017. ISEScan: automated identification of insertion sequence elements in prokaryotic genomes. *Bioinformatics*, **33**(21): 3340–3347.
Xu Q, Yuan X, Gu T, et al. 2017. Comparative characterization of bacterial communities in geese fed all-grass or high-grain diets. *PLOS ONE*, **12**(10): e0185590.
Yano S, Rattanakit N, Wakayama M, et al. 2005. Cloning and expression of a Bacillus circulans KA-304 gene encoding chitinase I, which participates in protoplast formation of Schizophyllum commune. *Biosci Biotechnol Biochem*, **69**(3): 602–609.
Youngblut ND, Reischer GH, Walters W, et al. 2019. Host diet and evolutionary history explain different aspects of gut microbiome diversity among vertebrate clades. *Nat Commun*, **10**(1): 2200.

Zachos FE. 2016. Handbook of the Mammals of the World. vol. 6. Lagomorphs and Rodents I: DE Wilson, TE Lacher, Jr., RA Mittermeier (chief editors). Lynx Edicions, Barcelona (2016). 987 pp., 60 colour plates, 737 colour photographs, 850 distribution maps, hardback.€ 160, ISBN: 978-84-941892-3-4. Springer.

Zayulina KS, Kochetkova TV, Piunova UE, et al. 2019. Novel Hyperthermophilic Crenarchaeon *Thermofilum adornatum sp.* nov. Uses GH1, GH3, and Two Novel Glycosidases for Cellulose Hydrolysis. *Front Microbiol*, **10**: 2972.

Zhang Q, Han S, Liu K, et al. 2019. Occurrence of Selected Zoonotic Fecal Pathogens and First Molecular Identification of Hafnia paralvei in Wild Taihangshan Macaques (*Macaca mulatta tcheliensis*) in China. *Biomed Res Int*, **2019**: 2494913.

Zhou S, Luo R, Gong G, et al. 2020. Characterization of Metagenome-Assembled Genomes and Carbohydrate-Degrading Genes in the Gut Microbiota of Tibetan Pig. *Front Microbiol*, **11**: 595066.

Zhou Y, Liu M, Yang J. 2022. Recovering metagenome-assembled genomes from shotgun metagenomic sequencing data: Methods, applications, challenges, and opportunities. *Microbiol Res*, **260**: 127023.

Zmora N, Suez J, Elinav E. 2019. You are what you eat: diet, health and the gut microbiota. *Nat Rev Gastroenterol Hepatol*, **16**(1): 35–56.

Zu K, Luo A, Shrestha N, et al. 2019. Altitudinal biodiversity patterns of seed plants along Gongga Mountain in the southeastern Qinghai-Tibetan Plateau. *Ecol Evol*, **9**(17): 9586–9596.

贡嘎山野生小型兽类的宿主系统发育和食性对肠道微生物组和病毒组的影响

何乐为[1,2,#]，唐瑞祥[1,#]，刘少英[3,#]，张镇杰[1]，李艺[1]，王旭明[3]，岳碧松[1]，范振鑫[1,2,*]

[1]四川大学生命科学学院，生物资源与生态环境教育部重点实验室，四川成都 610065

[2]四川大学生命科学学院，珍稀濒危野生动物保护四川省重点实验室，四川成都 610064

[3]四川省林业科学研究院，四川成都 610068

#这些作者对这项工作贡献均等

*这些作者是本文的通讯作者

摘要

肠道微生物群在宿主适应过程中发挥着关键作用，然而在高海拔小型兽类中，其组成与功能特征仍未得到充分阐明。本研究以青藏高原生物多样性热点区域——贡嘎山为研究区域，探究了宿主的系统发育（目水平）和食性如何影响三类哺乳动物（劳亚食虫目、啮齿目、兔形目）的肠道微生物组与病毒组。对来自 22 个物种的 219 份样本进行宏基因组测序后，研究发现了具有目特异性的肠道微生物特征：劳亚食虫目（肉食性）动物肠道中潜在病原体（如螺杆菌属、哈夫尼亚菌属）和逆转录病毒科的相对丰度更高；兔形目（草食性）动物肠道中富含纤维素分解菌（如毛螺菌科、普雷沃氏菌属）和碳水化合物活性酶（CAZymes）；啮齿目（杂食性）动物则表现出介于两者之间的特征。本研究还通过组装获得了 1385 个高质量宏基因组组装基因组（MAGs），其中 1328 个可能为新物种；同时鉴定出了 749 个病毒操作分类单元（vOTUs），其中超过 80%属于有尾噬菌体纲。有趣的是，劳亚食虫目动物肠道中逆转录病毒科的高相对丰度提示存在人畜共患病传播风险。肠道噬菌体-宿主互作分析表明，有尾噬菌体纲对兔形目动物肠道中的纤维素分解菌具有调控作用。宿主系

Science Press ZR Zoological Research

统发育与食性共同驱动小型兽类肠道微生物组的分化。本研究建立了贡嘎山高海拔地区首个小型兽类肠道微生物组与病毒组资源库，并推测劳亚食虫目动物可能是人畜共患病病原体的潜在传播载体。

关键词：小型兽类；肠道微生物组；MAGs；病毒组

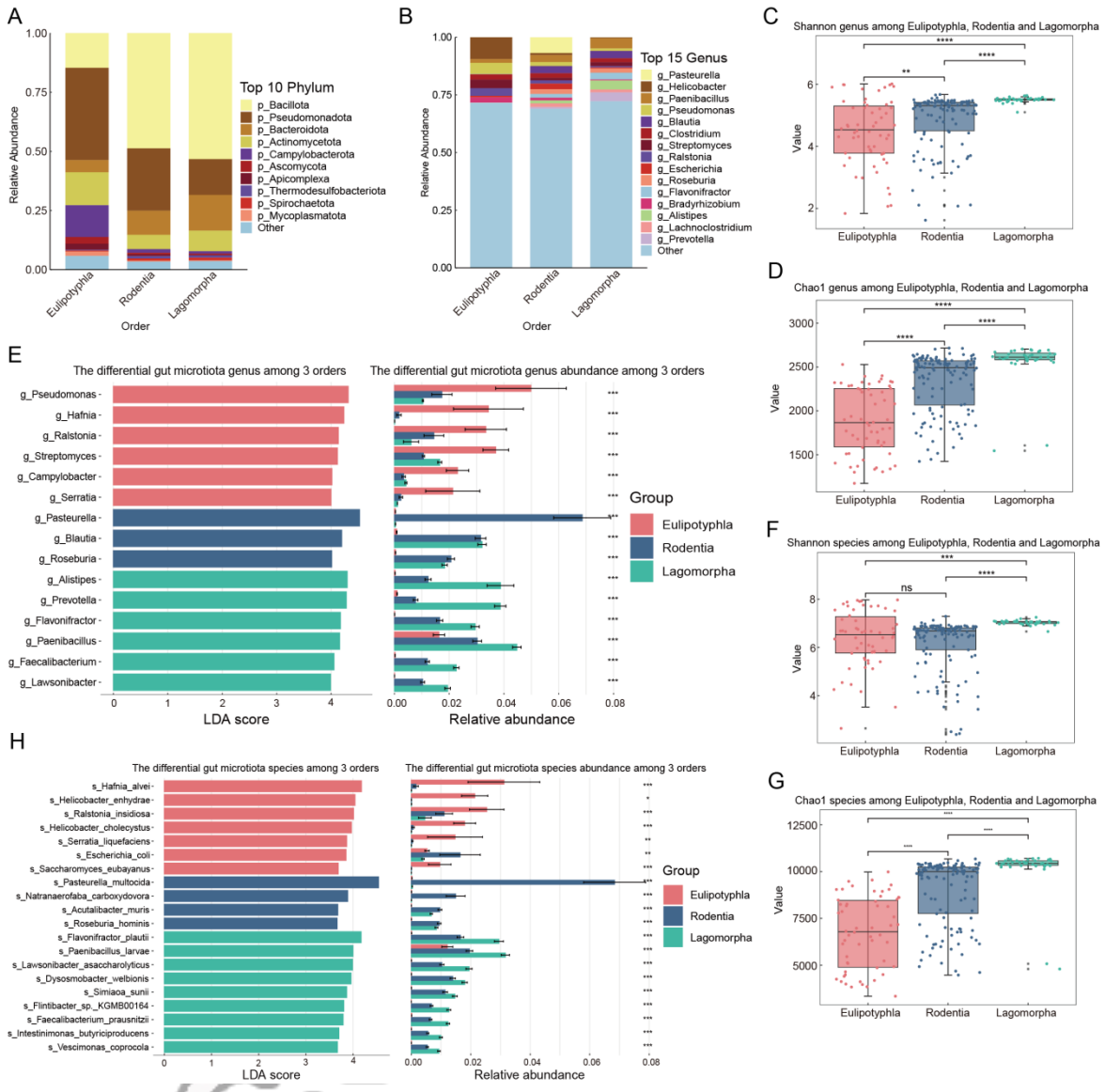

Figure 1. Differences in the composition and abundance of gut microbiota among small mammals of different orders. A: The composition of the top 10 gut microbiota in terms of abundance at the phylum level. **B:** The composition of the top 15 gut microbiota in terms of abundance at the genus level. **C:** The α-diversity (Shannon index) of gut microbiota at the genus level among small mammals of different orders. **D:** The α-diversity (Chao1 index) of gut microbiota at the genus level among small mammals of different orders. **E:** Differences in the abundance of gut microbiota at the genus level among small mammals of different orders. **F:** The α-diversity (Shannon index) of gut

microbiota at the species level among small mammals of different orders. **G:** The α-diversity (Chao1 index) of gut microbiota at the species level among small mammals of different orders. **H:** Differences in the abundances of gut microbiota at the species level among small mammals of different orders. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

Science Press ZR Zoological Research

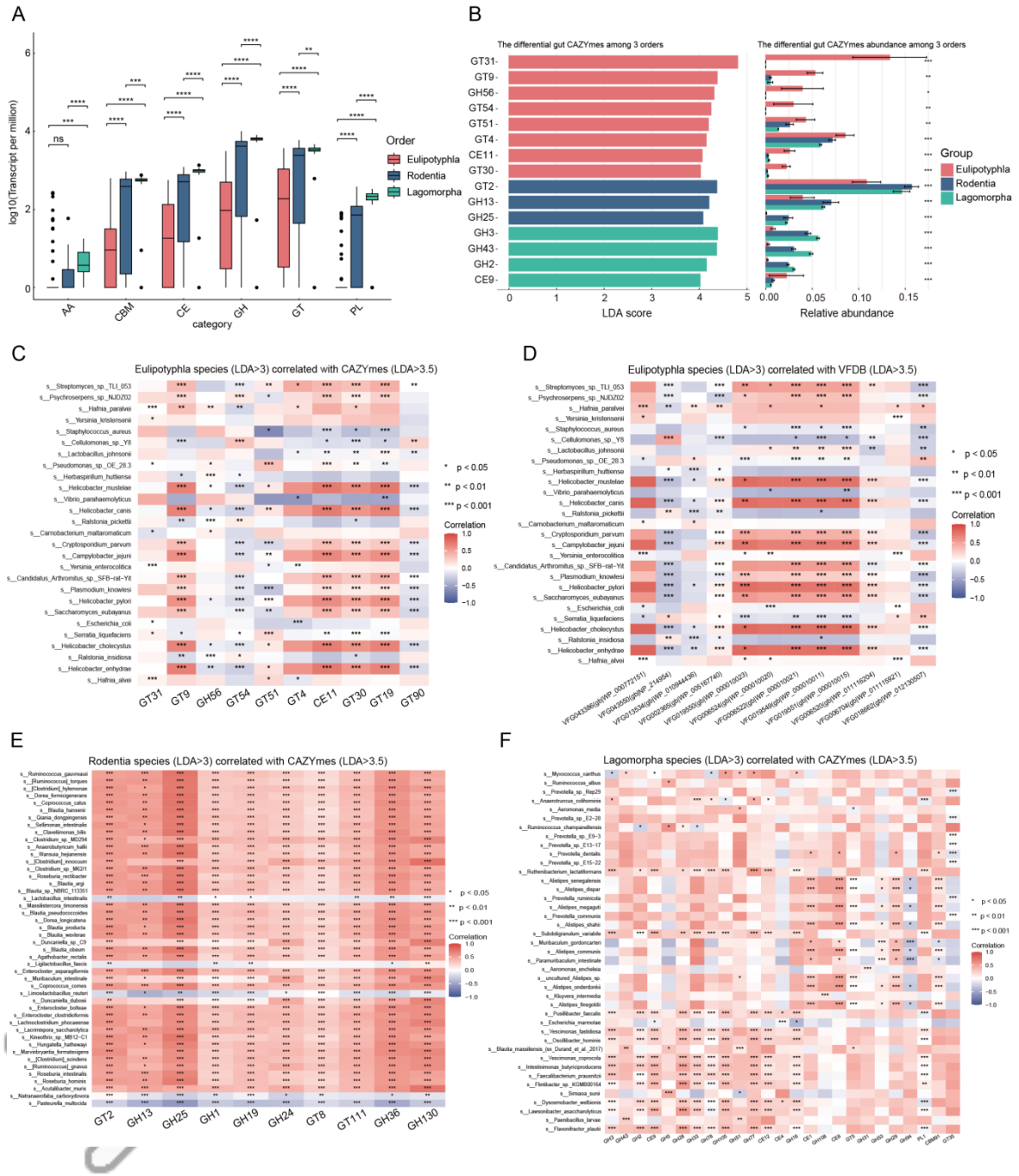

Figure 2. Functional profiles of gut microbiota in small mammals of different orders.

A: Comparison of the total abundances of major categories of CAZYmes in the gut microbiota among small mammals of different orders. **B:** Differences in the abundance of gut microbial CAZYmes among small mammals of different orders. **C:** The association between the species markers of the gut microbiota and the CAZYmes markers in

Science Press
Zoological Research

Eulipotyphla. **D:** The association between the species markers of the gut microbiota and the virulence gene markers in Eulipotyphla. **E:** The association between the species markers of the gut microbiota and the CAZYmes markers in Rodentia. **F:** The association between the species markers of the gut microbiota and the CAZYmes markers in Lagomorpha. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

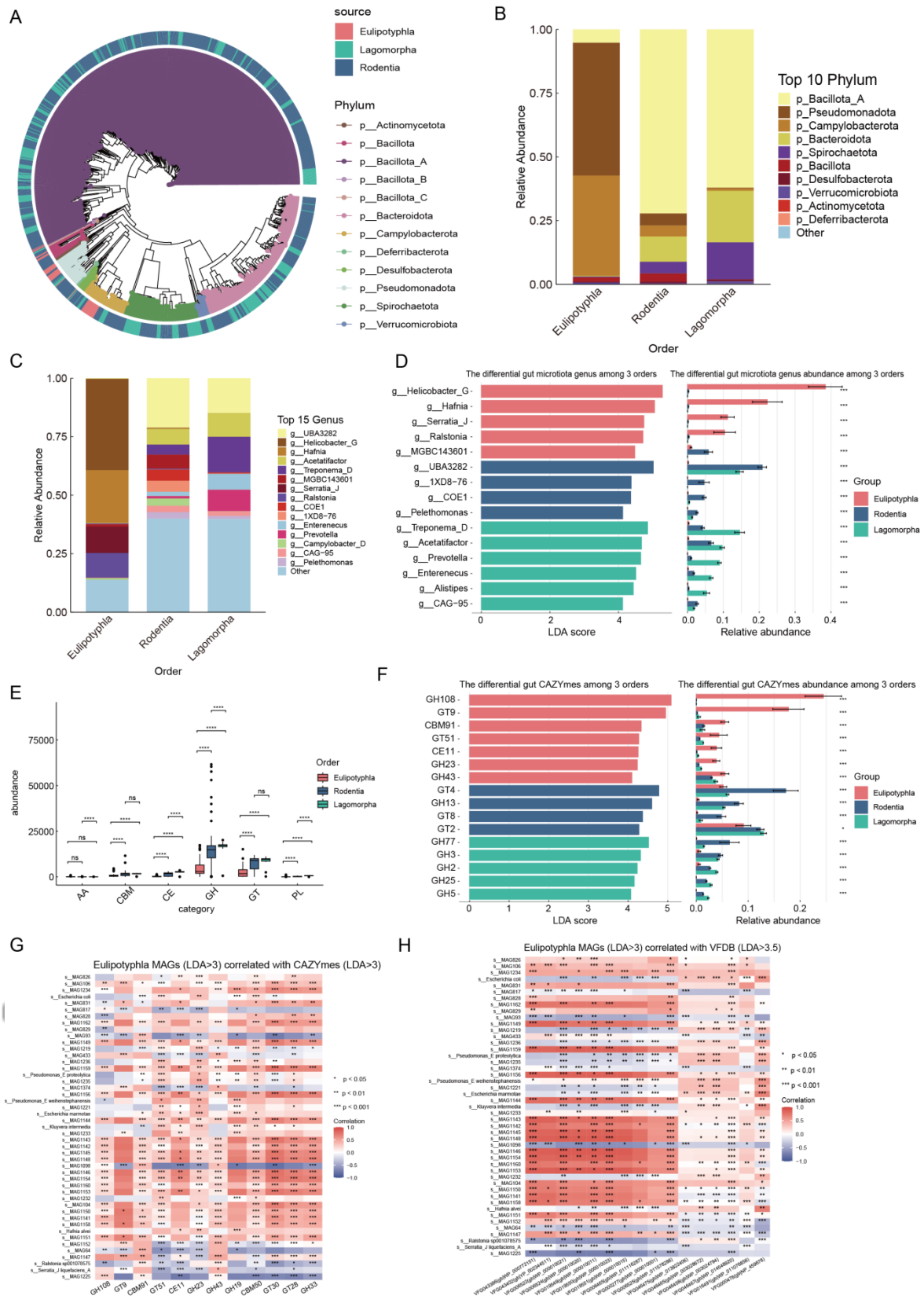

Figure 3. Reconstruction of MAGs in the gut metagenome of small mammals. A: Phylogenetic tree of 1,384 bacterial MAGs derived from small mammals. **B:** The

composition and distribution of the gut microbiota among the top 10 in terms of taxonomic abundance at the phylum level of MAGs. **C:** The composition and distribution of the gut microbiota among the top 15 in terms of taxonomic abundance at the genus level of MAGs. **D:** Differences in the abundances of gut microbiota classified at the genus level of MAGs. **E:** Comparison of the total abundances of each major category of CAZYmes in the MAGs of the gut microbiota in small mammals of different orders. **F:** Differences in CAZYmes in the MAGs of the gut microbiota in small mammals of different orders. **G:** Association between the markers of MAGs of the gut microbiota in Eulipotyphla and the CAZYmes markers of MAGs. **H:** Association between the markers of MAGs of the gut microbiota in Eulipotyphla and the virulence gene markers of MAGs. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

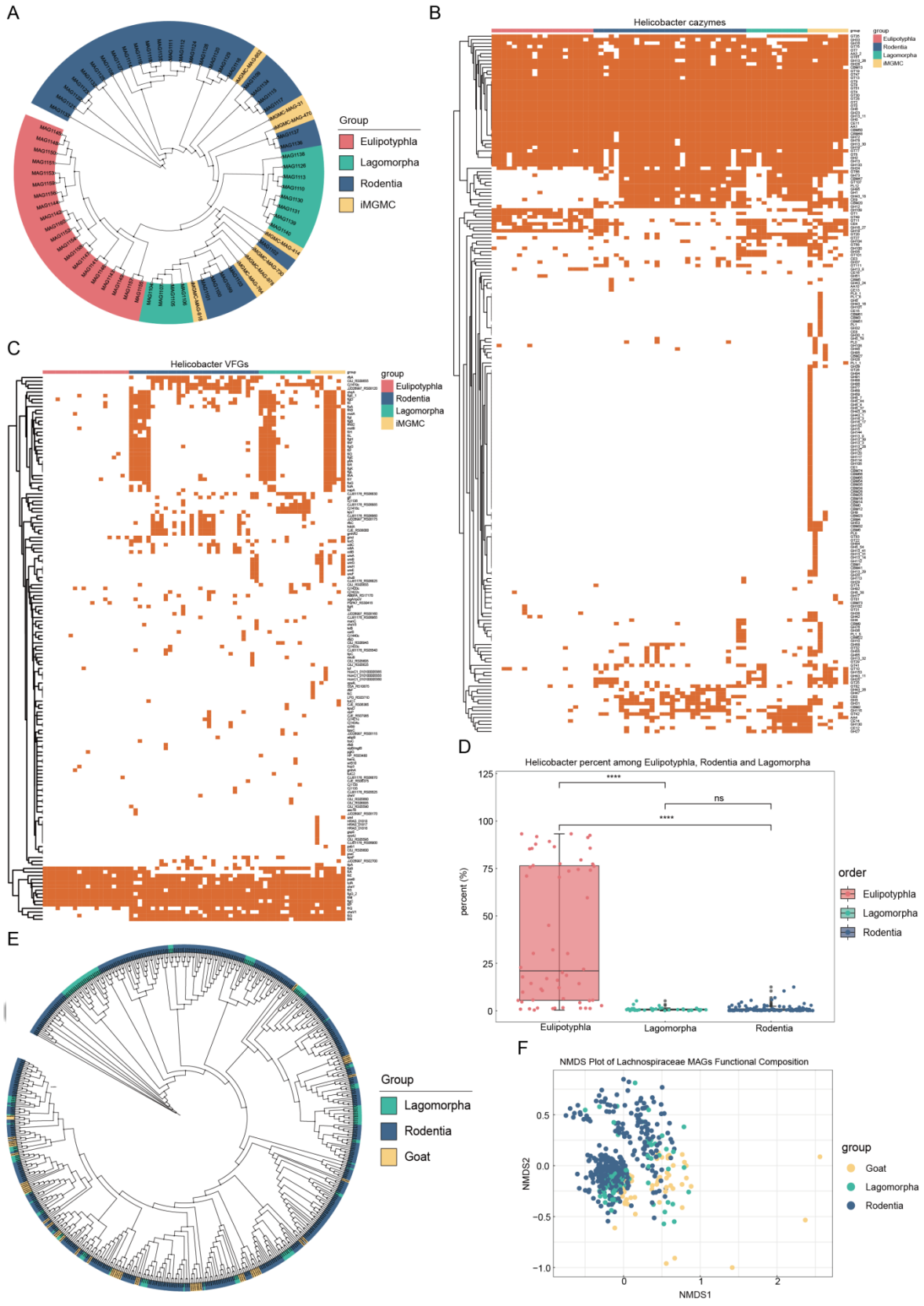

Figure 4. Comparison of MAGs from Helicobacter spp. and cellulolytic bacteria of different sources. A: The phylogenetic tree of the Helicobacter's MAGs. **B:** Heatmap of

the annotation of CAZYmes genes in MAGs of Helicobacter. **C:** Heatmap of the annotation of virulence genes in MAGs of Helicobacter. **D:** Comparison of the relative abundances of Helicobacter's MAGs among small mammals from different orders. **E:** The phylogenetic tree of the Lachnospiraceae's MAGs. **F:** Sort the functional annotations of Lachnospiraceae's MAGs according to different sources. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

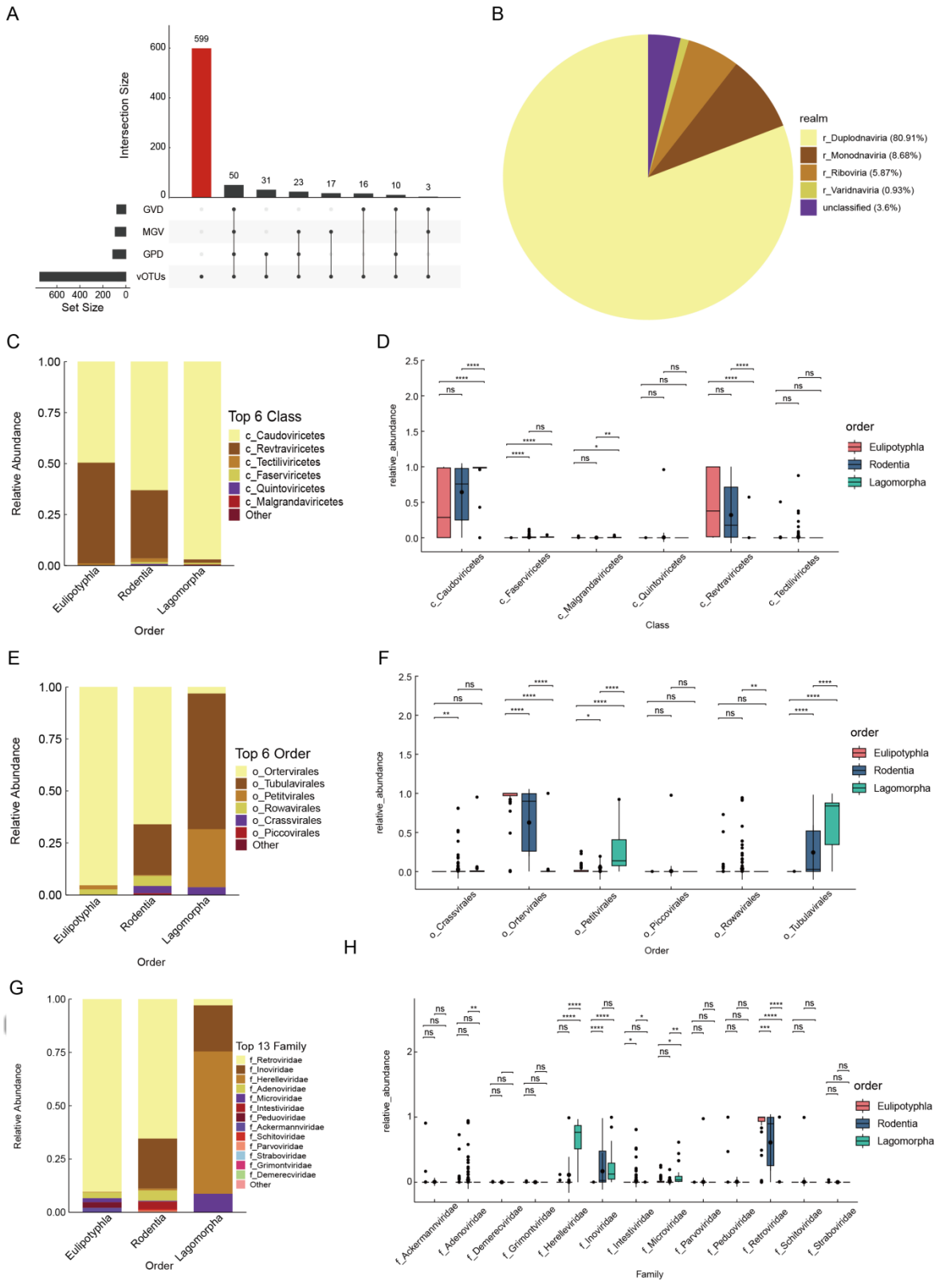

Figure 5. Prediction and characterization of gut viral genomes in small mammals. A: Comparison of gut vOTUs in small mammals with other gut virus genome databases. **B:** Taxonomic classification of vOTUs at the realm level. **C:** The compositional distribution

at the class level of gut vOTUs in small mammals of different orders. **D:** Differences in the abundances at the class level of gut vOTUs in small mammals of different orders. **E:** The compositional distribution at the order level of gut vOTUs in small mammals of different orders. **F:** Differences in the abundances at the order level of gut vOTUs in small mammals of different orders. **G:** The compositional distribution at the family level of gut vOTUs in small mammals of different orders. **H:** Differences in the abundances at the family level of gut vOTUs in small mammals of different orders. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

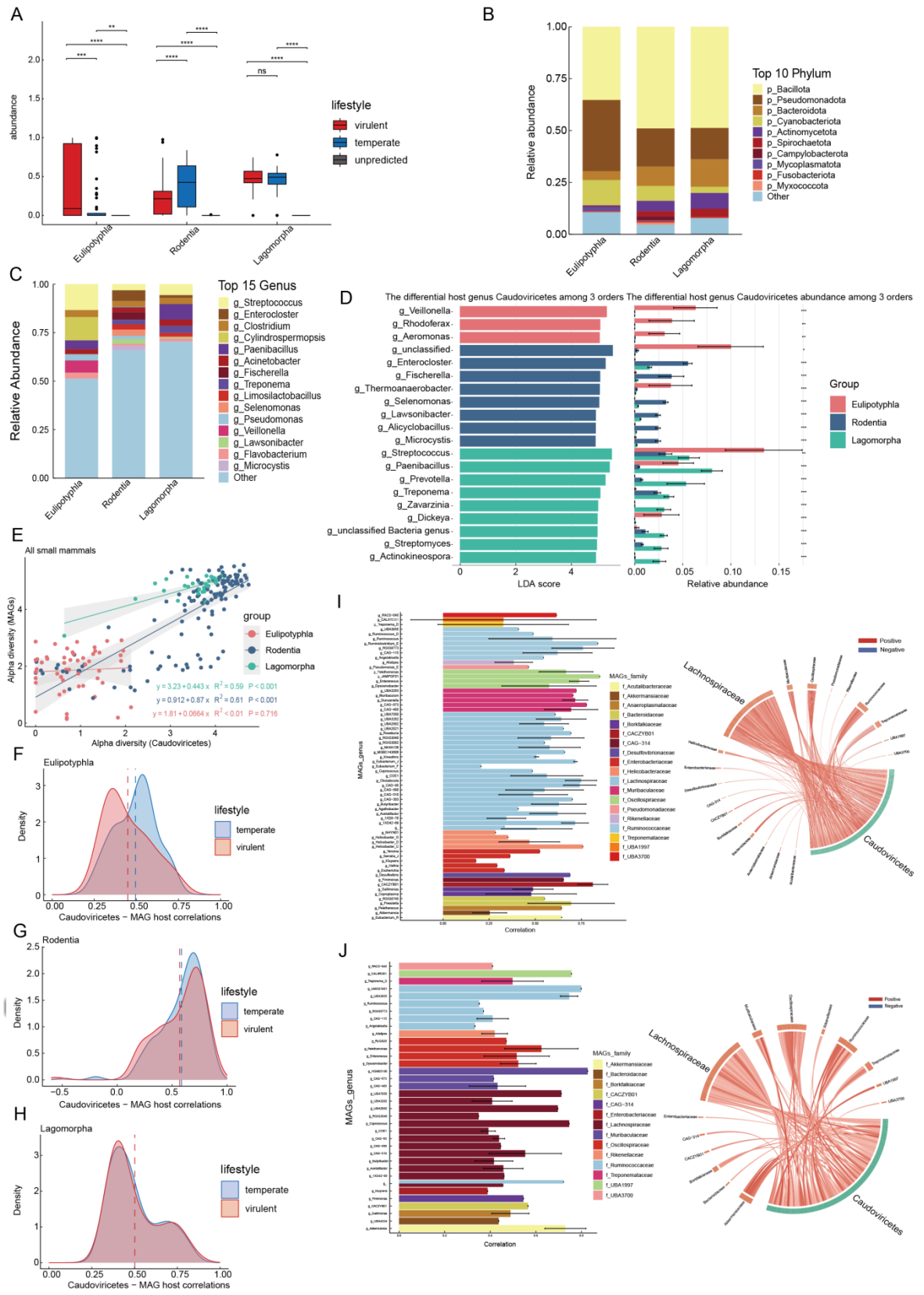

Figure 6. Interaction between Caudoviricetes and hosts in the gut of small mammals.

A: Differences in the lifestyles of the Caudoviricetes in small mammals of different orders.

B: Microbial hosts at the phylum level of Caudoviricetes in the gut of small mammals from different orders, and the abundance distribution of corresponding vOTUs. **C:** Microbial hosts at the genus level of Caudoviricetes in the gut of small mammals from different orders, and the abundance distribution of corresponding vOTUs. **D:** Differences in the abundances of corresponding vOTUs of microbial hosts at the genus level of Caudoviricetes in the gut of small mammals from different orders. **E:** At the species level, the α-diversity between gut Caudoviricetes vOTUs and MAGs in three orders was compared respectively. **F:** Distribution of the correlation between Caudoviricetes vOTUs and host MAGs in the gut samples of Eulipotyphla. The dashed line represents the average correlation coefficient in the distribution. **G:** Distribution of the correlation between Caudoviricetes vOTUs and host MAGs in the gut samples of Rodentia. The dashed line represents the average correlation coefficient in the distribution. **H:** Distribution of the correlation between Caudoviricetes vOTUs and host MAGs in the gut samples of Lagomorpha. The dashed line represents the average correlation coefficient in the distribution. **I:** Correlations between Caudoviricetes vOTUs and predicted bacterial MAG hosts in Rodentia gut samples. The left bar plot shows the Spearman correlation coefficients between Caudoviricetes vOTUs and predicted bacterial MAG hosts at the genus level. The right chord diagram displays the Spearman correlation coefficients between Caudoviricetes vOTUs and predicted bacterial MAGs at the family level. **J:** Correlations between Caudoviricetes vOTUs and predicted bacterial MAG hosts in Lagomorpha gut samples. The left bar plot shows the Spearman correlation coefficients between Caudoviricetes vOTUs and predicted bacterial MAG hosts at the genus level. The

Science Press ZR Zoological Research

right chord diagram displays the Spearman correlation coefficients between Caudoviricetes vOTUs and predicted bacterial MAGs at the family level. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

Supplementary Materials

Supplementary Tables

Supplementary Table S1. Sampling information of gut microbiota from small mammals in Gongga Mountain.

Supplementary Table S2. Sequencing information of metagenomic samples.

Supplementary Table S3. Closely related species of small mammals and reference genomes.

Supplementary Table S4. Differential microorganisms at the phylum level (LDA>3) among small mammals of different orders.

Supplementary Table S5. Differential microorganisms at the genus level (LDA>3) among small mammals of different orders.

Supplementary Table S6. Differential microorganisms at the species level (LDA>3) among small mammals of different orders.

Supplementary Table S7. Differential CAZYmes (LDA>3) among small mammals of different orders.

Supplementary Table S8. Differential virulence genes (LDA>3) among small mammals of different orders.

Supplementary Table S9. Taxonomic informations of MAGs.

Supplementary Table S10. Differential microorganisms at the genus level of MAGs (LDA>3) among small mammals of different orders.

Supplementary Table S11. Differential microorganisms at the species level of MAGs (LDA>3) among small mammals of different orders.

Supplementary Table S12. Differential CAZYmes (LDA>3) in gut microbial MAGs of small

mammals across different orders.

Supplementary Table S13. Differential virulence genes (LDA>3) in gut microbial MAGs of small mammals across different orders.

Supplementary Table S14. Taxonomic informations of vOTUs.

Supplementary Table S15. Lifestyle prediction results of the Caudovirales vOTUs.

Supplementary Table S16. Host prediction results of Caudoviricetes vOTUs based on CHERRY.

Supplementary Table S17. Host prediction results of MAGs for Caudovirales vOTUs.

Supplementary Table S18. Correlation based on relative abundances between Caudovirales vOTUs and their corresponding MAGs hosts in Rodentia.

Supplementary Table S19. Correlation based on relative abundances between Caudovirales vOTUs and their corresponding MAGs hosts in Eulipotyphla.

Supplementary Table S20. Correlation based on relative abundances between Caudovirales vOTUs and their corresponding MAGs hosts in Lagomorpha.

Supplementary Figures

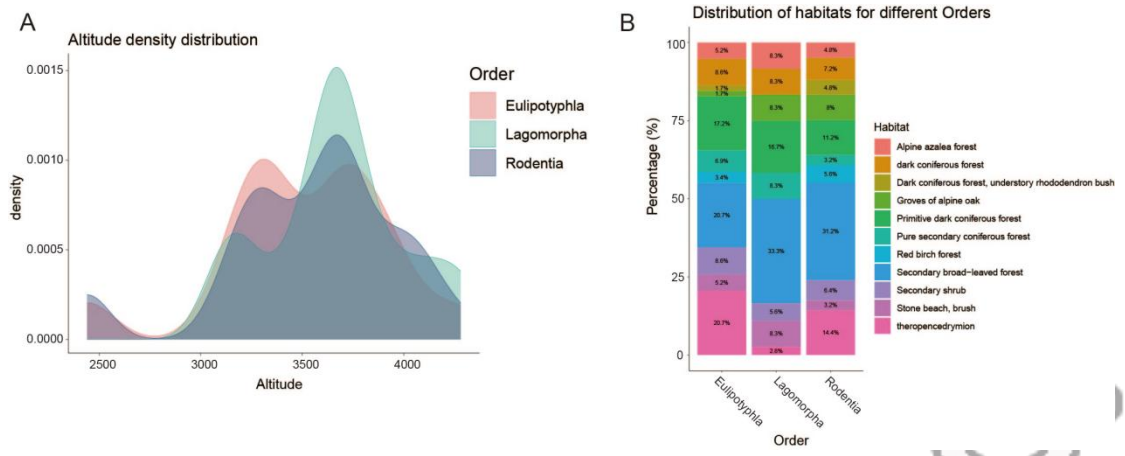

Figure S1. The differences in altitude and habitat environments of sampling locations among the small mammals of different orders. A: The differences in altitude of sampling locations among small mammals of different orders. **B:** The differences in habitat environments of sampling locations among small mammals of different orders.

Science Press
ZR Zoological Research

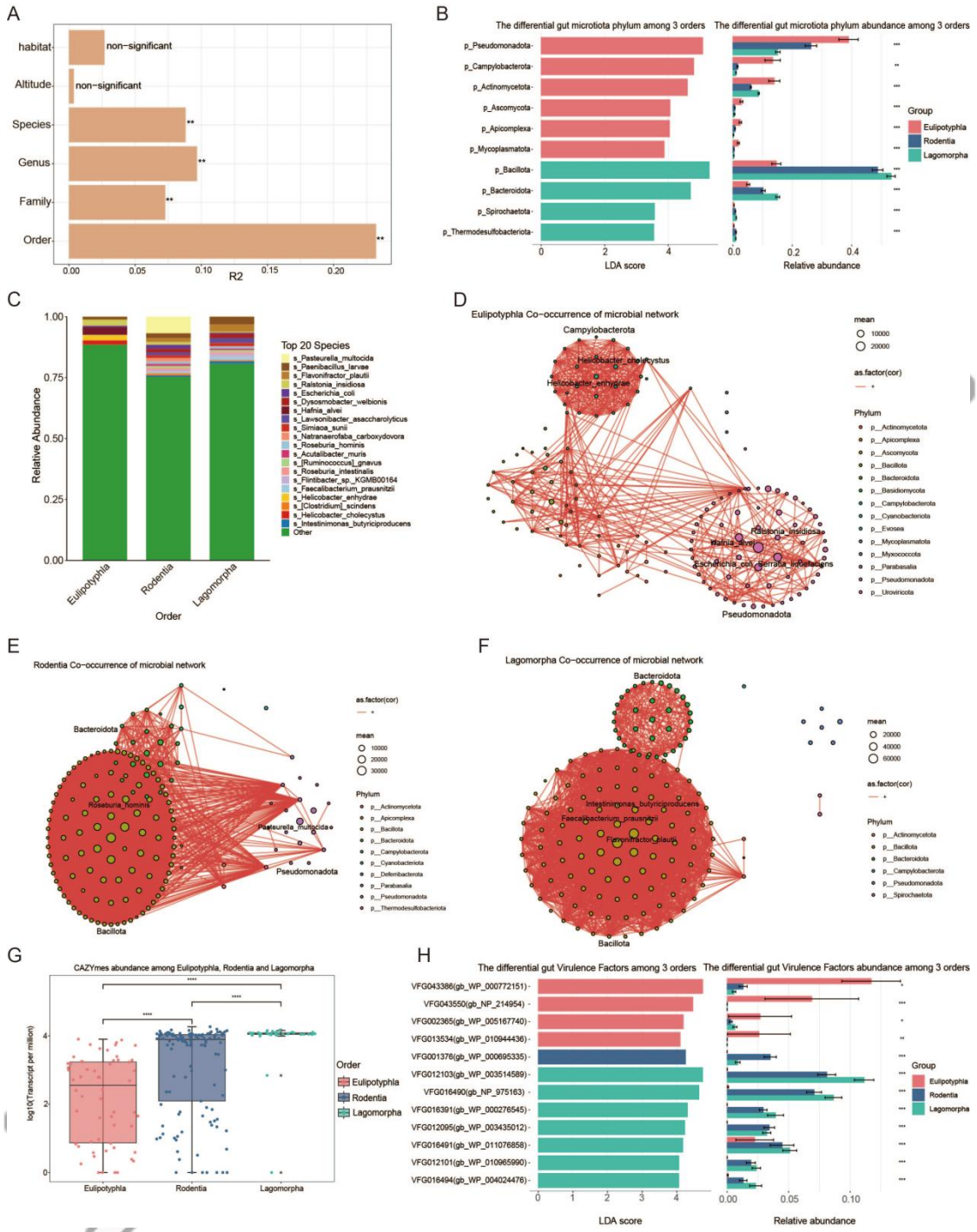

Figure S2. Metagenome-based differences in the composition and functions of gut microbiota among small mammals of different orders. A: The impacts of taxonomic units at all levels of small mammals, as well as the altitude and habitat types of sampling sites, on the β-diversity of gut microbiota. **B:** Differences in the abundance of gut microbiota at the phylum level among small mammals of different orders. **C:** The composition of the top 20 gut microbiota in terms of abundance at

the species level. **D:** Co-occurrence network of gut microbiota in Eulipotyphla. **E:** Co-occurrence network of gut microbiota in Rodentia. **F:** Co-occurrence network of gut microbiota in Lagomorpha. **G:** Comparison of the total abundances of CAZYmes in the gut microbiota among small mammals of different orders. **H:** Differences in the relative abundances of gut microbial virulence genes among small mammals of different orders. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

Science Press ZR Zoological Research

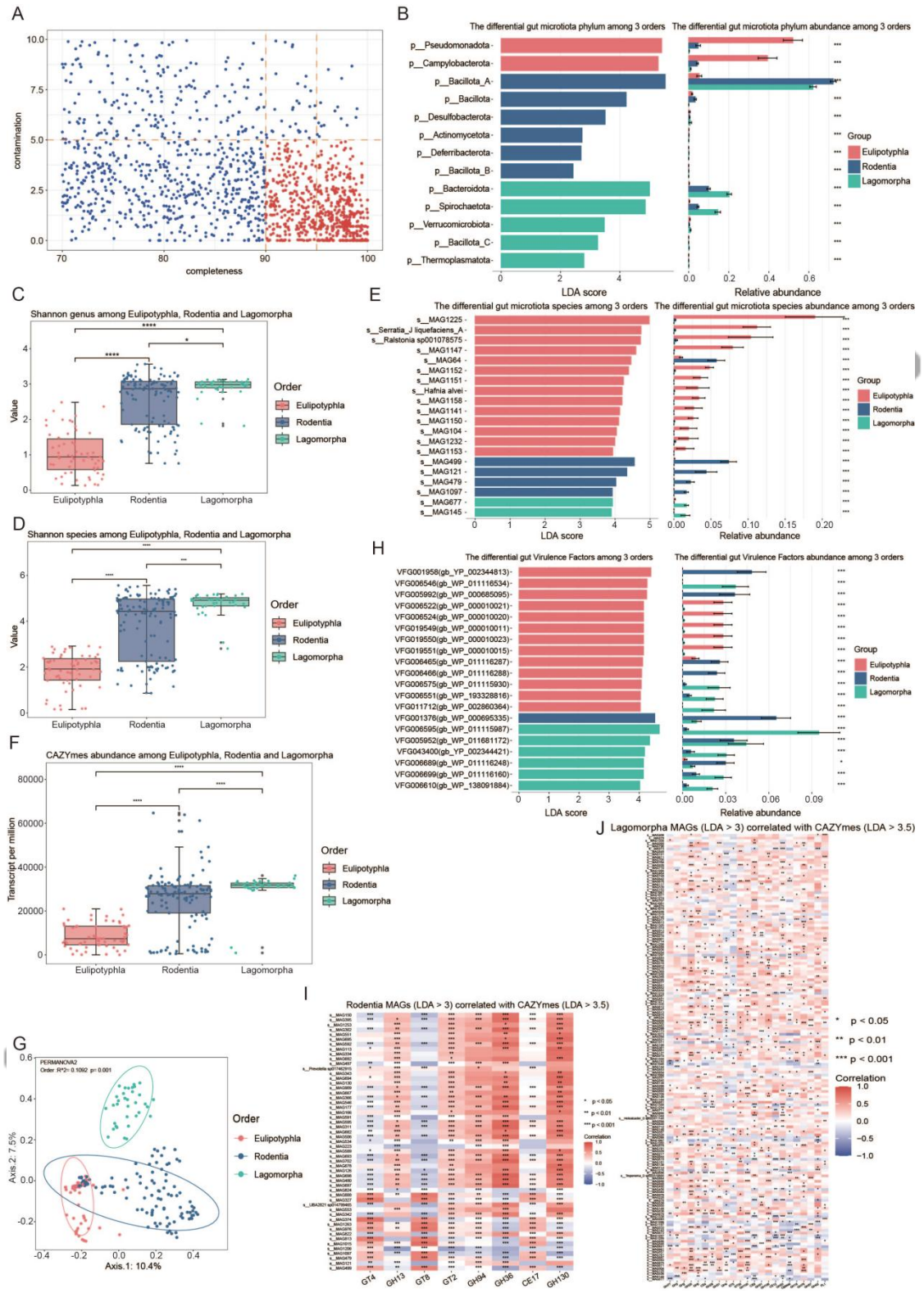

Figure S3. Differences in the composition and functions of gut microbiota among small mammals of different orders based on MAGs. A: The completeness and contamination levels of MAGs of small mammals. **B:** Differences in the abundances of gut microbiota classified at the phylum level of

MAGs. **C:** Differences in α-diversity (Shannon) of the classification at the genus level of MAGs. **D:** Differences in α-diversity (Shannon) of the classification at the species level of MAGs. **E:** Differences in the abundances of gut microbiota classified at the species level of MAGs. **F:** Comparison of the total abundance of CAZYmes in the MAGs of the gut microbiota in small mammals of different orders **G:** PCoA analysis of virulence genes in small mammals of different orders. **H:** Differences in the relative abundances of virulence genes in the MAGs of gut microbiota among small mammals of different orders. **I:** Association between the markers of MAGs of the gut microbiota in Rodentia and the CAZYmes markers of MAGs. **J:** Association between the markers of MAGs of the gut microbiota in Lagomorpha and the CAZYmes markers of MAGs. ns: Not significant; *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

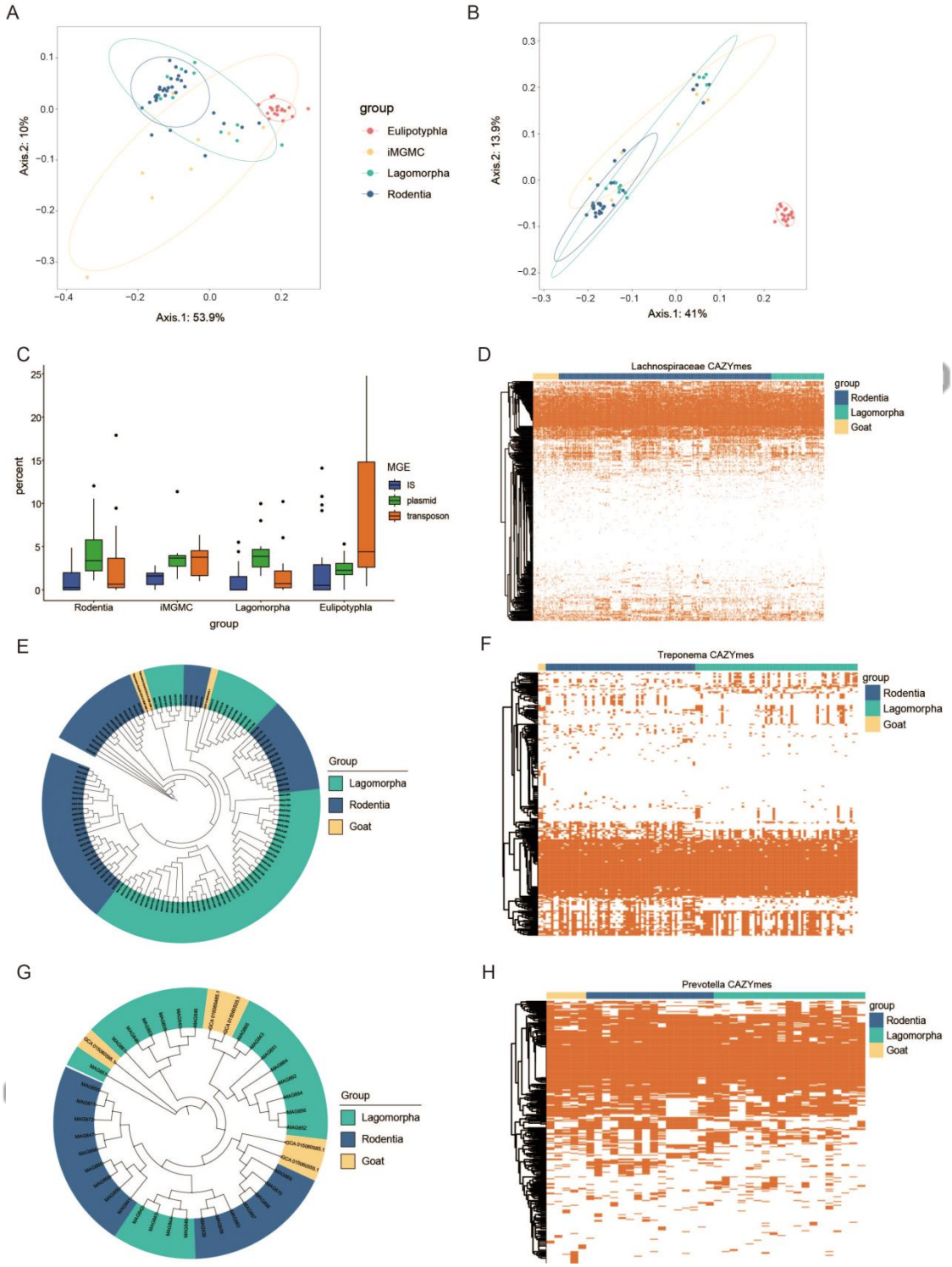

Figure S4. Comparison of MAGs in small mammals. A: PCoA analysis based on the Bray-Curtis distances calculated from the annotation of CAZYmes. **B:** PCoA analysis based on the Bray-Curtis distances calculated from the annotation of VFGs. **C:** The virulence genes in Helicobacter's MAGs from different sources are affected by mobile genetic elements. **D:** Heatmap of the annotation of

1445 CAZYmes genes in MAGs of Lachnospiraceae. **E:** The phylogenetic tree of the Treponema's MAGs. **F:**

1446 Heatmap of the annotation of CAZYmes genes in MAGs of Treponema. **G:** The phylogenetic tree of

1447 the Prevotella's MAGs. **H:** Heatmap of the annotation of CAZYmes genes in MAGs of Prevotella.

1448

Science Press
ZR Zoological Research

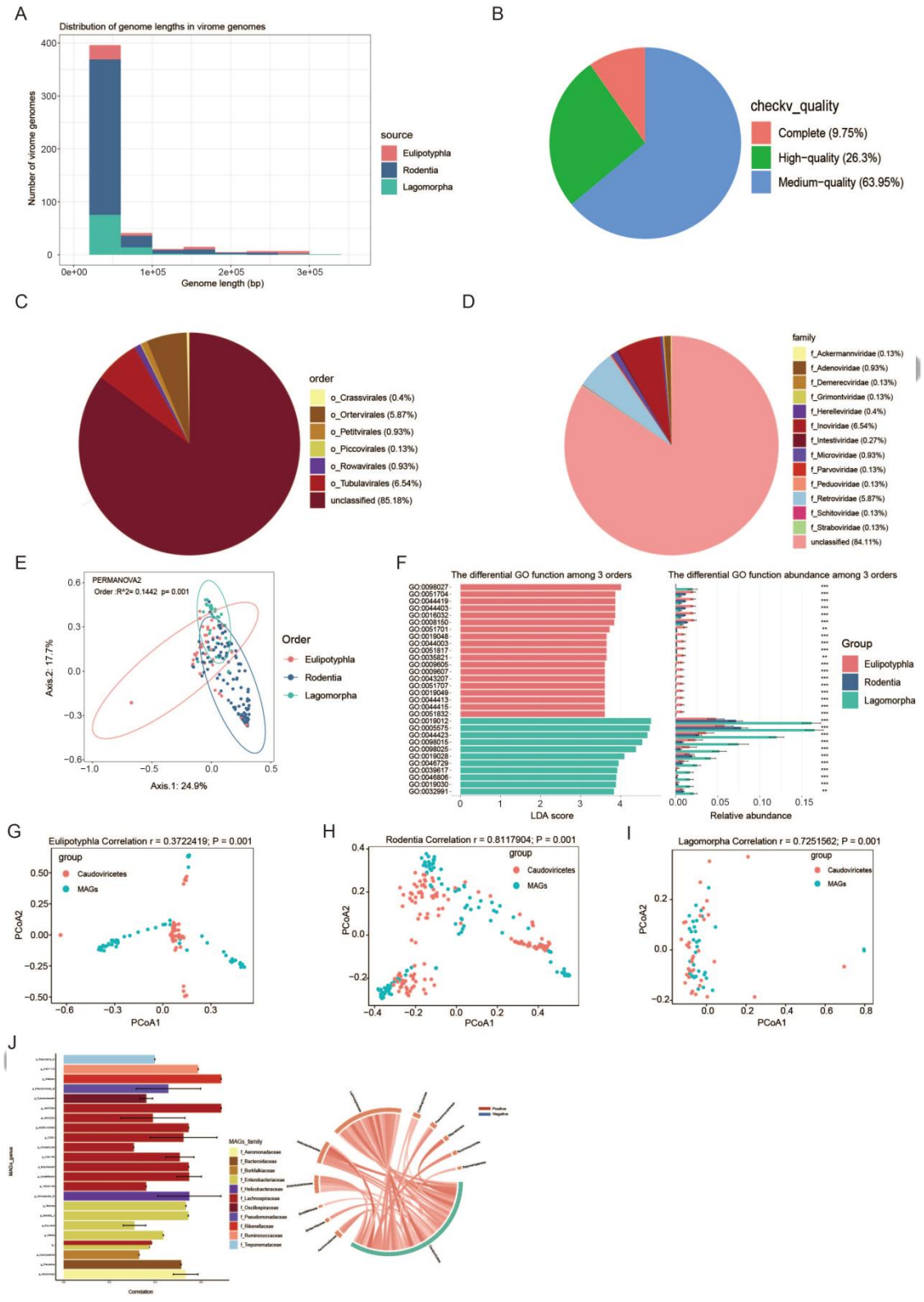

Figure S5. Genomic characteristics of gut viruses in small mammals. A: Histogram of the frequency distribution of the sequence lengths of intestinal vOTUs in small mammals. **B:** Quality distribution of gut vOTUs in small mammals. **C:** Taxonomic classification of vOTUs at the order level.

Science Press
ZR Zoological Research

D: Taxonomic classification of vOTUs at the family level. **E:** PCoA analysis based on the Bray-Curtis distances calculated from the abundances of GO terms. **F:** GO terms with significant abundance differences among different orders. **G:** Procrustes analysis of association between Caudoviricetes vOTUs and its predicted MAGs host of Eulipotyphla gut samples. **H:** Procrustes analysis of association between Caudoviricetes vOTUs and its predicted MAGs host of Rodentia gut samples. **I:** Procrustes analysis of association between Caudoviricetes vOTUs and its predicted MAGs host of Lagomorpha gut samples. **J:** Correlations between Caudoviricetes vOTUs and predicted bacterial MAG hosts in Eulipotyphla gut samples. The left bar plot shows the Spearman correlation coefficients between Caudoviricetes vOTUs and predicted bacterial MAG hosts at the genus level. The right chord diagram displays the Spearman correlation coefficients between Caudoviricetes vOTUs and predicted bacterial MAGs at the family level.