

Exploring microbial diversity and biosynthetic potential in zoo and wildlife animal microbiomes

Received: 18 December 2023

Accepted: 13 September 2024

Published online: 26 September 2024



Georges P. Schmartz^{1,7}, Jacqueline Rehner^{2,7}, Miriam J. Schuff²,
Leidy-Alejandra G. Molano¹, Sören L. Becker², Marcin Krawczyk³,
Azat Tagirdzhanov^{1,4}, Alexey Gurevich^{4,5}, Richard Francke⁶, Rolf Müller⁴,
Verena Keller^{1,7} & Andreas Keller^{1,4,7} ✉

Understanding human, animal, and environmental microbiota is essential for advancing global health and combating antimicrobial resistance (AMR). We investigate the oral and gut microbiota of 48 animal species in captivity, comparing them to those of wildlife animals. Specifically, we characterize the microbiota composition, metabolic pathways, AMR genes, and biosynthetic gene clusters (BGCs) encoding the production of specialized metabolites. Our results reveal a high diversity of microbiota, with 585 novel species-level genome bins (SGBs) and 484 complete BGCs identified. Functional gene analysis of microbiomes shows diet-dependent variations. Furthermore, by comparing our findings to wildlife-derived microbiomes, we observe the impact of captivity on the animal microbiome, including examples of converging microbiome compositions. Importantly, our study identifies AMR genes against commonly used veterinary antibiotics, as well as resistance to vancomycin, a critical antibiotic in human medicine. These findings underscore the importance of the ‘One Health’ approach and the potential for zoonotic transmission of pathogenic bacteria and AMR. Overall, our study contributes to a better understanding of the complexity of the animal microbiome and highlights its BGC diversity relevant to the discovery of novel antimicrobial compounds.

Microorganisms and the microbiomes they shape wield considerable influence on broader ecological dynamics despite their small scale. While much attention in clinical microbial research focuses on human pathogens and associated microbiomes, there is a growing recognition of the interconnectedness among animal, human, and environmental health, underscored by the ‘One Health’ paradigm^{1–7}. Accordingly, there is an increasing interest in exploring environmental and animal-

associated microbial ecosystems. At the forefront of the ‘One Health’ paradigm lies the pressing issue of antimicrobial resistance (AMR)⁸. The emergence and spread of AMR pose a significant threat to public health worldwide, with escalating concerns about its impact on a population scale^{9,10}. While extensive literature has documented the prevalence of AMR in farm animals, shedding light on the consequences of intensive animal farming, comparatively scarce data is

¹Chair for Clinical Bioinformatics, Saarland University, 66123 Saarbrücken, Germany. ²Institute of Medical Microbiology and Hygiene, 66421 Saarland University, Homburg, Germany. ³Department of Medicine II, 66421 Saarland University, Homburg, Germany. ⁴Helmholtz Institute for Pharmaceutical Research Saarland, Helmholtz Center for Infection Research, 66123 Saarbrücken, Germany. ⁵Department of Computer Science, Saarland University, 66123 Saarbrücken, Germany. ⁶Zoo Saarbrücken, 66121 Saarbrücken, Germany. ⁷These authors contributed equally: Georges P. Schmartz, Jacqueline Rehner, Verena Keller, Andreas Keller. ✉e-mail: andreas.keller@ccb.uni-saarland.de

samples²⁸. However, challenges faced during sampling can introduce various confounding factors that may significantly disturb downstream analysis and impact final conclusions. To mitigate some of the previously mentioned challenges, zoo animals, serving as ambassadors for their wild counterparts, yet existing in a controlled environment, have been explored^{29–31}. However, most studies working on captive animals focus only on one animal species and often limit themselves to their gut microbiome^{32–35}.

In this study, we focus on the oral and intestinal microbiota of captive animals, all derived from the same zoo, whose environments are regulated and influenced by the presence of zookeepers and visitors. Our aim is not only to characterize the composition of these microbial communities but also to understand their functional roles, document BGCs, and measure antimicrobial resistance genes. Our investigation extends beyond the confines of captive environments, as we performed a comparative analysis between the microbiomes of captive zoo animals and those of their wild counterparts (Fig. 1a). By scrutinizing the microbiota of these captivated creatures, our study

not only aims to enrich our understanding of the microbiome's complexity but also holds the promise of unearthing novel antimicrobial compounds sourced from animal microbiota.

Results

Deep sequencing and quality control results in 64 metagenomes from 45 species

First, we assessed the quality of the metagenomic sequencing results in light of the diversity of species and sample types included and characterized the robustness of our data. We collected a total of 55 stool and 16 saliva samples, representing an extensive range of 48 and 15 distinct zoo animal species (mammals, birds, and reptiles), respectively (Fig. 1b). Subsequently, after sequencing and quality control, we obtained a final dataset comprising 52 stool and 14 saliva samples, reflecting 45 and 13 species (Fig. 1c). Our quality control measures, including host DNA decontamination, yielded minimal read losses during the process, with an average loss of only 6.6% and a standard deviation (SD) of $\pm 13.2\%$. We retained an average of 5.3 gigabases of sequencing data (SD: 1.7 GB), ensuring a reliable dataset for further analysis.

To account for the species for which a reference assembly was not available on RefSeq, we employed substitute assemblies that were taxonomically close. Notably, this substitution did not significantly impact the relative number of filtered reads (two-sided Wilcoxon p -value of 0.36, Supplementary Data 1), supporting our methodology. Utilizing reference-free ordination analysis, we performed an in-depth examination of the cleaned reads, unveiling distinct patterns of sample clustering primarily based on biospecimen (PERMANOVA p -value < 0.001 , Supplementary Fig. 1). This finding underscores the significance of differentiating between stool and saliva samples and highlights the influence of the animal's specific microbiota on each biospecimen.

De-novo analysis reveals 585 novel genomes and enhances taxonomic assignment

We encountered an expected—yet significant—challenge when performing taxonomic profiling based on the Genome Taxonomy Database (GTDB) (21). The assignment rate using this database was low, with an average of less than 17% (SD: $\pm 16.6\%$) matches. This scarcity of read assignments prompted us to adopt a de-novo analysis workflow. Applying this de-novo analysis workflow proved to be instrumental in overcoming some limitations of the taxonomic profiling from existing databases and uncovered the hidden microbial diversity within our dataset. Through this approach, we successfully recovered a total of 786 dereplicated species-level genome bins (SGBs) exceeding the criteria of at least medium MIMAG quality (namely, less than 10% contamination and a minimum of 50% completeness) (22). Among these SGBs, 585 genomes (74%) had no representatives in the GTDB with ANI (Average Nucleotide Identity) less than 95% (Fig. 2a, Supplementary Fig. 2, Supplementary Data 2). Specifically, when examining the stool samples, we found that out of the 616 dereplicated SGBs, 446 had no representatives (72%). In the case of saliva samples, the ratio increased to 139 out of 170 (82%). Saliva samples, accounting for 21% of the overall samples, contributed 22% and 23% of all the dereplicated SGBs and novel dereplicated SGBs, respectively, suggesting the importance of the oral microbiome in uncovering microbial diversity to be on par with the gut microbiome. Analyzing all the recovered SGBs, we observed an average scaffold length of 13 kb (SD: ± 2.5 kb). Additionally, we conducted searches for tRNA sequences as well as 5S, 16S, and 23S rRNA sequences within the SGBs. In total, 11,801 tRNAs and 205 rRNAs were detected in the SGBs, averaging 15 tRNAs and 0.3 rRNAs per SGB. Whereas these functional gene statistics are indicative of the overall quality of the assemblies, they also highlight the challenges of reliably assembling ribosomal RNA genes.

Importantly, the integration of our SGBs into the GTDB prior to taxonomic profiling yielded a substantial improvement in the read assignment rate (paired two-sided Wilcoxon p -value $< 1.7 \times 10^{-12}$, Supplementary Fig. 3). Nevertheless, for 17 samples, the assignment rate remained below the low threshold of 20%. This highlights the significance of including the novel microbial genomes discovered in this study to enhance the accuracy and comprehensiveness of taxonomic assignments. This analysis is also necessary to assess compositional and functional differences between microbiomes and to uncover the distribution of BGCs.

Culture-based taxonomic assignment yields differences between herbivores and carnivores

In our metagenomic data, the measured alpha diversity, a sign of the microbial complexity of a sample, appears stable for biological replicates. In contrast, the alpha-diversity fluctuates significantly across species (Fig. 3). Astonishingly, we observed a negative Spearman correlation of -0.38 between assignment rates and diversity. Moreover, the reference-based ordination analysis does not yield clear clusters, reflecting neither zoological classification nor diet compositions. Nevertheless, specific zoological proximities are reflected in the clustering hierarchy, such as similar patterns between sheep and goat or between zebra and horse. But in sum, the overall assessment is that the reference-based ordination analysis remains inconclusive with respect to identifying sub-groups of animals. One likely reason for this result is the high variability of assignment rates and missing SGBs. Because differences in the gut microbiota between herbivores and carnivores are known from the literature, we asked whether a more targeted approach involving culturing of bacteria highlights such differences^{36,37}.

Culturing of 11 saliva and 49 stool samples on TSA, Chocolate blood, Columbia, and MacConkey agar, followed by subsequent MALDI-TOF analysis, enabled the identification of 79 different bacterial species (Supplementary Fig. 4, Supplementary Data 3). While we identified a total of 29 species in saliva (37%), only 8 of them (28%) were also detected in stool samples, where 6 of these 8 were of the genus *Staphylococcus*. In total, 32 species (40%) were only detected in the 38 samples of herbivore animals (including species such as *Enterococcus mundtii*, *Bacteroides ovatus*, and *Bacillus pumilus*). In contrast, 8 species (10.1%) were observed only in the 7 samples from carnivore animals (including *Citrobacter braakii*, *Plesiomonas shigelloides*, and *Staphylococcus simulans*). Moreover, 17 bacterial species (21.3%) were uniquely detected in 15 samples of omnivore animals (including *Neisseria zoodegmatis* and *Staphylococcus hominis* depicting the highest frequency across samples). Across all samples, 7 species (8.8%) are present in all three diet forms, including prevalent intestinal microbiota such as *Enterococcus faecalis*, *Escherichia coli*, and *Enterococcus faecium*, as well as *Clostridium perfringens* and *Bacillus cereus*. Before adjustment for multiple hypothesis testing, 11 species were significantly unevenly distributed within the cohorts (χ^2 test p -value < 0.05). After the Benjamini–Hochberg adjustment, no p -value remained significant. Performing the same test over all stool samples and cohorts did not display significant differences between diets (χ^2 test p -value = 0.51). As for culturing, only a selection of media was used, bias is introduced by excluding the growth of certain bacteria, that cannot grow on the selected media. However, all samples were treated the same, which makes these results at least comparable. It is worth mentioning that not every microorganism is cultivatable under laboratory conditions, making the metagenomic analysis a more powerful and more precise tool to investigate the microbiome. Nevertheless, the considerably different repertoire of microbiota suggests unique functional characteristics that might be connected to the dietary origin. We thus performed a functional in-silico gene analysis of the respective microbiota.

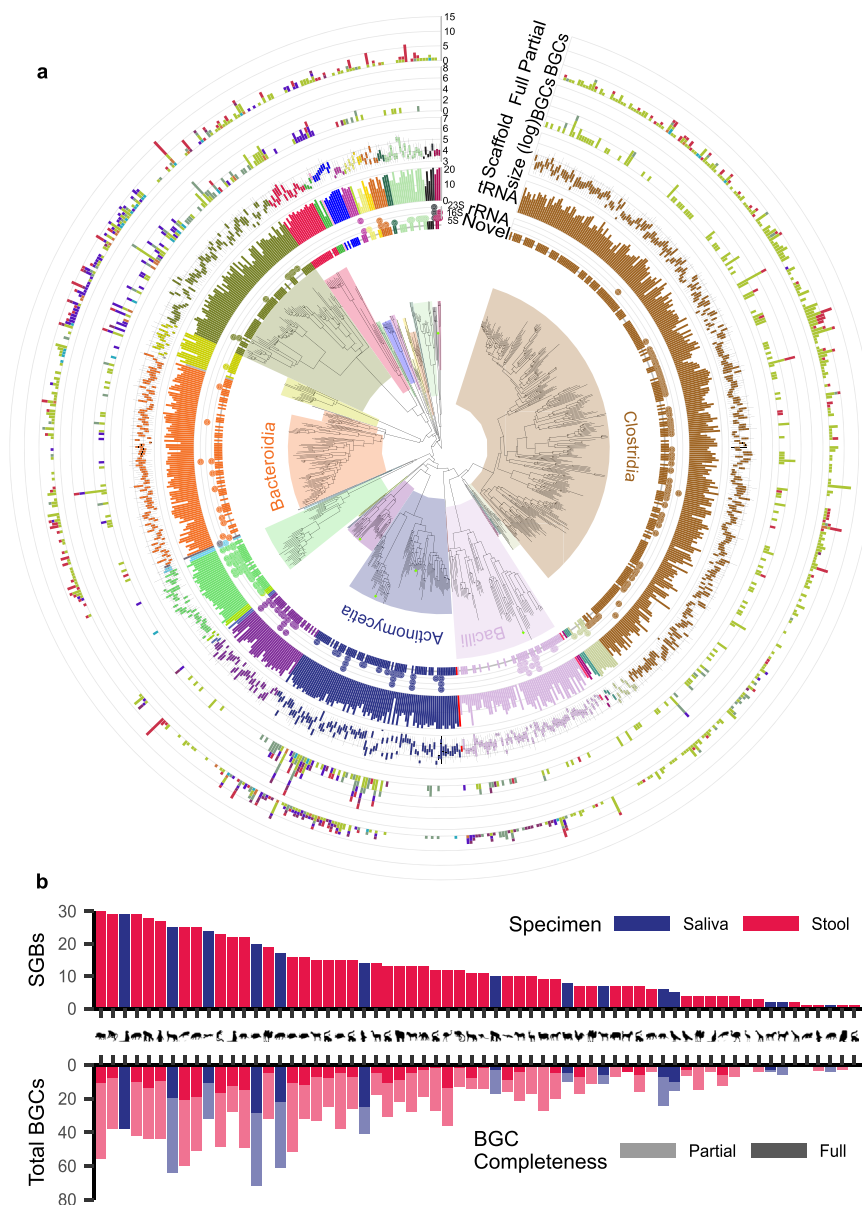


Fig. 2 | Species-level genome bins. **a** Phylogenetic tree of species-level genome bins as classified by the GTDB-Tk. The colored background of clades indicates class ranks. The innermost ring, named *Novel*, indicates if the GTDB-Tk found a species-level assignment. The second, third, and fourth rings discuss bin quality by displaying detected rRNAs, tRNAs, and scaffold length distribution, respectively. The

two outer rings indicate the BGCs that were detected in the respective bins. BGCs are classified by type and by completion. A more richly annotated version of this visualization is available in Supplementary Fig. 2. **b** Number of SGBs and BGCs recovered from each sample.

The statistically significant results from this functional gene analysis highlight elevated creatinine degradation I pathway in herbivore animals (Supplementary Fig. 5). Contrastingly, the super pathway of tetrahydrofolate biosynthesis and salvage are more prevalent in microbiota from carnivore animals. Enriched in both carnivore and omnivore animals are bacteria carrying the genomic information for flavin-dependent thymidylate synthase (*thyX*), which is required to synthesize pyrimidine deoxyribonucleotides de novo. Most notably, this gene and the encoded protein are present in human and animal pathogens, such as *Helicobacter pylori*, *Borrelia burgdorferi*, and *Chlamydia trachomatis*^{38–40}.

Differences in 484 complete biosynthetic gene clusters depending on the diet

After the initial general functional gene analysis of the different animal microbiota, we looked into the specific metabolite landscapes of

individual members of the microbiomes. We performed genome mining of the previously defined SGBs and identified 1588 potential BGCs. Of those, 1104 remained partial, and 484 were identified as full BGC clusters of various categories (Fig. 2a, Fig. 2b, Supplementary Data 4). Further analysis with BiG-SCAPE categorized BGCs into 1482 families, out of which 1407 families were singletons containing only one BGC⁴¹. A total of five families comprising six BGCs are linked to annotated gene clusters from the MIBiG 3.1 database⁴². But interestingly, BiG-SCAPE did not form any clans of the families. Together with a high number of singleton families, this suggests a high diversity of BGCs in the collected dataset.

With 604 (38%) BGCs, *Clostridia* was the class where we predicted most BGCs. However, this is mostly due to *Clostridia* making up about 36% of our recovered dereplicated SGBs. If we look at the average number of BGCs per SGB and exclude singletons, we observe that, on average, most BGCs were predicted for the class of *Planctomycetia*.

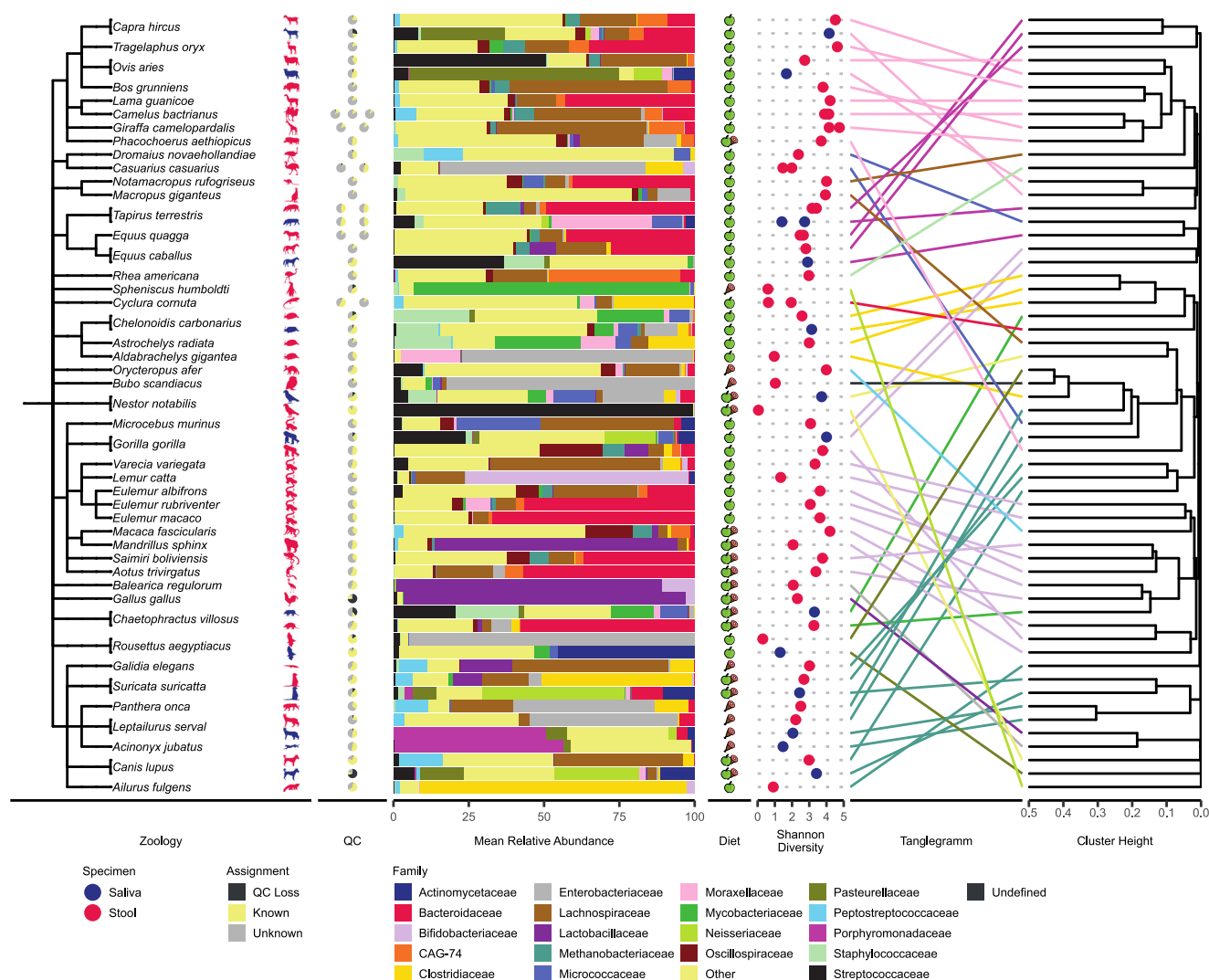


Fig. 3 | Reference-based analysis. Summary statistics on quality, diversity, composition, and compositional similarity of microbiomes. Starting from the left, the taxonomic classification of host animals is displayed. Silhouettes represent the host species, and their color represents the different specimens. If multiple replicates were available, multiple pie charts are displayed, where each pie chart indicates the overall quality of the reference-based analysis. Further, diet classification is provided for each species which is consistently used throughout the paper. Three diets

are being distinguished: herbivore, carnivore, and omnivore. Alpha-diversity of each sample is indicated using the Shannon index, to visualize microbiome complexity. On the rightmost side, hierarchical clustering based on Bray-Curtis distances are displayed. The optimized tanglegram displays the accordance between taxonomic class and membership based on predicted microbial composition. The edges are colored by the taxonomic class of the host.

Averaged over 15 genomes, we observed 3.73 BGCs per SGB. With only 4 BGCs in 33 SGBs, *Saccharimonadia* had the lowest non-singleton ratio of BGCs to SGBs. Concerning disparities between the oral and gut microbiome, we observed a total of 450 BGCs (28%) in the 170 saliva-derived SGBs averaging 2.65 BGCs per SGB, which compares to 1,138 BGCs in 616 SGBs at a ratio of 1.85 in the stool samples. Focusing only on the stool-derived BGCs, we observed an average of 2.01, 1.65, and 1.47 BGCs per SGB for herbivores, omnivores, and carnivores, respectively. These differences were confirmed to be significant (Kruskal-Wallis p -value < 0.0053). Specifically, the average number of nonribosomal peptides (NRPs) was 2.87 and 3.46 times higher in herbivore SGBs compared to carnivores and omnivores, respectively.

Through a comparative analysis of predicted BGCs and known annotated BGCs from the MIBiG database, we observed 37 BGCs (2%) within our SGBs that shared a similarity of over 50% with known entries. Among these annotations, various

compounds may be of relevance to the host organism (Supplementary Fig. 6). We detected virulence factors, such as the toxin *toluasin I*, within an SGB derived from tapir saliva. Furthermore, we uncovered various annotations associated with health benefits, including the bacteriocin *salivaricin CRL 1328*, present in an SGB derived from a mandrill stool sample⁴³. We encountered two further compounds with noteworthy properties: α -galactosylceramide, an immunostimulating compound found in an SGB derived from horse stool, and *rhizomide*, identified in an SGB derived from tapir saliva, exhibiting anti-tumor and antimicrobial properties *in vitro*⁴⁴.

Having captured differences in the repertoire of bacteria from animals with different diets in the gut and oral cavity along with unique functional characteristics and novel BGCs raises the question of whether captivity has an influence on the microbiota or whether wildlife animals reveal similar patterns.

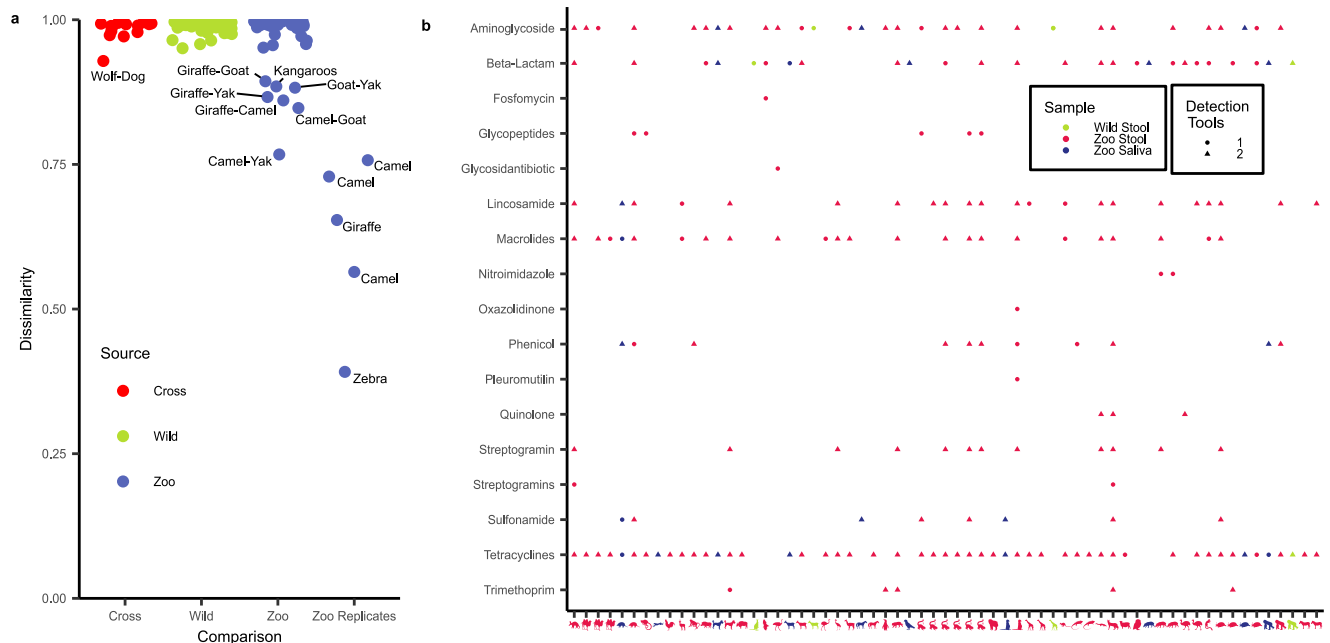


Fig. 4 | Potential consequences of captivity. **a** FracMinHash dissimilarity between samples within our dataset and the dataset of Youngblut et al.²⁸. The cross-comparison matches sample pairs as elaborated in Supplementary Data 1. For reference, zoo replicates and their dissimilarity are visualized alongside. **b** Presence

of antimicrobial resistance genes for each of the zoo and wildlife samples classified by antimicrobial compound class. Visualized results derive from ResFinder and AMRFinderPlus.

Animals in captivity present different antimicrobial resistance gene patterns

Comparing microbiome differences between captive and wildlife animals and addressing the complexities in the sample extraction process, we conducted a comparative analysis with data from Youngblut et al.²⁸, the—as of now—most complete study of animal gut microbiota. Their dataset consisted of 289 samples from 180 different host species, including humans. The large differences between both studies in the selection of animal species call for a balanced and stratified analysis approach. Therefore, we implemented a matching scheme that carefully selects a subset of samples with close zoological similarity from both studies (Supplementary Data 1). We excluded the oral microbiomes of the zoo animals from this analysis because no oral microbiota from wildlife animals were present.

It is important to acknowledge that differences in sample processing between the two studies, such as different DNA extraction kits, can be reflected in the data⁴⁵. Similarly, the easier collection process in a controlled environment, such as a zoo in comparison to a wildlife setting, likely leads to differences in the sample quality. To quantify these differences and ensure methodological consistency, we thus applied our analysis workflow to the selected metagenomes from Youngblut et al.²⁸. We observed a significant decrease in read quantity after decontamination compared to the present data, which is explained by the above-mentioned challenges in wildlife sampling (Supplementary Fig. 7a). This reduction also influences the assembly quality, which was lower in the wildlife samples, finally leading to overall shorter fragments (Supplementary Fig. 7b). Consequently, fewer SGBs were recovered in the wildlife samples compared to the zoo dataset (Supplementary Fig. 7c). While the samples from animals kept in captivity retained an average of 9.8 SGBs per sample, the wildlife dataset yielded just 1.9 SGBs. Similar differences also apply to the number and abundance distribution of BGCs. Here, SGBs derived from wildlife animals present, on average, 13 fewer BGCs per SGB (Supplementary Fig. 7d). We were only able to recover partial BGCs in the wildlife samples compared to 50 complete BGCs in the matching zoo samples. Further, only one BGC was annotated to have a similarity

>10% to any known MIBiG BGC. It has a 28% similarity to a carotenoid cluster derived from an *Algoriphagus* species. Again, the latter results might seem counterintuitive, and we might expect more BGCs in wildlife, yet the results are likely biased by the challenges of wildlife sampling and different sample processing. Most importantly, the quality of the wildlife samples is still sufficient to enable reference-free comparison.

As one first aspect, we asked whether the microbiomes between zoo and wildlife animals present a conserved proximity-dependent on the relatedness of host animal species. For the selected samples, we thus performed reference-free FracMinHash comparisons (Fig. 4a). On average, we computed a large dissimilarity between any compared pairs. In detail, the average dissimilarity amounts to 0.98 (SD: 0.018), which is close to the maximal dissimilarity value of 1. Importantly, the dissimilarity distributions within the wildlife and zoo animals do not differ significantly (two-sided Wilcoxon p -value > 0.37). Nevertheless, zoo animals display several strong similarities between gut microbiota. These include mostly inter-replicate comparisons of zebra, camel, and giraffes, yielding an overall significantly lower dissimilarity index as compared to the other zoo animals (two-sided Wilcoxon p -value < 9.44×10^{-7}). Of note, no replicates for the wildlife animals are available, explaining the missing similarities within those samples. Interestingly, several of the zoo animal species, including the yak, giraffe, camel, and goat, displayed increased similarities in gut microbiota. The same applies to two kangaroo species that also show similarities in the gut microbiota. Of note, such similarities are not present in the wildlife animals and may suggest an influence, e.g., of the nutrition in this controlled environment. Further, the results clearly argue for combining the advantages of studies in wildlife animals (being closer to nature) and controlled environments (facilitating higher sample quality).

One immediate question in comparing wildlife to captivity setups concerns the presence of AMR. AMR gene analysis of zoo animals revealed potential resistances against antibiotics that are commonly used in veterinary medicine, such as tetracyclines, macrolides, and lincosamides, with *tetW* being the most prominent

gene detected encoding for resistance against tetracyclines, and *lnu(C)* being the most frequent gene encoding for lincosamide resistance when looking at the results using AMRFinderPlus⁴⁶ (Fig. 4b, Supplementary Data 5). Genes conferring resistance to fluoroquinolones were rare (*qnrS1*). However, we also observed resistance genes against vancomycin, a broad spectrum antibiotic against Gram-positive bacteria⁴⁷. Specifically, we documented the well-known resistance clusters *vanD* and *vanG*, which are uncommon in humans^{48,49}. However, we also detected the *vanO* operon, which has only been identified in captive elephants in Africa^{50,51}. Furthermore, we observed a high number of genes encoding beta-lactamases (e.g., *blaEC* and *blaTEM*), some of which may even confer resistance to carbapenems. We were able to detect 21 different genes encoding for resistance mechanisms against aminoglycosides, with the most prominent one being *aadE* when looking at the results generated with AMRFinderPlus. We stratified all beta-lactamases according to the Ambler classification and found most prominently beta-lactamases belonging to Ambler class A and C. We did not observe any beta-lactamases belonging to Ambler class B or D. As outlined in the 'One Health' concept, such resistant bacteria could be transferred from zoo animals to zookeepers, increasing the global spreading of such organisms. When we compared our matching stool zoo samples to the wildlife samples, we observed a significantly smaller number of antimicrobial compound classes that are targeted by at least one resistance gene in the wildlife samples (two-sided Wilcoxon *p*-value < 0.036). Overall, we only observed a total of five potential resistances in all analyzed wildlife samples. This suggests that wild animals overall harbour less AMR. Nevertheless, we want to highlight that this result is again to be interpreted in the light of the inferior assembly quality of the wildlife samples, which impacts the quality of AMR gene detection.

Discussion

Our findings, in line with the study by Youngblut et al.²⁸, indicate that the microbial dark matter within animal microbiomes remains inadequately characterized in existing data repositories. Despite our extensive efforts and the generation of several novel SGBs, we encountered 17 samples with a low estimated assignment rate below 20%. This deficiency significantly impacts state-of-the-art reference-based analysis, as evident in our own investigation.

The microbial richness we detect, despite the accompanying challenges, presents an intriguing opportunity for the discovery of BGCs associated with antimicrobial natural compounds within these samples. In this context, it is worth emphasizing the advantages of combining different study setups. While our focus lies on samples from a highly controlled environment, specifically a zoo, complementary studies like that of Youngblut et al.²⁸, provide valuable insights into wildlife microbiomes, which are closer to the natural microbiota. By integrating findings from diverse settings, we can gain a more comprehensive understanding of the animal microbiome and potentially uncover novel microbial resources with therapeutic potential.

Specifically, the zoo animals present higher numbers of SGBs and BGCs per SGBs but also higher proximity of gut microbiota as compared to the wildlife animals. It is important to acknowledge that the number of BGCs within SGBs can vary, depending on the specific species discovered. However, the improved assembly statistics highlight the advantages of easier sample collection in captivity compared to wild animals, at the cost of BGCs that might only be present in wildlife animals.

When comparing studies, one limitation we encountered was the need to perform inter-species comparisons, which involved species from different continents with potentially diverse diets. This aspect adds complexity to the analysis, as the microbiomes of zoo animals, despite sharing similar diets such as local seasonal vegetables, still

exhibit considerable differences. The convergence of microbiome composition across zoo animals appears to be limited, yet measurable.

Furthermore, the presence of AMR genes in animal microbiomes is of considerable importance from the 'One Health' perspective. A previous study, for example, screened captive animals in a zoo in Seoul, South Korea, for particular AMR patterns carried by *Escherichia coli* and *Enterococcus faecalis*. However, the research assessed resistance phenotypically only without analyzing the responsible genes. They found ampicillin resistance in most *E. coli* isolates, and also described multidrug resistance in 50% of isolated *E. coli*⁵². As we wanted to monitor all AMR genes found in captive animals, we followed a metagenomic approach. While it is not uncommon to detect antibiotic resistance genes both in wildlife animals and in zoo animals, the distribution of these genes is of enormous relevance, e.g., to track possible associations between commonly used antibiotics in veterinary and/or human medicine and to decipher potential transmission chains. Also, resistance towards 'last-line antibiotics' in animals might constitute a potential threat to humans⁵³. In that regard, it is important to note that we also identified resistance genes against vancomycin in certain animals, including prosimians. Considering their close contact with zookeepers, there is a potential risk of transferring vancomycin-resistant bacteria to humans. A closer look into resistances against beta-lactam antibiotics revealed mainly beta-lactamases belonging to Ambler class A and C, such as *blaEC* and *blaTEM*. Recent findings from studies conducted in Africa revealed mainly *blaOXA*, *blaKPC*, *blaNDM*, *blaSHV*, and *blaVIM* to be found in animals, food, and environmental samples⁵⁴. These genes encoding for carbapenemases were absent in our study but are also becoming a major threat in human medicine^{55–57}, and hence, are listed among the World Health Organization's (WHO) 'priority pathogen list in the highest category' as 'critical'⁵⁸. Therefore, longitudinal screening of captive animals which are in close contact to humans should be employed to notice such a trend before the spread of bacteria carrying such resistances cannot be stopped. The comparatively frequent detection of resistance to tetracycline antibiotics is a further concern and future studies should ideally employ strategies to document the previous antibiotic intake of zoo animals to gain insights on potential associations with a rise in resistance rates. Indeed, as transmission of multi-resistant bacteria has been observed in clinical settings, our findings emphasize the need for comprehensive surveillance and management of AMRs in zoo settings to mitigate potential health risks and maintain a safe environment for both animals and humans^{59–61}. Our AMR analysis is limited by the fact that individual resistance genes could not be assigned to specific bacterial species; e.g., *Enterobacter cloacae* complex frequently carries AmpC beta-lactamases in clinical practice, and it would have been interesting to see whether such associations also hold true for captive animals⁶².

Methods

Study design

For docile animals such as horses, dwarf goats, and tapirs, buccal swabs were easily taken from the oral cavity to collect saliva samples. Concurrently, fresh fecal samples were collected from the enclosures or stables maximum two hours after defecation and immediately transported to the veterinary station. Using a spoon from a stool sample tube, feces from the inner portion of the excreta were transferred into sample tubes. Subsequently, all samples were promptly frozen at -20 °C in the freezer compartment of a refrigerator. Typically, samples were frozen within 30 min of collection.

For non-docile animals, such as primates and large or small carnivores, the same sample collection methods were employed during necessary anesthesia, which occurred for veterinary examinations, treatment, transport, or sex determination. For small animals, fecal samples were collected rectally as swabs, following the same protocol described above, and stored frozen until further analysis. Due to the non-invasive sampling procedure, no ethical approval was required.

DNA extraction

We extracted whole-genome DNA from all fecal and salivary swabs using the Qiagen QiAamp Microbiome Kit (Qiagen, Hilden, Germany)⁸. The DNA extraction procedure was conducted according to the manufacturer's protocol. Briefly, all swabs containing native samples were vortexed in 1 ml PBS for 2 minutes. The PBS containing the microbes from each sample was then used for DNA extraction according to the manufacturer's recommendation. We used the MP Biomedicals™ FastPrep-24™ 5 G Instrument (FisherScientific GmbH, Schwerte, Germany) for mechanical lysis of bacterial cells. The velocity and duration were adjusted to the 'hard-to-lyse' protocol, meaning 6.5 m/s for 45 s 2 times and 5 min storage on ice in between each lysis step. DNA was eluted in 50 µl elution buffer. The DNA concentration after elution was determined via NanoDrop 2000/2000c (ThermoFisher Scientific, Wilmington, DE) full-spectrum microvolume UV–Vis measurements⁴⁵.

Library preparation and sequencing

Extracted whole-genome DNA was sent to Novogene Company Limited (Cambridge, UK) for library preparation and sequencing. For quality control of the samples, potential genomic DNA degradation was measured using the fragment analyzer platform AATI (Agilent Technologies, CA, USA). The DNA concentration was measured using Qubit (Thermo Fisher, Wilmington, DE) before library preparation. Briefly, samples were subjected to metagenomic library preparation and further sequenced via paired-end Illumina Novaseq X plus Sequencing PE150. For library preparation, the Novogene NGS DNA Library Prep Set (Cat No.PT004) was used. Genomic DNA was sheared into short fragments in random positions and fragmented DNA was subjected to end-repair and A-tailing, as well as Illumina adapter ligation. Fragments with the appropriate size of 500 bp were selected via beads-based size selection of libraries and amplified via PCR. The PCR products underwent quality control and quantification using the Qubit system and bioanalyzer to visualize the generated fragment sizes. All samples were pooled and sequenced on the Illumina Novaseq X plus Sequencer. For all samples, 5 Gb reads per sample were generated.

Culturing of bacteria

Native fecal samples were streaked out using the swab they were taken with, on three different agar plates: TSA with 5% sheep blood (TSA), MacConkey (MC), and Columbia (Co) agar plates (Becton, Dickinson and Company, Heidelberg, Germany). Oral samples were streaked out on TSA, Co and Chocolate blood (CB) agar plates (Becton, Dickinson and Company, Heidelberg, Germany). All TSA, CB, and MC agar plates were incubated at 35.6 °C and 5% CO₂ for a minimum of 18 h and a maximum of 24 h. Co agar plates were used for the cultivation of anaerobic bacteria and therefore incubated in an anaerobic environment for a minimum of 48 h at 35.6 °C⁴⁵.

Mass spectrometry-based identification

Bacterial colonies obtained by culturing native fecal and oral samples on different agar plates were subjected to species identification using matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry. To this end, colonies from overnight growth on tryptic soy agar plates containing 5% sheep blood, from Columbia agar plates, and MacConkey agar plates were taken with a sterile toothpick and spotted on the MALDI-TOF target plate by smearing one colony on one spot of the target, dried, and then overlaid with 1 µl of 70% formic acid. This step aids in the cell lysis and makes peptides and proteins available for ionization. After drying, 1 µl of α-cyano-4-hydroxycinnamic acid (CHCA) matrix solution (Bruker Daltonics, Bremen, Germany) was pipetted on top of the bacterial matter and formic acid and set to dry. The matrix solution is composed of saturated CHCA dissolved in 50% (v/v) acetonitrile, 47.5% (v/v) LC-MS grade water, and 2.5% (v/v) trifluoroacetic acid. After drying the matrix solution at room temperature, each spot was overlaid with 70 %

formic acid to pre-disrupt the cells. Followed by drying at room temperature, the plate was placed into the Microflex LT Mass Spectrometer (Bruker Daltonics) for MALDI-TOF MS. All measurements were performed with the AutoXecute algorithm in the FlexControl© software version 3.4 (Bruker Daltonics). Each spot was excited with 240 laser shots in six random positions. Measurements were carried out automatically to generate protein mass profiles in linear positive ion mode using a laser frequency of 60 Hz, high voltage of 20 kV, and pulsed ion extraction of 180 ns. Mass charge ratio ranges (m/z) were measured between 2 kDa and 20 kDa. We identified bacterial species using the software MALDI BioTyper compass explorer (v.3.0). The database used was Bruker's commercial database: Bruker BDAL database (10,148 species-specific main spectra profiles). Identification scores above 2.0 were considered a precise identification of proteins and peptides on the species level, scores between 1.7 and 1.99 were considered as possible species identification and precise genus identification, and all identification scores below 1.7 were considered unsuccessful identification. In this study, we only considered scores ≥2 for analyses³⁶.

Next-generation sequencing preprocessing

The first step of data analysis was host read removal with KneadData (version (v):0.10.0; command line arguments (cla): "--trimmomatic-options='LEADING:3 TRAILING:3 MINLEN:50' --bowtie2-options='--very-sensitive --no-discordant --reorder'"") using the respective genomes as specified in Supplementary Data 1⁶³. The selected, publicly available, host genomes were downloaded with the ncbi-datasets-cli (v13.35.0). For several animal species, no exact sequenced genome of sufficient quality was available and instead, a taxonomically close substitute was selected. Bowtie2 (v2.4.5; -s) databases were prepared for each reference⁶⁴. After decontamination, we performed sequence overrepresentation analysis and quality assurance with fastp (v:0.23.2; cla: "--overrepresentation_analysis") and visualized results with MultiQC (v1.13a)^{65,66}. The two-sided Wilcoxon rank sum test was performed on the relative loss attributed to host DNA removal. To reduce bias, replicates were averaged. Saliva and stool samples were not averaged.

Metagenome assembly

We assembled each sample with SPAdes (v3.15.4; cla: --meta) and monitored assembly quality with QUAST (v5.0.2; cla: -s)^{67,68}. Next, we aligned each host decontaminated sample against each set of assembled scaffolds with BWA-MEM2 (v2.2.1) and generated abundance profiles for each combination⁶⁹. We extracted coverage information to bin scaffolds with MetaBAT2 (v2.15; cla: -l --seed 420 --unbinned), MaxBin2 (v2.2.7), and DAS Tool(v1.1.5; --search_engine diamond)^{70–72}. MAGs across all samples were aggregated and dereplicated with dRep (v:3.4.0; cla: -comp 50 -con 10 --checkM_method lineage_wf --S_algorithm fastANI --S_ani 0.95 -nc 0.5). At last, we used GTDB-Tk (v:2.1.1; cla: classify_wf), tRNAscanSE (v:2.0.11; --brief -Q), and barrnap(v:0.9; cla: -q) to taxonomically classify MAGs and annotate them with tRNA and rRNA information based on their classified kingdom^{73,74}.

Reference-based compositional analysis

FracMinHash profiles were computed for all samples with sourmash (v:4.4.3; cla: -k51)⁷⁵. After FracMinHash profile generation, samples were compared with sourmash compare. Dissimilarities were computed by subtracting the resulting similarities from one. Samples were embedded with UMAP (v:0.2.8)⁷⁶. Further, for each SGB, FracMinHash profiles were computed as well, and an index was generated. The PERMANOVA analysis treated samples and replicates as independent⁷⁷. Taxonomic profiling was performed with sourmash (cla: -k51) our previously generated indices, GTDB (v:GTDB R07-RS207 all genomes k51), and host decontaminated reads. Shannon index was used as the alpha-diversity measure and computed with phyloseq (v:1.40.0)^{78,79}. Relative abundances were averaged if replicates were available. Clustering was

performed with average hierarchical clustering on Bray-Curtis distances computed with the *vegan* package on mean relative abundances (v:2.6.2)⁸⁰. Tanglegram was optimized for visual clarity with “step2side” algorithm of the R *dendextend* package (v:1.16.0)⁸¹. Differential abundance analysis was made with ANCOMBC (v:1.6.2) comparing herbivores and the union of omnivores and carnivores⁸².

Functional analysis

In order to incorporate our own SGBs into the functional profiling step, we updated an existing GTDB207-based database with Struo2 (v:2.3.0)⁸³. After database generation, functional profiling was performed with HUMAnN 3 (v:3.6; cla: --bypass-nucleotide-index)⁶³. We also used ANCOMBC for exploration of differences in function. The default setting of Holm–Bonferroni *p*-value adjustment was employed. Genes were predicted with prodigal (cla: -p meta) and passed to antiSMASH (v:6.1.1; cla: -cb-knownclusters --cb-subclusters --asf) for BGC detection^{22,84}. A BGC was classified as partial if it is shorter than 5 kbp or located on a contig edge and as full otherwise. Clustering of all BGCs was performed with BiG-SCAPE (v:1.1.5; cla: --mibig) using Pfam (v:35.0). BiG-SCAPE failed to process two BGCs and removed them from further analysis^{42,85}.

Antimicrobial resistance gene analysis

Antimicrobial resistance gene assessment was performed with AMR-FinderPlus (v:3.11.26; database v:2023-11-15.1; cla: --report_all_equal --plus --coverage_min 0.9 --ident_min 0.95), DeepARG (v:1.0.4; database v:2; cla: --model SS --type nucl --min-prob 0.8 --arg-alignment-identity 95 --arg-alignment-evalue 1e-10 --arg-num-alignments-per-entry 1000), and ResFinder (v:4.4.2; database v:2.3.0; clr: --threshold 0.95 --min_cov 0.95). After metagenomic assembly, the contigs of each sample were passed to each of the aforementioned tools, grouping contigs by sample. Results across all samples, as well as tools were aggregated using hamronizer (<https://github.com/pha4ge/hAMRonzation>; v:1.1.4)^{86–88}. DeepARG predictions were rejected from further analysis due to exceptionally high numbers of detected resistance genes and divergence from manually inspected results. AMR-FinderPlus predictions were discussed in detail in the paper. For visualization, ResFinder and AMR-FinderPlus results were unified if genes that provide resistance against the same group of antimicrobial compounds were predicted on the same contig within a 50 bp interval by both tools.

Wildlife comparison

Samples specified in Supplementary Data 1 were downloaded from the European Nucleotide Archive and processed identically to our dataset, from host DNA removal to BGC prediction⁸⁹. We subsetted our data to only the paired samples specified in the aforementioned table. Pairings were manually selected based on taxonomic similarity. Paired comparison to our data was done based on FracMinHash dissimilarities.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw unfiltered sequencing reads as well as dereplicated SGBs generated in this study have been deposited in the Sequence Read Archive under the accession [PRJNA983076](https://www.ncbi.nlm.nih.gov/sra/PRJNA983076).

References

- Gacesa, R. et al. Environmental factors shaping the gut microbiome in a Dutch population. *Nature* **604**, 732–739 (2022).
- Tun, H. M. et al. Exposure to household furry pets influences the gut microbiota of infant at 3–4 months following various birth scenarios. *Microbiome* **5**, 40 (2017).
- Kuthyar, S. & Reese, A. T. Variation in microbial exposure at the human-animal interface and the implications for microbiome-mediated health outcome. *mSystems* **6**, e0056721 (2021).
- Aggarwal, D. & Ramachandran, A. One health approach to address zoonotic diseases. *Indian J. Commun. Med.* **45**, S6–S8 (2020).
- Integrative HMP RNC. The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. *Cell Host Microbe* **16**, 276–289 (2014).
- Becker, A. et al. Effects of resistant starch on symptoms, fecal markers, and gut microbiota in Parkinson’s disease—the RESISTA-PD trial. *Genomics Proteom. Bioinforma.* **20**, 274–287 (2022).
- Rup, L. The human microbiome project. *Indian J. Microbiol.* **52**, 315 (2012).
- Djordjevic, S. P. et al. Genomic surveillance for antimicrobial resistance—a One Health perspective. *Nat. Rev. Genet.* **25**, 142–157 (2024).
- Aljeldah, M. M. Antimicrobial Resistance and Its Spread Is a Global Threat. *Antibiotics (Basel)* **11**, 1082 (2022).
- Berglund, F. et al. Identification and reconstruction of novel antibiotic resistance genes from metagenomes. *Microbiome* **7**, 52 (2019).
- Van Boeckel, T. P. et al. Global trends in antimicrobial resistance in animals in low- and middle-income countries. *Science* **365**, eaaw1944 (2019).
- Rosch, M. E. G. et al. Time series of chicken stool metagenomics and egg metabolomics in changing production systems: preliminary insights from a proof-of-concept. *One Health Outlook* **6**, 4 (2024).
- Ikhimiukor, O. O., Odih, E. E., Donado-Godoy, P. & Okeke, I. N. A bottom-up view of antimicrobial resistance transmission in developing countries. *Nat. Microbiol.* **7**, 757–765 (2022).
- Zhao, C., Wang, Y., Mulchandani, R. & Van Boeckel, T. P. Global surveillance of antimicrobial resistance in food animals using priority drugs maps. *Nat. Commun.* **15**, 763 (2024).
- Palmeira, J. D. et al. Emergence and spread of cephalosporinases in wildlife: a review. *Animals* **11**, 1765 (2021).
- Benavides, J. A., Salgado-Caxito, M., Torres, C. & Godreuil, S. Public health implications of antimicrobial resistance in wildlife at the one health interface. *Med. Sci. Forum* **25**, 1 (2024).
- García-Gutiérrez, E., Mayer, M. J., Cotter, P. D. & Narbad, A. Gut microbiota as a source of novel antimicrobials. *Gut Microbes* **10**, 1–21 (2019).
- Behsaz, B. et al. Integrating genomics and metabolomics for scalable non-ribosomal peptide discovery. *Nat. Commun.* **12**, 3225 (2021).
- Galvan, A. E. et al. Identification of the biosynthetic gene cluster for the organoarsenical antibiotic arsinothricin. *Microbiol. Spectr.* **9**, e0050221 (2021).
- Martinet, L. et al. A single biosynthetic gene cluster is responsible for the production of bagremycin antibiotics and ferroverdin iron chelators. *mBio* **10**, e01230–19 (2019).
- Miethke, M. et al. Towards the sustainable discovery and development of new antibiotics. *Nat. Rev. Chem.* **5**, 726–749 (2021).
- Blin, K. et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res.* **49**, W29–W35 (2021).
- Scherlach, K. & Hertweck, C. Mining and unearthing hidden biosynthetic potential. *Nat. Commun.* **12**, 3864 (2021).
- Mullowney, M. W. et al. Artificial intelligence for natural product drug discovery. *Nat. Rev. Drug Discov.* **22**, 895–916 (2023).
- Peixoto, R. S. et al. Harnessing the microbiome to prevent global biodiversity loss. *Nat. Microbiol.* **7**, 1726–1735 (2022).
- Donia, M. S. et al. A systematic analysis of biosynthetic gene clusters in the human microbiome reveals a common family of antibiotics. *Cell* **158**, 1402–1414 (2014).

27. Charlop-Powers, Z. et al. Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proc. Natl Acad. Sci. USA* **113**, 14811–14816 (2016).
28. Youngblut, N. D. et al. Large-scale metagenome assembly reveals novel animal-associated microbial genomes, biosynthetic gene clusters, and other genetic diversity. *mSystems* **5**, e01045–20 (2020).
29. de Jonge, N., Carlsen, B., Christensen, M. H., Pertoldi, C. & Nielsen, J. L. The gut microbiome of 54 mammalian species. *Front. Microbiol.* **13**, 886252 (2022).
30. McKenzie, V. J. et al. The effects of captivity on the mammalian gut microbiome. *Integr. Comp. Biol.* **57**, 690–704 (2017).
31. Chong, R., Cheng, Y., Hogg, C. J. & Belov, K. Marsupial gut microbiome. *Front. Microbiol.* **11**, 1058 (2020).
32. Guo, W. et al. Comparative study of gut microbiota in wild and captive giant pandas (*Ailuropoda melanoleuca*). *Genes (Basel)* **10**, 827 (2019).
33. Jia, T. et al. Characterization of the gut microbiome and resistomes of wild and zoo-captive macaques. *Front. Vet. Sci.* **8**, 778556 (2021).
34. San Juan, P. A., Castro, I. & Dhami, M. K. Captivity reduces diversity and shifts composition of the Brown Kiwi microbiome. *Anim. Microbiome* **3**, 48 (2021).
35. Sun, C. H., Liu, H. Y., Liu, B., Yuan, B. D. & Lu, C. H. Analysis of the gut microbiome of wild and captive pere David's deer. *Front. Microbiol.* **10**, 2331 (2019).
36. Rehner, J. et al. The effect of a planetary health diet on the human gut microbiome: a descriptive analysis. *Nutrients* **15**, 1924 (2023).
37. Nishida, A. H. & Ochman, H. Rates of gut microbiome divergence in mammals. *Mol. Ecol.* **27**, 1884–1897 (2018).
38. Zhong, J., Skouloubris, S., Dai, Q., Myllykallio, H. & Barbour, A. G. Function and evolution of plasmid-borne genes for pyrimidine biosynthesis in *Borrelia* spp. *J. Bacteriol.* **188**, 909–918 (2006).
39. Sodolescu, A. et al. Structural and functional insight into serine hydroxymethyltransferase from *Helicobacter pylori*. *PLoS ONE* **13**, e0208850 (2018).
40. Escartin, F., Skouloubris, S., Liebl, U. & Myllykallio, H. Flavin-dependent thymidylate synthase X limits chromosomal DNA replication. *Proc. Natl Acad. Sci. USA* **105**, 9948–9952 (2008).
41. Bowers, R. M. et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **35**, 725–731 (2017).
42. Navarro-Munoz, J. C. et al. A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
43. Vera Pingitore, E., Hebert, E. M., Nader-Macias, M. E. & Sesma, F. Characterization of salivaricin CRL 1328, a two-peptide bacteriocin produced by *Lactobacillus salivarius* CRL 1328 isolated from the human vagina. *Res Microbiol* **160**, 401–408 (2009).
44. Wang, X. et al. Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in Burkholderiales species. *Proc. Natl Acad. Sci. USA* **115**, E4255–E4263 (2018).
45. Rehner, J. et al. Systematic cross-biospecimen evaluation of DNA extraction kits for long- and short-read multi-metagenomic sequencing studies. *Genomics Proteom. Bioinforma.* **20**, 405–417 (2022).
46. De Briyne, N., Atkinson, J., Pokludova, L. & Borriello, S. P. Antibiotics used most commonly to treat animals in Europe. *Vet. Rec.* **175**, 325 (2014).
47. Wilhelm, M. P. Vancomycin. *Mayo Clin. Proc.* **66**, 1165–1170 (1991).
48. Depardieu, F., Kolbert, M., Pruul, H., Bell, J. & Courvalin, P. VanD-type vancomycin-resistant *Enterococcus faecium* and *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **48**, 3892–3904 (2004).
49. McKessar, S. J., Berry, A. M., Bell, J. M., Turnidge, J. D. & Paton, J. C. Genetic characterization of vanG, a novel vancomycin resistance locus of *Enterococcus faecalis*. *Antimicrob. Agents Chemother.* **44**, 3224–3228 (2000).
50. Gudeta, D. D., Moodley, A., Bortolaia, V. & Guardabassi, L. vanO, a new glycopeptide resistance operon in environmental *Rhodococcus equi* isolates. *Antimicrob. Agents Chemother.* **58**, 1768–1770 (2014).
51. Feng, X. et al. Comparison of the gut microbiome and resistome in captive African and Asian elephants on the same diet. *Front. Vet. Sci.* **10**, 986382 (2023).
52. Kim, M., Kim, M., Yeo, Y. G., Lee, Y. T. & Han, J. I. Antimicrobial resistance of commensal *Escherichia coli* and *Enterococcus faecalis* isolated from clinically healthy captive wild animals in Seoul zoo. *Front. Vet. Sci.* **10**, 1283487 (2023).
53. Olaru, I. D., Walther, B. & Schaumburg, F. Zoonotic sources and the spread of antimicrobial resistance from the perspective of low and middle-income countries. *Infect. Dis. Poverty* **12**, 59 (2023).
54. Dossouvi, K. M. & Ametepe, A. S. Carbapenem resistance in animal-environment-food from Africa: a systematic review, recommendations and perspectives. *Infect. Drug Resist.* **17**, 1699–1728 (2024).
55. Hou, C. & Yang, F. Drug-resistant gene of blaOXA-23, blaOXA-24, blaOXA-51 and blaOXA-58 in *Acinetobacter baumannii*. *Int. J. Clin. Exp. Med.* **8**, 13859–13863 (2015).
56. Bassetti, M. et al. Management of KPC-producing *Klebsiella pneumoniae* infections. *Clin. Microbiol. Infect.* **24**, 133–144 (2018).
57. Kumari, M. et al. Emergence of blaNDM-1 and blaVIM producing Gram-negative bacilli in ventilator-associated pneumonia at AMR Surveillance Regional Reference Laboratory in India. *PLoS ONE* **16**, e0256308 (2021).
58. Tacconelli, E. et al. Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).
59. Thorpe, H. A. et al. A large-scale genomic snapshot of *Klebsiella* spp. isolates in Northern Italy reveals limited transmission between clinical and non-clinical settings. *Nat. Microbiol.* **7**, 2054–2067 (2022).
60. van Schaik, W. Baas Becking meets One Health. *Nat. Microbiol.* **7**, 482–483 (2022).
61. Parajuli, A. et al. Drivers of antimicrobial resistance within the communities of Nepal from One Health perspective: a scoping review. *Front. Public Health* **12**, 1384779 (2024).
62. Guerin, F., Isnard, C., Cattoir, V. & Giard, J. C. Complex regulation pathways of AmpC-mediated beta-lactam resistance in *Enterobacter cloacae* complex. *Antimicrob. Agents Chemother.* **59**, 7753–7761 (2015).
63. Beghini, F. et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *Elife* **10**, e65088 (2021).
64. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
65. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
66. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
67. Nurk, S., Meleshko, D., Korobeynikov, A. & Pevzner, P. A. metaSPAdes: a new versatile metagenomic assembler. *Genome Res.* **27**, 824–834 (2017).
68. Mikheenko, A., Saveliev, V. & Gurevich, A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics* **32**, 1088–1090 (2016).
69. Vasimuddin M., Misra S., Li H., Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) (2019).

70. Kang, D. D. et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
71. Wu, Y. W. & Singer, S. W. Recovering individual genomes from metagenomes using MaxBin 2.0. *Curr. Protoc.* **1**, e128 (2021).
72. Sieber, C. M. K. et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* **3**, 836–843 (2018).
73. Chaumeil, P. A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* **38**, 5315–5316 (2022).
74. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096 (2021).
75. Irber, L. et al. Lightweight compositional analysis of metagenomes with FracMinHash and minimum metagenome covers. *bioRxiv* <https://doi.org/10.1101/2022.01.11.475838> (2022).
76. McInnes, et al. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **3**, 861 (2018).
77. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**, 32–46 (2001).
78. Parks, D. H. et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **50**, D785–D794 (2021).
79. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**, e61217 (2013).
80. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
81. Galili, T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
82. Lin, H. & Peddada, S. D. Analysis of compositions of microbiomes with bias correction. *Nat. Commun.* **11**, 3514 (2020).
83. Youngblut, N. D. & Ley, R. E. Struo2: efficient metagenome profiling database construction for ever-expanding microbial genome datasets. *PeerJ* **9**, e12198 (2021).
84. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinforma.* **11**, 119 (2010).
85. Mistry, J. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **49**, D412–D419 (2021).
86. Feldgarden, M. et al. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Sci. Rep.* **11**, 12728 (2021).
87. Arango-Argoty, G. et al. DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**, 23 (2018).
88. Florensa, A. F., Kaas, R. S., Clausen, P., Aytan-Aktug, D. & Aarestrup, F. M. ResFinder—an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microb. Genom.* **8**, 748 (2022).
89. Burgin, J. et al. The European Nucleotide Archive in 2022. *Nucleic Acids Res.* **51**, D121–D125 (2023).

Acknowledgements

The compute infrastructure for this project was funded by the DFG [469073465], received by AK. This work was supported financially by Saarland University, the Uds-HIPS TANDEM initiative and the TALENTS

Marie Skłodowska-Curie COFUND-Action of the European Union [101081463], received by LAGM. The views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the European Union, which cannot be held responsible for them.

Author contributions

A.K., V.K., and R.F. had the idea for this study. G.P.S., J.R., S.L.B., and A.K. developed the methodology. R.F. collected samples and managed the animals. J.R. and M.J.S. performed whole-genome DNA extraction, culturing, and MALDI-TOF analysis. B.G.C. interpretation was performed by G.P.S., A.G., and R.M. S.L.B., V.K., R.M., M.K., and A.K. provided financial resources and supervised the project. J.R. supervised experimental investigations and project coordination. G.P.S., A.T., L.A.G.M., and A.K. performed computational analysis and data processing. G.P.S. and J.R. drafted the paper. All authors have read and critically reviewed the paper, in particular intellectual content, and agreed to the submission of the paper.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

G.P.S., R.M., and A.K. are co-founders of MooH GmbH, a company developing metagenomic-based oral health tests. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-52669-9>.

Correspondence and requests for materials should be addressed to Andreas Keller.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024