



## Data Article

# Gut microbiomes of wild and domesticated mammals and birds in Slovenia, Europe: 16S rRNA sequencing data

Tanja Zlender<sup>a,b,\*</sup>, Rupnik Maja<sup>a,b</sup><sup>a</sup> Department for microbiological research, National laboratory of health, environment and food, 2000 Maribor, Slovenia<sup>b</sup> Department of microbiology, Faculty of Medicine, University of Maribor, 2000 Maribor, Slovenia

## ARTICLE INFO

*Article history:*

Received 26 November 2025

Revised 15 January 2026

Accepted 2 February 2026

Available online 9 February 2026

Dataset link: [Gut microbiomes of mammals and birds from Slovenia: 16S rRNA sequencing data \(Original data\)](#)*Keywords:*Gut microbiota  
16S rRNA gene  
Metagenome  
Animal feces

## ABSTRACT

From a One Health perspective, the gut microbiota of animals acts as a major driver of microbial exchange between animals and the environment. Animals continuously release gut microbes into their surroundings, shaping environmental and human microbial communities and potentially dispersing pathogens. Characterizing gut microbiota across diverse animal hosts is therefore critical for understanding the patterns of microbial spread through ecosystems and their impact on animal, human and environmental health.

Here, we introduce a large, taxonomically diverse dataset of fecal microbiomes from 715 individual animals representing over 50 mammalian and avian species. We collected samples from both wild and domestic animals with an emphasis on capturing microbial diversity across a wide range of taxa and ecological contexts. The samples were subjected to 16S rRNA gene sequencing, targeting the V3–V4 hypervariable region. Bioinformatic analysis was performed using Usearch to generate zero-radius operational taxonomic units (ZOTUs). This dataset was generated primarily for the development of microbial source tracking (MST) assays used for identifying the sources of fecal pollution in contaminated water. However, it provides a valuable resource for broader microbiome research. It enables comparative studies across host species,

\* Corresponding author.

E-mail address: [tanja.zlender@nlzoh.si](mailto:tanja.zlender@nlzoh.si) (T. Zlender).

trophic guilds, and environmental contexts such as domestication.

© 2026 The Author(s). Published by Elsevier Inc.  
This is an open access article under the CC BY license  
(<http://creativecommons.org/licenses/by/4.0/>)

## Specifications Table

Subject	Environmental genomics and metagenomics
Specific subject area	Gut microbiomes of mammals and birds
Type of data	Tables, figures, raw 16S rRNA sequences and analyzed ZOTU data
Data collection	A total of 715 mammalian and avian fecal samples were collected in Slovenia between December 2020 and September 2023. DNA was extracted and subjected to amplicon sequencing targeting the V3-V4 hypervariable regions of the 16S rRNA gene using next-generation sequencing (NGS) on the Illumina platform. Zero-radius operational taxonomic units (ZOTUs) were generated using Usearch and taxonomy assignment was performed with the RDP training set. The resulting ZOTU table was rarefied to 28,000 reads.
Data source location	Institution: National laboratory for health, environment and food (NLZOH) City: Maribor Country: Slovenia Latitude and longitude: Provided for each sample in Mendeley data repository (10.17632/b43trthrmg.2)
Data accessibility	Raw sequences Repository name: Sequence Read Archive (SRA) Data identification (accession) number: PRJNA1191222 Direct URL to data: <a href="https://dataview.ncbi.nlm.nih.gov/object/PRJNA1191222?reviewer=a1jbni6dtkjattgv7f3roh9ij9">https://dataview.ncbi.nlm.nih.gov/object/PRJNA1191222?reviewer=a1jbni6dtkjattgv7f3roh9ij9</a> Analyzed ZOTU data Repository name: Mendeley data Data identification number: 10.17632/b43trthrmg.2 Direct URL to data: <a href="https://data.mendeley.com/datasets/b43trthrmg/2">https://data.mendeley.com/datasets/b43trthrmg/2</a>
Related research article	Zlender T, Brezočnik L, Podgorelec V, Rupnik M. MicrobiomePrime: A primer pair selection tool for microbial source tracking validated on a comprehensive collection of animal gut and fecal waste microbiomes. <i>Water Research</i> 2026;289:124,990. <a href="https://doi.org/10.1016/j.watres.2025.124990">https://doi.org/10.1016/j.watres.2025.124990</a> .

## 1. Value of the Data

- This dataset provides a large collection of 16S sequencing data from fecal samples of 54 animal species, including some less-characterized hosts.
- The dataset provides information on animal fecal microbiomes of animals from a so far less covered geographical region.
- Important metadata for reuse include sample freshness, living conditions (wild/domestic), diet type, taxonomic ranks (species, genus, family, order, class), sampling location, key dates (sampling, freezing, sequencing), and sequencing platform/polymerase.
- Users can apply these metadata to filter samples (e.g. exclude DMVA samples or samples from captive animals) or account for potential biases (batch effects).
- The data enables comparative analyses of gut microbiota across diverse animal taxa and trophic guilds including alpha and beta diversity, differential abundance, functional prediction and taxonomic composition; exclusion of DMVA samples is recommended.
- The inclusion of animals from the same or closely related species under different living conditions allows for an investigation into how varying environments and domestication influence gut microbiota (e.g. wild boar versus domesticated pig and ibex versus domesticated goat).
- The dataset can be used for cross-host marker discovery, e.g. in library-independent MST.

- The dataset can serve as a reference library for library-dependent (community-based) MST approaches.

## 2. Background

Studying animal gut microbiota is crucial from a One Health perspective, as it is closely linked to environmental microbial ecosystems and the transmission of zoonotic diseases [1]. Moreover, animal gut microbiota represents an under-studied but important reservoir of microorganisms that can influence human microbiome composition through everyday human-animal interactions [2]. In this study, fecal samples from over 50 wild and domestic mammalian and avian species (totaling 715 samples) were collected primarily for the development of library-independent MST assays. These assays are designed to identify sources of fecal contamination in environmental, recreational, and drinking waters by detecting microbial nucleic acids unique to the feces of specific hosts, such as humans, livestock or wildlife. Such nucleic acids serve as biomarkers, which can be identified using a specially designed PCR assay [3,4]. While the MST application is discussed in a separate publication [5], this dataset also provides valuable insights into the gut microbiota of various animal species, some of which have not been well characterized. By understanding how diet, taxonomy, domestication, and habitat shape these communities, we can gain valuable insights for conservation strategies and agricultural practices, as well as public health management.

## 3. Data Description

This dataset consists of high-throughput sequencing data obtained from 515 mammalian and 200 avian fecal samples. In total, 51,345,923 sequences were generated from mammalian samples (mean = 99,700 ± 48,646 per sample) and 17,522,474 from avian samples (mean = 87,612 ± 60,465 per sample).

A summary of the mammalian fecal samples is shown in Table 1. Most samples originate from domesticated animals including cattle, pig, horse, sheep, goat, dog, cat, donkey and rabbit. However, there is also a significant number of wild mammalian fecal samples, especially from ungulates (different species of deer, ibex, chamois, wild boar) and rodents (nutria, beaver, mouse). Other wild animals sampled include hedgehogs, wild carnivorans (fox, otter, badger, bear) and bats. Some of the samples were collected from wild animals temporarily held in captivity for veterinary care, as well as from deer, which are being raised in controlled environments for meat production and tourism, as noted in Mendeley data.

The majority of avian fecal samples are from anatids (swan, duck, goose), columbids (pigeon, dove), gulls and domesticated birds (chicken, quail and turkey). The remaining samples include those from passerines and predatory birds including different species of owls, buzzard, kestrel and stork (Table 2). A proportion of birds sampled in our study (including all predatory birds) were held in captivity for veterinary care (Mendeley data).

Sample sizes per group varied substantially, with some mammalian and avian groups represented by only a few samples. This variability should be considered when interpreting comparative analyses.

The rarefaction curves for different animal groups approached a plateau at different points as illustrated for both mammals (Fig. 1A) and birds (Fig. 1B). Most mammalian samples displayed a general trend of capturing microbial richness at given sequencing depths. However, some fecal samples from animals in the order *Artiodactyla*, including cattle, horse, sheep, goat, ibex and different species of deer did not reach the plateau (Fig S1).

In comparison to mammalian samples, a significant proportion of avian samples had very low concentrations of DNA before sequencing, with some being under the detection limit (Mendeley data). This resulted in very low sequencing depths in some avian samples, with 30 out of 200

**Table 1**

Overview of sampled animal groups from class *Mammalia*. The table includes group names, total sample counts, the number of samples after rarefaction, the number of DMVA samples, the number of sampling locations, and taxonomic classification at the genus, family, and order levels.

Group	Total samples	Rarefied samples	DMVA samples	No. Locations	Species	Genera	Families	Order
Cattle	79	79	2	17	<i>Bos taurus</i>	<i>Bos</i>	<i>Bovidae</i>	<i>Artiodactyla</i>
Ibex	15	15	11	2	<i>Capra ibex</i>	<i>Capra</i>	<i>Bovidae</i>	<i>Artiodactyla</i>
Sheep	14	14	0	5	<i>Ovis aries</i>	<i>Ovis</i>	<i>Bovidae</i>	<i>Artiodactyla</i>
Goat	13	13	3	6	<i>Capra aegagrus hircus</i>	<i>Capra</i>	<i>Bovidae</i>	<i>Artiodactyla</i>
Chamois	10	10	4	1	<i>Rupicapra rupicapra</i>	<i>Rupicapra</i>	<i>Bovidae</i>	<i>Artiodactyla</i>
Fallow deer	34	32	21	4	<i>Dama dama</i>	<i>Dama</i>	<i>Cervidae</i>	<i>Artiodactyla</i>
Roe deer	33	32	20	9	<i>Capreolus capreolus</i>	<i>Capreolus</i>	<i>Cervidae</i>	<i>Artiodactyla</i>
Red deer	17	17	14	3	<i>Cervus elaphus</i>	<i>Cervus</i>	<i>Cervidae</i>	<i>Artiodactyla</i>
Pig	45	45	4	16	<i>Sus scrofa domestica</i>	<i>Sus</i>	<i>Suidae</i>	<i>Artiodactyla</i>
Wild boar	17	17	2	4	<i>Sus scrofa</i>	<i>Sus</i>	<i>Suidae</i>	<i>Artiodactyla</i>
Horse	36	36	4	12	<i>Equus ferus caballus</i>	<i>Equus</i>	<i>Equidae</i>	<i>Perissodactyla</i>
Donkey	2	2	0	2	<i>Equus africanus asinus</i>	<i>Equus</i>	<i>Equidae</i>	<i>Perissodactyla</i>
Beaver	15	15	5	1	<i>Castor fiber</i>	<i>Castor</i>	<i>Castoridae</i>	<i>Rodentia</i>
Nutria	23	23	9	4	<i>Myocastor coypus</i>	<i>Myocastor</i>	<i>Echimyidae</i>	<i>Rodentia</i>
Mouse	26	26	0	1	<i>Apodemus agrarius</i> , <i>Apodemus flavicollis</i> , <i>Apodemus sylvaticus</i> , <i>Apodemus</i> indet.	<i>Apodemus</i>	<i>Muridae</i>	<i>Rodentia</i>
Bat	9	7	6	9	<i>Rhinolophus hipposideros</i> , <i>Myotis myotis</i> , <i>Rhinolophus ferrumequinum</i> , <i>Eptesicus serotinus</i>	<i>Rhinolophus</i> , <i>Myotis</i> , <i>Eptesicus</i>	<i>Rhinolophidae</i> , <i>Vespertilionidae</i>	<i>Chiroptera</i>
Hedgehog	31	28	24	3	<i>Erinaceus concolor</i>	<i>Erinaceus</i>	<i>Erinaceidae</i>	<i>Erinaceomorpha</i>
Rabbit	4	4	0	3	<i>Oryctolagus cuniculus</i>	<i>Oryctolagus</i>	<i>Leporidae</i>	<i>Lagomorpha</i>
Dog	43	43	1	19	<i>Canis lupus familiaris</i>	<i>Canis</i>	<i>Canidae</i>	<i>Carnivora</i>
Cat	33	33	1	9	<i>Felis catus</i>	<i>Felis</i>	<i>Felidae</i>	<i>Carnivora</i>
Fox	4	4	0	2	<i>Vulpes vulpes</i>	<i>Vulpes</i>	<i>Canidae</i>	<i>Carnivora</i>
Otter	5	5	3	3	<i>Lutra lutra</i>	<i>Lutra</i>	<i>Mustelidae</i>	<i>Carnivora</i>
Badger	2	1	2	1	<i>Meles meles</i>	<i>Meles</i>	<i>Mustelidae</i>	<i>Carnivora</i>
Bear	3	3	1	2	<i>Ursus arctos</i>	<i>Ursus</i>	<i>Ursidae</i>	<i>Carnivora</i>
Unidentified	2	1	0	2	<i>Canidae</i> indet., <i>Carnivora</i> indet.	<i>Canidae</i> indet., <i>Carnivora</i> indet.	<i>Canidae</i> indet., <i>Carnivora</i> indet.	<i>Carnivora</i>

Samples with an undetermined host genus are labelled as "Unidentified". The number of samples included in the analysis after rarefaction to 28,000 reads is indicated in the 'Rarefied Samples' column. Legend: DMVA – feces with decreased moisture content or visibly aged appearance, indet. – indeterminate, indicating that the taxon could not be identified beyond the listed taxonomic level.

having fewer than 10,000 reads. Similarly to rarefaction curves of mammals, rarefaction curves for avian groups reached a plateau at different points (Fig. 1B). Only a small proportion of avian samples, primarily those with lower sequencing depth, failed to reach a plateau (Fig. S2).

For the following representation of this dataset, we rarefied the ZOTU table to 28,000 sequences per sample, excluding samples with fewer reads. This threshold was chosen to balance sample size and microbial diversity representation. The rarefaction resulted in the exclusion of 10 mammalian and 44 avian samples, among which 70 % of samples had DNA concentrations below 1 ng/μl and 48 % of samples had DNA concentrations under the detection limit.

Alpha diversity was assessed using a measure of ZOTU richness, with results for individual animal groups and trophic guilds shown in Fig. 2. Notably, some samples with lower moisture content or a visibly aged appearance (e.g. from bear, bat, nutria, and hedgehog) stood out as

**Table 2**

Overview of sampled animal groups from class Aves. The table includes group names, total sample counts, the number of samples after rarefaction, the number of DMVA samples, the number of sampling locations, and taxonomic classification at the genus, family, and order levels.

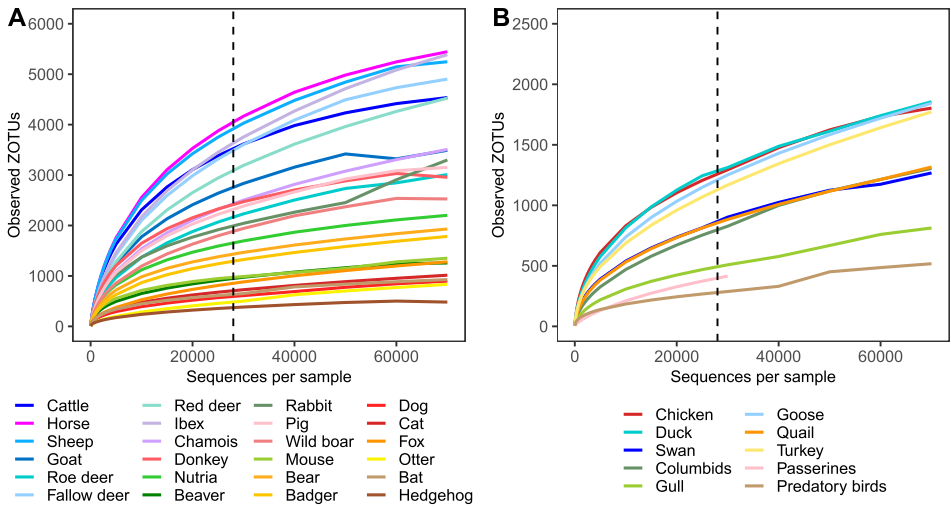
Group	Total sam- ples	Rarefied samples	DMVA sam- ples	No. Lo- cations	Species	Genera	Families	Order
Chicken	35	35	0	15	<i>Gallus gallus domesticus</i>	<i>Gallus</i>	<i>Phasianidae</i>	<i>Galliformes</i>
Quail	4	4	0	1	<i>Coturnix coturnix</i>	<i>Coturnix</i>	<i>Phasianidae</i>	<i>Galliformes</i>
Turkey	2	2	0	1	<i>Meleagris gallopavo domesticus</i>	<i>Meleagris</i>	<i>Phasianidae</i>	<i>Galliformes</i>
Swan	43	38	5	4	<i>Cygnus olor</i>	<i>Cygnus</i>	<i>Anatidae</i>	<i>Anseriformes</i>
Duck	30	23	0	4	<i>Anas platyrhynchos, Anas platyrhynchos domesticus – Cairina moschata domestica, Anas indet.</i>	<i>Anas, Cairina</i>	<i>Anatidae</i>	<i>Anseriformes</i>
Goose	4	4	1	1	<i>Anser indet.</i>	<i>Anser</i>	<i>Anatidae</i>	<i>Anseriformes</i>
Gull	24	15	0	3	<i>Chroicocephalus ridibundus, Larus indet.</i>	<i>Chroicocephalus, Larus</i>	<i>Laridae</i>	<i>Charadriiformes</i>
Columbids	29	20	4	6	<i>Columbia livia domestica, Streptopelia decaocto</i>	<i>Columbia, Streptopelia</i>	<i>Columbidae</i>	<i>Columbiformes</i>
Passerines	7	1	0	1	<i>Parus major, Poecile palustris, Fringilla coelebs, Cyanistes caeruleus</i>	<i>Parus, Fringilla, Cyanistes</i>	<i>Paridae, Fringillidae</i>	<i>Passeriformes</i>
Predatory birds	13	5	1	1	<i>Athene noctua, Otus scops, Tyto alba, Strix uralensis, Buteo buteo, Falco tinnunculus, Strix aluco, Bubo bubo, Ciconia ciconia, Falco naumanni, Ciconia nigra</i>	<i>Athene, Otus, Tyto, Strix, Buteo, Falco, Bubo, Ciconia</i>	<i>Strigidae, Tytonidae, Accipitridae, Falconidae, Ciconiidae</i>	<i>Strigiformes, Accipitriformes, Falconiformes, Ciconiiformes</i>
Unidentified	9	9	0	1	<i>Aves indet.</i>	<i>Aves indet.</i>	<i>Aves indet.</i>	<i>Aves indet.</i>

Samples with an undetermined host genus are labelled as "Unidentified". The number of samples included in the analysis after rarefaction to 28,000 reads is indicated in the 'Rarefied Samples' column. Legend: DMVA – feces with decreased moisture content or visibly aged appearance, indet. – indeterminate, indicating that the taxon could not be identified beyond the listed taxonomic level.

outliers in comparison to freshly collected samples, a pattern also reflected in the beta diversity analysis (Fig. 3). With sample freshness specified in the metadata (Mendeley Data), any samples showing reduced moisture or aging characteristics (samples labeled as DMVA) can be easily excluded from the analysis, if needed.

In mammalian fecal samples, the predominant bacterial phyla were *Bacillota* and *Bacteroidota*. Although bacteria from phylum *Pseudomonadota* were present in all mammalian groups, they were especially highly abundant in hedgehogs, bats, bears and otters. Carnivorous mammals exhibited a visibly higher relative abundance of *Fusobacteria* compared to other trophic guilds (Fig. 4A).

On the family level, the most predominant bacterial taxa in herbivorous and omnivorous mammals were *Oscillospiraceae* and *Lachnospiraceae*. In rodents (nutria, beaver and mouse) there was a higher percentage of *Muribaculaceae* compared to other animal groups. The most predominant families in insectivorous animals were *Enterobacteriaceae*, *Enterococcaceae* and *Clostridiaceae*. In all carnivorous animals apart from otter, which feeds on fish, the predominant families were *Bacteroidaceae*, *Lachnospiraceae*, *Fusobacteriaceae* and *Peptostreptococcaceae*.



**Fig. 1.** Mean rarefaction curves for animal groups. Mean rarefaction curves for animal groups from the orders (A) *Mammalia* and (B) *Aves*. Rarefaction depth selected for a general dataset overview is indicated by a black dashed line. Unidentified animal samples were excluded from this analysis.

The most abundant bacterial phylum in bird samples was *Bacillota*. A notable proportion of *Pseudomonadota* and *Bacteroidota* was also observed in all bird groups except for passerines. On the family level, the most common bacterial taxa in avian feces were *Enterobacteriaceae*, *Clostridiaceae*, *Lactobactillaceae*, *Bacteroidaceae*, *Lachnospiraceae*, *Peptostreoticicaceae* and *Streptococcaceae*. Bacteria from the family *Carnobacteriaceae* were especially highly abundant in gull and predatory birds.

The core microbiome of animal orders *Mammalia* and *Aves* are presented in Fig. 5, showing ZOTUs that are present in all animal groups with a mean prevalence higher than 70 %. In mammals, the most prevalent ZOTUs were Zotu90027 (classified as *Escherichia/Shigella*) and Zotu2 (classified as *Romboutsia*). In birds, Zotu2 was also highly prevalent, alongside Zotu4 (classified as *Clostridium sensu stricto*).

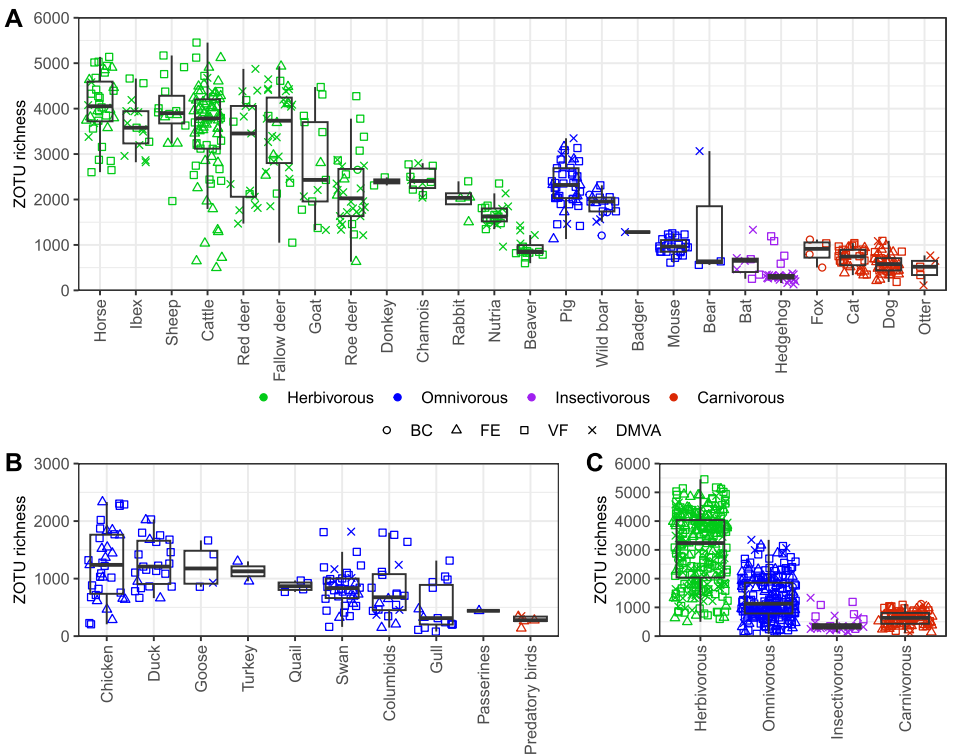
### 3.1. Mendeley data repository structure

The Mendeley Data repository contains both raw and processed versions of the ZOTU table, ZOTU sequences in FASTA format, taxonomy table and sample metadata. The raw data is located in the "Raw" folder, while the processed data (filtered to remove non-bacterial sequences and rarefied) can be found in the "Filtered and rarefied" folder. All scripts used for data processing, analysis, and figure generation are provided in the "Code" folder. A complete list of all files and folders in the Mendeley data repository is provided in Table 3.

The raw data consists of a ZOTU table (zotutab\_raw.tsv), ZOTU sequences in FASTA format (zotus\_raw.fasta), a corresponding taxonomy file (taxonomy\_raw.tsv), and sample metadata (metadata\_raw.xlsx).

The processed dataset has been filtered to remove non-bacterial and chloroplast sequences and rarefied to 28,000 reads. Samples with fewer than 28,000 reads were removed. This version also includes a rarefied ZOTU table (zotutab\_rare.tsv), sequences in FASTA format (zotus\_rare.fasta), a corresponding taxonomy file (taxonomy\_rare.tsv), and sample metadata (metadata\_rare.xlsx).

The metadata files include detailed information on each individual sample including animal taxonomy, breed (if applicable), trophic guild, age and sex of the animal, sample freshness, sam-



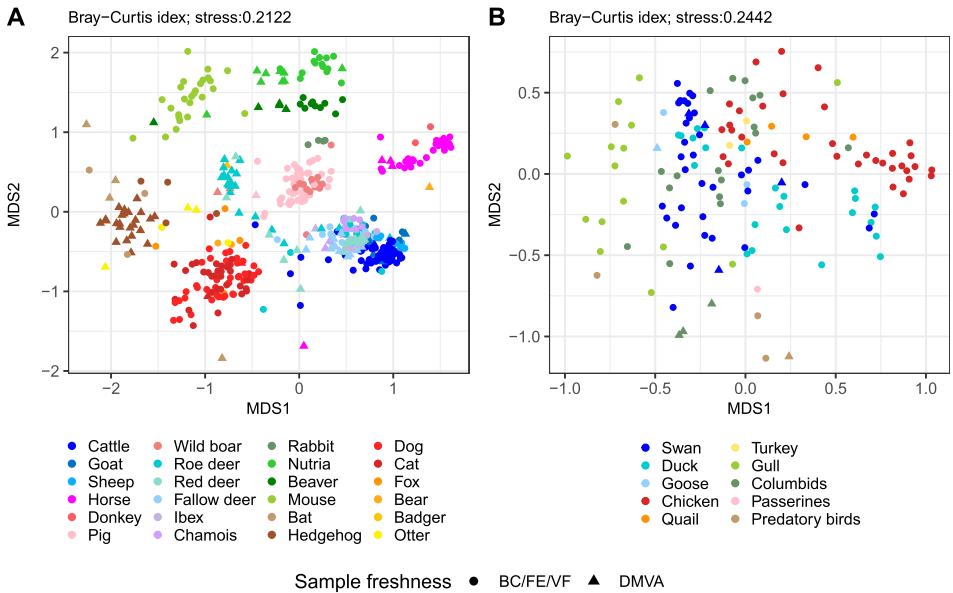
**Fig. 2.** Alpha diversity analysis. ZOTU richness across (A) animal groups from the order *Mammalia*, (B) animal groups from the order *Aves* and (C) animals categorized by trophic guild. Unidentified animal samples were excluded from this analysis. Note that a part of samples (including all predatory bird samples) were obtained from wild animals in captivity, which may influence their microbial profiles. Legend: BC – bowel contents, FE – freshly excreted feces, VF – visibly fresh feces, DMVA – feces with decreased moisture content or visibly aged appearance.

pling location, sampling dates, as well as the dates of sample freezing, DNA isolation, and sequencing. It also provides the information on animal living conditions (wild/domestic), with wild animals further classified into one of the three categories: free-ranging, temporarily captive for veterinary care, and bred and raised in captivity. Furthermore, the metadata provides the details about the polymerase and sequencing platform used.

## 4. Experimental Design, Materials and Methods

### 4.1. Sampling

Fecal samples wild and domestic mammals and birds were collected across Slovenia between December 2020 and September 2023. In total, we obtained 515 mammalian and 200 avian fecal samples. The samples were collected using sterile stool containers with a plastic spoon and immediately stored in ice-filled coolers. Where possible, only the interior of the fecal deposit was sampled to minimize contamination from external sources such as soil, litter and urine. The samples were transported to the laboratory within 24 h. When immediate transport was not possible, they were stored at  $-20\text{ }^{\circ}\text{C}$  and later transported to the laboratory on ice. The samples were stored at  $-80\text{ }^{\circ}\text{C}$  until further processing.



**Fig. 3.** Beta diversity analysis. Visualization of beta diversity across samples from orders (A) *Mammalia* and (B) *Aves* using non-metric multidimensional scaling (NMDS) based on the Bray-Curtis distance matrix. Legend: BC – bowel contents, FE – freshly excreted feces, VF – visibly fresh feces, DMVA – feces with decreased moisture content or visibly aged appearance.

#### 4.2. DNA isolation and amplicon sequencing

The DNA was isolated from 0.25 g of each sample using the QIAamp Fast Stool Mini Kit (Qiagen, Hilden, Germany) with mechanical disruption (MagNA Lyser; 7000 rpm for 70 s). The concentrations of DNA were measured using the Quant-iT PicoGreen dsDNA Kit (Thermo Fisher Scientific). The DNA was stored at  $-80^{\circ}\text{C}$  until further processing.

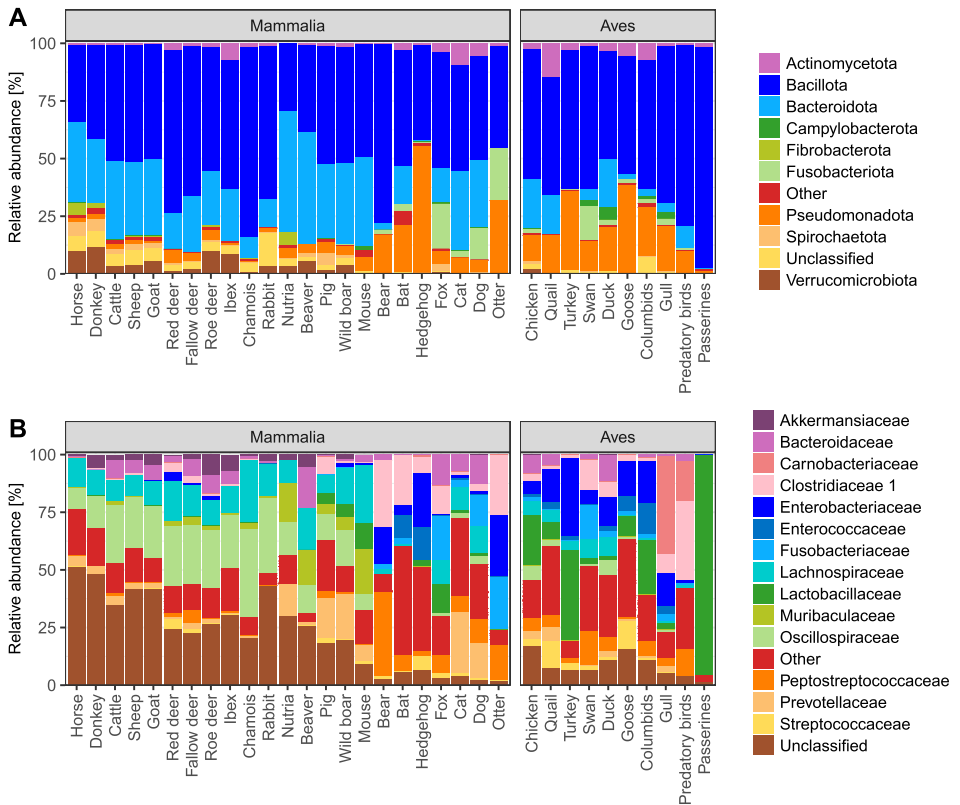
The V3-V4 hypervariable region of the 16S rRNA was amplified with the primers Bakt\_341F (5'-CCTACGGGNGGCWGCAG-3') and Bakt\_805R (5'-GACTACHVGGGTATCTAATCC-3'). Library preparation followed the Illumina 16S Metagenomic Sequencing Library Preparation protocol (Illumina, CA, USA) and was performed using either KAPA HiFi HotStart ReadyMix (Kapa Biosystems, MA, USA) or Q5 High-Fidelity DNA Polymerase (New England Biolabs, USA). After each PCR during library preparation, amplicons were run on a gel alongside a negative control, and no contamination was detected.

Sequencing was carried out on Illumina platforms (MiSeq or NextSeq 2000) using 600-cycle reagent kits. Batch effects may occur due to the use of two sequencing platforms (MiSeq, NextSeq) and two polymerases (KAPA HiFi HotStart ReadyMix, Kapa Biosystems; Q5 High-Fidelity DNA Polymerase, New England Biolabs). Relevant information is provided for each sample in the metadata\_raw.xlsx file (Mendeley Data), allowing these effects to be accounted for during downstream analysis. A 10 % PhiX control was included in each run for quality control.

Raw sequencing data is available in the NCBI SRA under accession number PRJNA1191222.

#### 4.3. Raw sequence analysis

Raw sequences were processed using Usearch version 11.0.667 [6,7]. Paired-end reads were merged with the maximum of 15 mismatches in the alignment. Primer sequences were removed



**Fig. 4.** Microbial composition of gut microbiota in mammals and birds. The mean relative abundances of (A) bacterial phyla and (B) bacterial families in the gut microbiota of different groups of mammals and birds. DMVA samples and unidentified animal samples were excluded from this analysis.

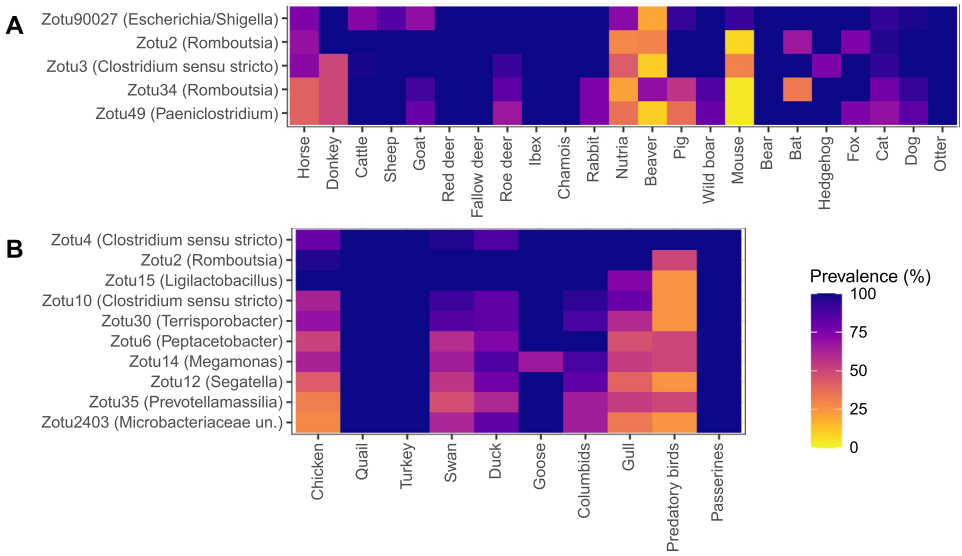
and sequences exceeding one expected error were discarded. Unique sequences underwent denoising with the UNOISE algorithm, generating zero-radius operational taxonomic units (ZOTUs). Sequences shorter than 400 bp were excluded, and taxonomy was assigned using the RDP training set (v.19) [8].

#### 4.4. Filtering and normalization

Further filtering was performed using R version 4.2.3 [9]. Taxa with confidence scores below 0.8 were categorized as unclassified. Non-bacterial ZOTUs, and sequences classified as chloroplasts at the class level were excluded from the dataset. The ZOTU table was rarefied to 28,000 reads per sample using the Vegan package v.2.6.4 [10] with seed set to 28. Samples with <28,000 reads were excluded.

#### 4.5. Animal grouping for microbial composition analysis

To explore patterns in microbial composition, we grouped animals into broad trophic guilds based on their predominant dietary tendencies: herbivorous, omnivorous, carnivorous, and in-



**Fig. 5.** Core ZOTUs shared across mammalian and avian host groups. ZOTUs present in all animal groups within (A) order *Mammalia* and (B) order *Aves*, with a mean prevalence above 70 % across groups. Unidentified animal samples and DMVA samples were excluded from this analysis.

**Table 3**  
Overview of folders and files in the Mendeley Data repository with brief descriptions.

Folder	File name	Content
Filtered and rarefied	metadata_rare.xlsx	Detailed metadata for samples retained after filtering and rarefaction
	taxonomy_rare.tsv	Taxonomic classification for each ZOTU retained after filtering and rarefaction
	zotus_rare.fa	Sequences of each ZOTU retained after filtering and rarefaction in a FASTA format
	zotutab_rare.tsv	Read counts of each ZOTU in each sample after filtering and rarefaction
Raw	metadata.xlsx	Detailed metadata for all samples included in the study
	taxonomy_raw.xlsx	Taxonomic classification for each ZOTU
	zotus_raw.xlsx	Sequences of each ZOTU in a FASTA format
	zotutab_raw.tsv	Read counts of each ZOTU in each sample
Code	usearch_script.sh	Script for running Usearch.
	filtering_and_rarefaction.R	R script for filtering and rarefying raw data generated by Usearch
	figure1_rarefaction_curves.R	R script for generating Figure 1
	figure2_boxplot.R	R script for generating Figure 2
	figure3_NMDS.R	R script for generating Figure 3
	figure4_microbiota_composition.R	R script for generating Figure 4
	figure5_core_microbiome.R	R script for generating Figure 5
	supplementary_figures.R	R script for generating Figures S1 and S2

sectivorous. While some species exhibit dietary flexibility, these categories reflect general feeding ecology trends.

Additionally, we grouped the samples into broader taxonomic or ecological categories to simplify data presentation. Details on group classifications, sample counts, and associated taxonomic levels for each animal group after rarefaction are provided in [Table 1](#), while information for individual samples can be found in the metadata\_raw.xlsx file (Mendeley Data).

#### 4.6. Visualizations and analysis

We analyzed the microbial community data using R v4.2.3 [9]. Rarefaction curves and core microbiome calculations were performed on raw data from samples with  $\geq 28,000$  reads. Chloroplasts and non-bacterial sequences were excluded for the core microbiome analysis. All other analyses and visualizations used filtered, rarefied data. The visualizations were created using ggplot2 v3.4.2 [11]. Alpha diversity was calculated as the number of unique ZOTUs per sample after rarefaction. Beta diversity was then assessed by generating a non-metric multidimensional scaling (NMDS) plot based on the Bray–Curtis distance matrix, using the vegan package v2.6.4 [10].

The core microbiome was determined at the ZOTU level for each animal order (*Mammalia* and *Aves*) separately, using filtered data after excluding non-bacterial sequences. ZOTUs that were present in all animal groups and had a mean prevalence greater than 70 % were considered part of the core microbiome. To calculate mean prevalence, we first determined the prevalence of each ZOTU within each animal group. The mean prevalence was then calculated across all groups.

Unidentified animal samples were excluded from all analyses. DMVA samples were further excluded from microbial composition and core microbiome analyses, as they tended to show different alpha and beta diversity patterns compared with other freshness categories.

#### Limitations

- Uneven sample sizes across species/animal groups.
- Variability in sample freshness, which appears to strongly influence microbial diversity and composition (DMVA samples, defined as feces with decreased moisture content or a visibly aged appearance, can be easily excluded from the analysis).
- Low sequencing depth in some samples (especially avian).
- Temporal span of sampling could introduce biases related to seasonal variation in microbial communities.
- The dataset was generated using two different polymerases and Illumina sequencing platforms, which may introduce variability in sequencing results.

#### Ethics Statement

The authors confirm that they have read and follow the ethical requirements for publication in Data in Brief. The current work does not involve human subjects, animal experiments, or data collected from social media platforms.

According to national guidelines, ethical approval was not required for the collection of animal fecal samples. Most samples were collected from the ground, if possible, immediately after defecation. Bowl contents were obtained post-mortem from some wild animals after being hunted down by members of the Slovenian Hunters Association. No animals were harmed or subjected to disruptions in their daily activities for the purposes of this study.

#### CRediT Authorstatement

**Tanja Zlender:** Investigation, Formal analysis, Writing – Original Draft, Visualization. **Maja Rupnik:** Conceptualization, Resources, Writing – Review & Editing, Supervision, Funding acquisition.

## Data availability

Gut microbiomes of mammals and birds from Slovenia: 16S rRNA sequencing data (Original data) (Mendeley Data).

## Acknowledgements

We would like to acknowledge all those who assisted in providing animal samples, with special thanks to Dr. Zlatko Golob for supplying samples from his wildlife sanctuary. We also thank Kaja Tominc, Masa Jarcic and Sabina Mlakar for their support in performing DNA isolations.

This work was supported by the [Slovenian Research and Innovation Agency](#) [research core funding [P3-0387](#) and the PhD program “[Mladi raziskovalci](#)”]. HPC computing cluster was obtained within Grant/[2021/PHF/23776](#) (EU/ECDC).

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.dib.2026.112564](https://doi.org/10.1016/j.dib.2026.112564).

## References

- [1] L. Ma, H. Zhao, L.B. Wu, Z. Cheng, C. Liu, Impact of the microbiome on human, animal, and environmental health from a one health perspective, *Sci. One Health* 2 (2023) 100037, doi:[10.1016/j.soh.2023.100037](https://doi.org/10.1016/j.soh.2023.100037).
- [2] S. Kuthyar, A.T. Reese, Variation in microbial exposure at the human-animal interface and the implications for microbiome-mediated health outcome, *mSystems* 6 (2021) e0056721, doi:[10.1128/mSystems.00567-21](https://doi.org/10.1128/mSystems.00567-21).
- [3] K. Demeter, R. Linke, E. Ballesté, G. Reischer, R.E. Mayer, J. Vierheilig, et al., Have genetic targets for faecal pollution diagnostics and source tracking revolutionized water quality analysis yet? *FEMS Microbiol. Rev.* 47 (2023) fuad028, doi:[10.1093/femsre/fuad028](https://doi.org/10.1093/femsre/fuad028).
- [4] T. Zlender, M. Rupnik, An overview of molecular markers for identification of non-human fecal pollution sources, *Front. Microbiol.* 14 (2023), doi:[10.3389/fmicb.2023.1256174](https://doi.org/10.3389/fmicb.2023.1256174).
- [5] T. Zlender, L. Brezocnik, V. Podgorelec, Rupnik M. MicrobiomePrime, A primer pair selection tool for microbial source tracking validated on a comprehensive collection of animal gut and fecal waste microbiomes, *Water Res.* 289 (2026) 124990, doi:[10.1016/j.watres.2025.124990](https://doi.org/10.1016/j.watres.2025.124990).
- [6] R.C. Edgar, Search and clustering orders of magnitude faster than BLAST, *Bioinformatics* 26 (2010) 2460–2461, doi:[10.1093/bioinformatics/btq461](https://doi.org/10.1093/bioinformatics/btq461).
- [7] R.C. Edgar, H. Flyvbjerg, Error filtering, pair assembly and error correction for next-generation sequencing reads, *Bioinformatics* 31 (2015) 3476–3482, doi:[10.1093/bioinformatics/btv401](https://doi.org/10.1093/bioinformatics/btv401).
- [8] Wang Q., Cole J.R. Updated RDP taxonomy and RDP Classifier for more accurate taxonomic classification. *Microbiol. Resour. Announc.* nd.;13:e01063-23. <https://doi.org/10.1128/mra.01063-23>
- [9] R Core Team. R: a language and environment for statistical computing 2023.
- [10] Oksanen J., Blanchet F.G., Kindt R., Legendre P., Minchin P., O'Hara B., et al. *Vegan: community ecology package*. R Package Version 22-1 2015;2:1–2.
- [11] H. Wickham, *Ggplot2: Elegant Graphics For Data Analysis*, 1st ed., Springer, New York, 2016.