

RESEARCH

Open Access



# Contrasting recovery of metagenome-assembled genomes and derived bacterial communities and functional profiles from lizard fecal and cloacal samples

Mauricio Hernández<sup>1,2\*</sup> , Jorge Langa<sup>3,4</sup> , Ostaizka Aizpurua<sup>4</sup> , Yendi E. Navarro-Noya<sup>5</sup>  and Antton Alberdi<sup>4\*</sup> 

## Abstract

Genome-resolved metagenomics, based on shotgun sequencing, has become a powerful strategy for investigating animal-associated bacterial communities, due its heightened capability for delivering detailed taxonomic, phylogenetic, and functional insights compared to amplicon sequencing-based approaches. While genome-resolved metagenomics holds promise across various non-lethal sample types, their effectiveness in yielding high-quality metagenome-assembled genomes remains largely unexplored. Our investigation of fecal and cloacal microbiota of the mesquite lizards (*Sceloporus grammicus*) using genome-resolved metagenomics revealed that fecal samples contributed 97% of the 127 reconstructed bacterial genomes, whereas only 3% were recovered from cloacal swabs, which were largely enriched with host DNA. Taxonomic, phylogenetic and functional alpha bacterial diversity was greater in fecal samples than in cloacal swabs. We also observed significant differences in bacterial community composition between sampling methods, and higher inter-individual variation in cloacal swabs. *Bacteroides*, *Phocaeicola* and *Parabacteroides* (all Bacteroidota) were more abundant in the feces, whereas *Hafnia* and *Salmonella* (both Pseudomonadota) increased in the cloaca. Functional analyses showed that metabolic capacities of the microbiota to degrade polysaccharides, sugars and nitrogen compounds were enriched in fecal samples, likely reflecting the role of intestinal bacteria in nutrient metabolism. Overall, our results indicate that fecal samples outperform cloacal swabs in characterizing bacterial assemblages within lizards using genome-resolved metagenomics.

**Keywords** Genome-resolved metagenomics, Non-invasive methods, MAG, Microbiome, *Sceloporus*, Shotgun sequencing

\*Correspondence:

Mauricio Hernández  
delmer.hernandez@unah.edu.hn  
Antton Alberdi  
antton.alberdi@sund.ku.dk

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. 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-nc-nd/4.0/>.

## Introduction

Microbial symbiotic communities exert a fundamental role in the ecology and evolution of many animals [34]. Therefore, there is an increasing interest in investigating how gut microbiota impacts host health and fitness. Currently, the two main approaches for analyzing the microbial communities are 16S rRNA gene amplicon sequencing and shotgun metagenomic sequencing. While most studies rely on the former for taxonomic classification and quantification of bacterial and archaeal taxa [17], 16S rRNA sequencing does not provide direct information of microbial gene contents, which prevents direct inference of metabolic functions [29]. In consequence, genome-resolved metagenomics (GRM), derived from shotgun sequencing, is becoming increasingly popular for its enhanced capacity to yield direct information on bacterial and archaeal functional capabilities through the reconstruction of metagenome-assembled genomes (MAGs) [14, 41]. Recent studies have evidenced its suitability to assess the functional potential of microbial communities and taxonomic identification up to species- or strain-level [7, 25, 28].

Since collecting intestinal contents usually requires euthanizing animals, researchers often need to rely on fecal samples or rectal/cloacal swabs as proxies for characterizing intestinal microbiomes. Benchmarking based on 16S rRNA sequencing revealed the importance of sample type for recovering microbial communities, highlighting that fecal microbiota is similar to hindgut microbiota [26, 52], whereas cloacal microbiota may be a mixture of microbes coming from reproductive and digestive systems [18, 56]. However, increasing evidence indicates that fecal samples comprise a reliable proxy for assessing intestinal microbial communities [18, 49, 52]. While the impact of sample type for characterizing intestinal microbiomes using 16S rRNA sequencing has been thoroughly studied across multiple taxa, similar studies based on GRM are less common [48]. Unlike 16S rRNA sequencing, which only targets a single microbial gene, GRM relies on the metagenomic assembly of the total DNA, including non-microbial DNA derived from the host or ingested prey. The variability in sample types (e.g. ileum versus caecum) and sampling procedures (e.g. digesta collection versus mucosal scraping) can lead to variations in the amounts of host DNA present, while factors such as diet type and digestive efficiency can influence the levels of dietary DNA [1]. Consequently, the proportion of non-microbial DNA in the analyzed sample can significantly impact downstream analyses. It is therefore imperative to test the suitability of different non-lethal sampling methods for analyzing microbial communities through the reconstruction of MAGs.

Aiming to explore the magnitude of the technical biases introduced by the sample type in shotgun sequencing-based microbiome analysis, we used GRM to generate the first catalog of bacterial genomes associated with the mesquite lizard (*Sceloporus grammicus*), and analyze the diversity, composition, and functional traits of the microbiota recovered from fecal samples and cloacal swabs.

## Materials and methods

### Sample data collection

The study was approved by the “*Secretaría de Medio Ambiente y Recursos Naturales* (SEMARNAT)” in Mexico under the collecting permit SGPA/DGVS/007736/20, and samples were collected following the Official Mexican Standard NOM-126-ECOL-2000 as a guideline to handle the reptiles.

Fieldwork was conducted in the La Malinche National Park (central Mexico), a high-mountain ecosystem that rises to 4461 m above sea level (m a.s.l.). Only adult males were considered in this study to avoid sex-dependent variation in lizard microbial communities (snout–vent length > 44.1 mm; [20]). A total of 10 fecal samples and 10 cloacal samples were collected from ten individuals (two samples per individual) living at ~2600 m a.s.l. (19°13′39.5″N, 97°54′44.1″W). Once captured, lizards were transported in cloth bags to La Malinche Scientific Station at 3100 m a.s.l., housed separately in plastic boxes (20×30×15 cm), and the next day, each individual was exposed to sunlight to induce their natural defecation. Fecal samples were immediately collected upon defecation using sterile tweezers and transferred into sterile 1.5 mL tubes. Collection time did not vary among individuals. Thereafter, the exterior of the cloaca was cleaned with alcohol to prevent contamination from exogenous microbes. Cloacal samples were obtained using sterile rayon swabs with a diameter of 1 mm (COPAN, Italy), inserted in the cloacal opening, gently rotated, and transferred to sterile 1.5 mL tubes. Due to *S. grammicus* being a small lizard, the swab size was appropriately adjusted to its size, trying to retrieve mostly luminal microbiota rather than host DNA from cloacal epithelium. In addition, to ensure less contamination of cloacal swabs with feces, we tried to insert swabs ~10 mm into the cloaca and did not reach the rectal section, avoiding collecting any residual material after defecation. Fecal and cloacal samples were transported to the laboratory at 4 °C in a cooler and stored at –20 °C until DNA extraction.

### DNA extraction and shotgun metagenomic sequencing

To obtain high-quality DNA that met the required standards for shotgun sequencing, the following extraction process was used: fecal samples were washed twice with 1 mL of 0.15 M decahydrated tetrasodium

pyrophosphate, followed by two washes with 0.15 M phosphate buffer (pH 8). Cell lysis was achieved using thermo/mechanical disruption. DNA from cloacal swabs was also extracted using thermo/mechanical disruption for cell lysis, followed by precipitation with cold isopropyl alcohol and glycogen (see Supporting Information; [18]). The extracted DNA was sequenced at the Roy J. Carver Biotechnology Center, University of Illinois (Champaign, IL, USA). Shotgun genomic libraries were constructed from 300 ng of DNA after sonication with a Covaris ME220 (Covaris, MA) to an average fragment size of 400 bp with the Hyper Library construction kit from Kapa Biosystems (Roche, CA). Libraries were electrophoresed on a 2% agarose gel. The size-selected libraries were amplified with 3 cycles of PCR and run on a Fragment Analyzer (AATI, Ankeny, IA) to confirm the absence of free primers and adaptor dimers, as well as the presence of DNA of the expected size range. Libraries were pooled in equimolar concentration and quantitated by qPCR on a Bio-Rad CFX Connect Real-Time System (Bio-Rad Laboratories, Inc. CA). The pooled shotgun libraries were sequenced on an Illumina NovaSeq 6000 SP lane with 2×150 nt paired-end configuration. Lastly, fastq read files were generated and demultiplexed with the bcl2fastq v2.20 Conversion Software (Illumina, San Diego, CA).

### Bioinformatic analysis

Raw metagenomic reads were processed following the bioinformatic workflow developed for the 3D'omics European Union Horizon 2020 Project (<https://3domics.eu/>), available online as a Snakemake pipeline [27] at GitHub ([https://github.com/3d-omics/mg\\_assembly](https://github.com/3d-omics/mg_assembly)), and based on the Earth Hologenome Initiative (<https://www.earthhologenome.org/>; see [31]). Briefly, Fastp v0.23.4 was used to remove adapters, and low-quality and short reads [10], and prokaryotic sequencing fractions were estimated using SingleM microbial fraction (Eisenhofer et al., preprint). Next, host-derived reads were removed by mapping against a reference host genome using Bowtie2 v.2.5.1 [30] and Samtools v.1.18 [13]. We used the reference genome of the phylogenetically related species *S. undulatus* (NCBI accession number PRJNA656311; [53]), due to the absence of a reference genome for *S. grammicus*. The unmapped metagenomic reads were assembled (samples were treated individually) and co-assembled (samples were pooled and processed together) into contigs by MEGAHIT v.1.2.9 (Li et al., 2016), and reads were mapped to all the contigs of the corresponding sample using Bowtie2 with default settings. Contigs were then binned using CONCOCT v.1.1 [4], MetaBAT2 v.1.7 [22] and MaxBin2 v.2.2.4 [54]. Bins were polished with MAG-Scot v.1.0.0 [46] and their quality was assessed using

CheckM2 [11]. MAGs were dereplicated at 95% average nucleotide identity (ANI) using dRep v.3.4.3 [36] to obtain genomically determined bacterial species representations [12, 45]. CoverM v.0.6.1 was used to calculate the percentage of reads mapping to each MAG (<https://github.com/wwood/CoverM>).

Dereplicated MAGs were taxonomically annotated using GTDB-Tk v.2.3.2 against the Genome Taxonomy Database (release 214) [9, 38]. The phylogenetic tree of the MAGs was generated leveraging the phylogenetic placement of the previous step, by pruning the reference GTDB-tk genomes using the function *keep.tip* in *ape* v.5.7-1 [37]. Functional prediction of the MAGs was performed with DRAM v.1.4.6 [47] against the Pfam, KEGG, UniProt, CAZY, and MEROPS databases. We employed distillR (available at <https://github.com/anttonalberdi/distillR>) to distill functional annotations into Genome-Inferred Functional Traits (GIFTs), which serve as quantitative indicators of the capacity of each MAG to either degrade or produce key biomolecules. The reference database, comprising 328 metabolic pathways and modules from KEGG [21] and Metacyc [23] databases, facilitated the conversion of raw annotations into 190 GIFTs.

### Statistical analyses

#### Alpha diversities

All statistical analyses were performed using the R software v.4.3.2 (R Core Team, 2023). We determined neutral, phylogenetic and functional diversities of bacterial communities using Hill numbers [19]. Neutral Hill numbers represent species diversity without considering the degree of relatedness among MAGs, while phylogenetic Hill numbers incorporate branch-length information of the phylogenetic tree of the MAGs and functional Hill numbers consider functional differences among MAGs. Hill numbers differ by the parameter  $q$ , which determines the sensitivity of the measure to the relative abundance [8]. For capturing effects of different components (neutral, phylogenetic and functional) and orders of diversity ( $q=0$  only accounts for presence/absence, and  $q=1$  weighs MAGs according to their relative abundances), species richness of  $q=0$ , neutral diversity of  $q=1$ , phylogenetic diversity of  $q=1$  and functional diversity of  $q=1$  were computed using *Hilldiv2* v.2.0.2 [3]. Linear mixed models were used to quantify differences in bacterial alpha diversity between sampling methods, with lizard ID included as random effect, using the function *lmer* from the *lme4* package v.1.1–35.1 [5].

#### Beta diversities

The dissimilarities between fecal and cloacal MAGs composition were calculated in terms of Hill numbers by computing the Sørensen-type turnover from neutral,

phylogenetic and functional beta diversities at order  $q=1$ , using the *hillpair* function in *Hilldiv2*. Nonmetric multidimensional scaling (NMDS) ordination plots based on derived distance matrices were performed to visualise bacterial composition variation. The *betadisper* function in *vegan* v.2.6–4 [35] was applied to test differences in dispersion within sampling methods. PERMANOVA was performed to test for differences in bacterial composition between sampling methods using the *adonis2* function in *vegan*, with individual identity as a blocking factor to control for repeated sampling using the *strata* function. To visualize bacteria according to their functional features, the MAGs were ordinated based on their GIFTs through a t-SNE analysis using the *Rtsne* package v.0.17 [24].

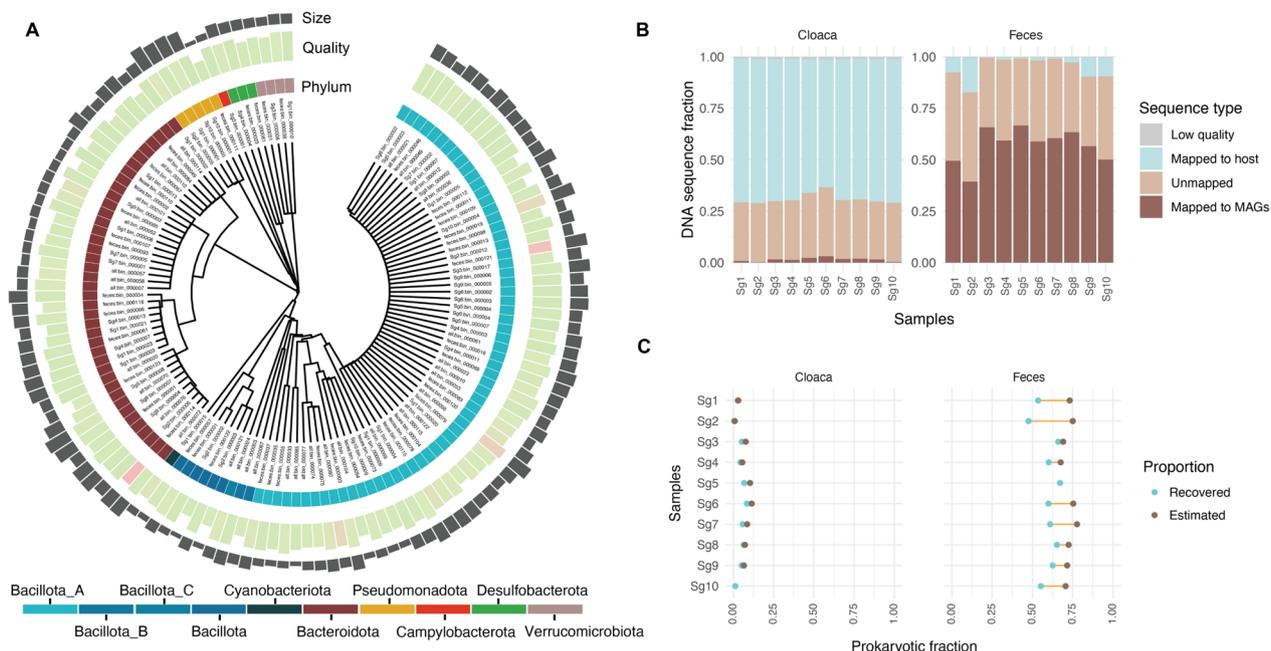
**Differential taxonomic and functional abundances**

Differential abundance analyses were performed to investigate which bacterial taxa significantly differ between sampling methods, using *ANCOM-BC2* package v.2.4.0 [33]. Considering the lizard ID random effect, differential abundance analyses with an alpha value of 0.05 and Bonferroni–Holm method for significance adjustment, were conducted at the MAG level, with log-fold changes between sample types and  $-\log(p\text{-values})$  visualized in a volcano plot, and at the phylum level, with results visualized in a bar chart. In addition, we

calculated community-weighted values of GIFTs before comparing values between sample types using linear mixed models including lizard ID as random effect: (e.g. `lmer(GIFT ~ samplotype + (1|individual), data=., REML=FALSE)`). The resulting  $p$ -values were adjusted using the Bonferroni method to account for multiple testing.

**Results**

We obtained a total of 387,468,658 raw sequencing reads from 10 fecal and 10 cloacal samples, with a mean depth of  $19,373,433 \pm 6,509,982$  reads per sample. Read mapping to the host genome revealed that cloacal swabs contained a significantly larger fraction of host DNA compared to fecal samples (cloaca:  $68.8 \pm 2.5\%$ ; feces:  $4.8 \pm 5.6\%$ ; LMM:  $p=0.001$ ). Metagenomic assemblies and binning yielded a total of 127 MAGs (Fig. 1A), with an average completeness of  $88.3 \pm 11.3\%$  and contamination of  $3.4 \pm 3.6\%$  (Supplementary Figure S1). Cloacal swabs contributed 3.0% of the MAGs, while fecal samples contributed 97.0% of the genomes included in the MAG catalog. The mapping rates of quality-filtered reads against the MAG catalog were  $1.5 \pm 0.9\%$  for cloacal and  $57.3 \pm 8.5\%$  for fecal samples (Fig. 1B). However, these values were close to the estimated microbial fractions calculated using SingleM (Fig. 1C), yielding domain-adjusted mapping rates



**Fig. 1** Microbial genome catalog and DNA fraction statistics. **A** Phylogenetic tree of the 127 metagenome-assembled genomes (MAGs) reconstructed, with their genome sizes, quality scores and taxonomic information. **B** DNA sequence fractions with low quality and assigned to the lizard host genome, MAG catalog and other unknown origins. **C** Sample-specific difference of the estimated and recovered prokaryotic fraction

(DAMR) of  $83.5 \pm 10.9\%$  and  $77.3 \pm 12.5\%$  in cloacal and fecal samples, respectively.

**Taxonomic composition of the microbial communities**

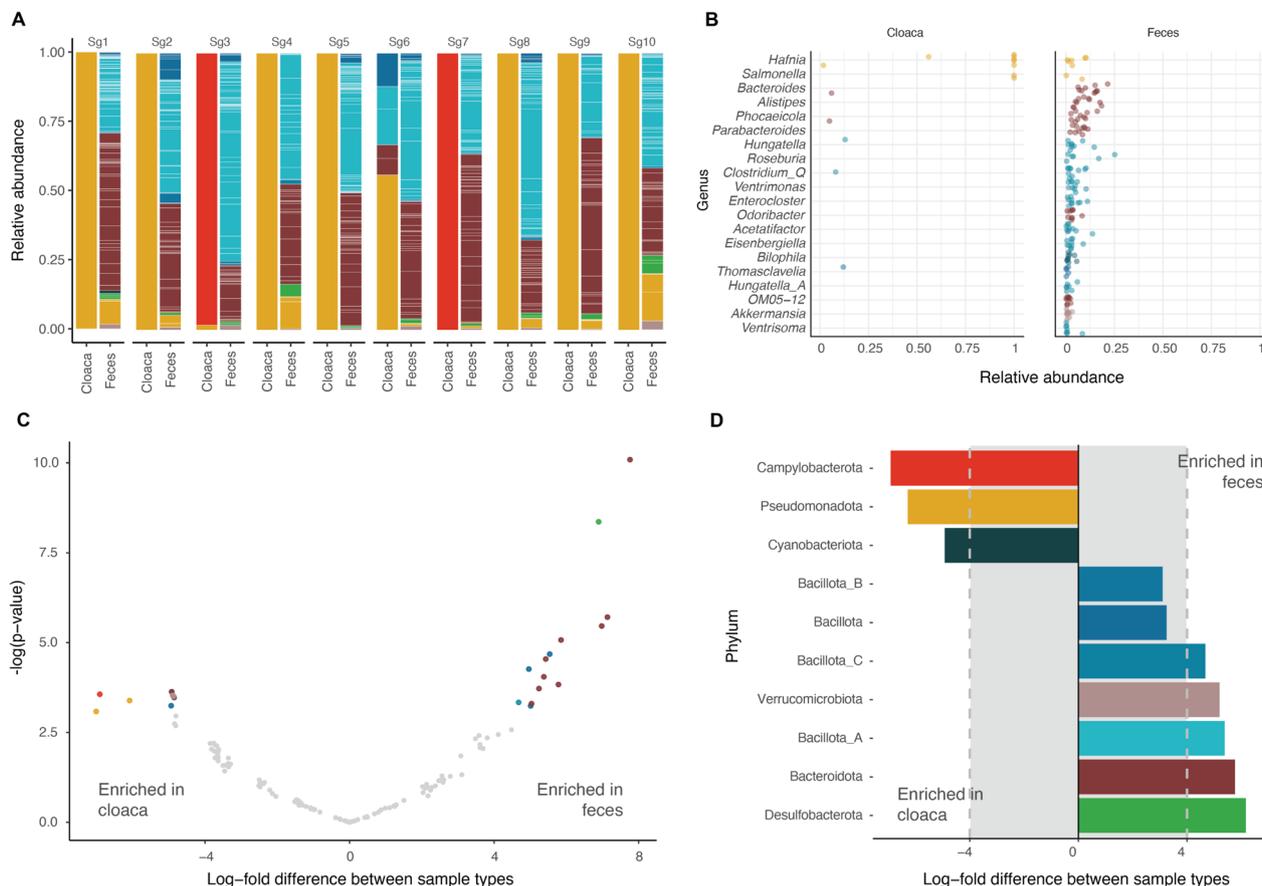
The MAGs were taxonomically assigned to 10 bacterial phyla, dominated by Pseudomonadota (40.43 ± 46.45%), Bacillota\_A (23.55 ± 24.46%), Bacteroidota (21.75 ± 23.60%), and Campylobacterota (9.94 ± 30.49%), with the remaining bacterial phyla representing less than 5% of reads mapped to the MAG catalog (Fig. 2A). At family level, Enterobacteriaceae (40.37 ± 46.50%), Lachnospiraceae (21.33 ± 22.54%), Bacteroidaceae (11.53 ± 12.33%), Helicobacteraceae (9.94 ± 30.49%), Rickenellaceae (4.86 ± 6.41%) and Tannerellaceae (3.78 ± 4.40%) were the most abundant in the analyzed samples (Supplementary Figure S2). Furthermore, most bacterial MAGs were assigned to the genera *Hafnia* (29.32 ± 43.63%), *Salmonella* (10.43 ± 30.69%), *Bacteroides* (6.79 ± 7.31%), *Alistipes* (4.70 ± 6.08%),

*Phocaeicola* (4.02 ± 4.74%) and *Parabacteroides* (3.50 ± 4.13%) (Fig. 2B).

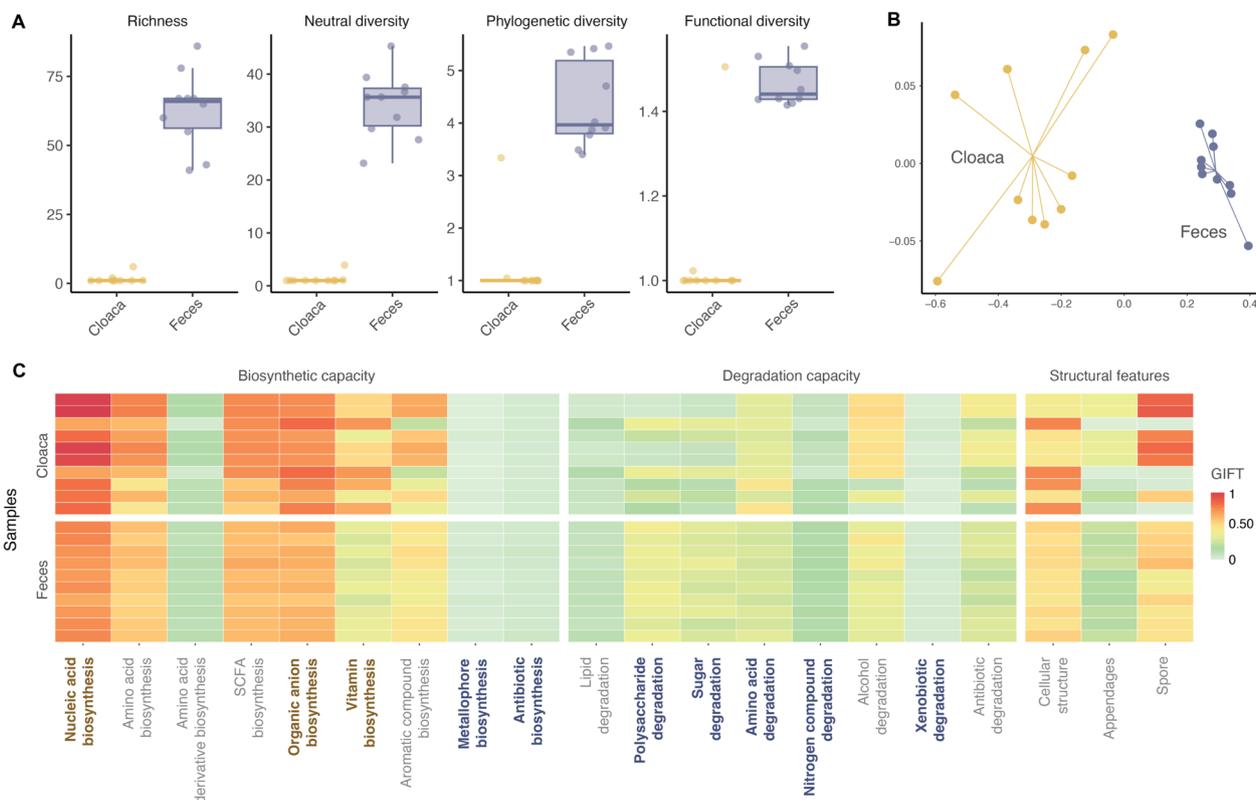
The representation of these taxa varied significantly between both sampling methods. We found that 9 MAGs were significantly more abundant in cloacal swabs, while 14 MAGs were more abundant in fecal samples (ANCOM-BC, Fig. 2C). Cloacal swabs exhibited an over-representation of Campylobacterota, Pseudomonadota and Cyanobacteriota, with a significantly higher proportion of genera *Hafnia* and *Salmonella* (Fig. 2D). In contrast, fecal samples yielded significantly elevated presence of Desulfobacterota, Bacteroidota, Bacillota\_A, Verrucomicrobiota, and Bacillota\_C, among which the genera *Bacteroides*, *Parabacteroides*, and *Phocaeicola* stood out prominently from cloacal swabs.

**Diversity of the microbial communities**

All four alpha diversity metrics analyzed showed that fecal samples were significantly more diverse than cloacal swabs (LMM:  $p=0.001$ ) (Fig. 3A). Beta diversities based



**Fig. 2** Taxonomic overview and differences between sample types. **A** Stacked barplot of the relative abundances of MAGs in each sample, coloured by phylum. **B** Relative abundances of the 20 most common genera split by sample type. **C** Differential abundance of MAGs between sample types, in which MAGs with significant log-fold differences between sample types are coloured according to their phyla. **D** Differential abundance of phyla, with vertical dashed lines indicating significance thresholds



**Fig. 3** Diversity and functional differences between sample types. **A** Alpha diversity differences between sample types under different Hill number metrics. Richness ( $q=0$ ) only considers the number of MAGs detected in each sample. Neutral diversity ( $q=1$  or exponential of Shannon index) accounts for relative abundances of the MAGs. Phylogenetic diversity ( $q=1$ ) accounts for both relative abundances and phylogenetic relationships among MAGs. Functional diversity ( $q=1$ ) accounts for both relative abundances and functional relationships among MAGs. **B** NMDS ordination plot derived from pairwise dissimilarity values between samples based on functional Hill numbers. **C** Weighed capacities of genome-inferred functional traits (GIFT) of bacterial communities to conduct diverse metabolic processes. Bolded trait names in the x-axis indicate statistical significance between cloacal and fecal samples, with yellow-colored functions enriched in cloaca and blue-colored functions enriched in feces

on different diversity metrics also varied considerably, clearly separating both types of samples (PERMANOVAs:  $p < 0.001$ ) and showing a greater variation in cloacal bacterial composition compared to fecal bacterial composition (Beta Dispersion  $p < 0.05$ ; Fig. 3B, Supplementary Figure S3). The in-depth analysis of genome-inferred functional traits derived from MAG annotations also indicated that the functional microbiome profiles recovered from both sampling methods differed, with 10 out of the 20 analyzed functional traits exhibiting significant differences between cloacal and fecal samples (Fig. 3C).

### Discussion

An increasing number of researchers are opting for GRM to characterize gut microbiomes [7, 25, 28, 50, 51], due to the improved resolution this approach offers in comparison to amplicon sequencing. GRM involves analyzing total DNA, encompassing host DNA, dietary DNA, viral DNA, and more, rather than just amplifying a bacterial genetic marker. Consequently, not all conclusions based

on 16S rRNA sequencing can be directly extrapolated to GRM [1]. Our comparison of sample types revealed drastic differences in the recovery of gut microbial communities associated with mesquite lizards. The diversity recovered in the fecal samples was several orders of magnitude larger than that in the cloacal swabs, mirroring patterns recently documented in a study encompassing more than 150 vertebrate species [39]. Bacterial assemblages reconstructed from fecal samples resembled typical gut communities dominated by anaerobes from the phyla Bacillota\_A and Bacteroidota, and a lower representation of Pseudomonadota and Desulfobacterota. In contrast, 90% of the cloacal swabs were represented by a single bacterium belonging to Pseudomonadota or Campylobacterota. While some recent amplicon studies based on cloacal swabs in birds and amphibians ([6, 56], as well as rectal swabs in humans [42], have reported otherwise, our findings were largely aligned with a previous 16S rRNA sequencing-based analysis in this same species [18]. This study, indicated that fecal samples

mirrored bacterial communities found in the mid- and hind-gut, while cloacal swabs exhibited a substantial depletion in bacterial diversity [18]. Similarly, in juvenile ostriches, fecal samples had significantly greater bacterial alpha diversity compared to the cloacal swabs [52]. Some of the variation in diversity and composition may be related to differences in the DNA extraction techniques used for the two sample types. These variations were necessary to obtain DNA of sufficient quality and quantity for library construction and shotgun sequencing. However, in a previous study using 16S rRNA gene sequencing, we found a strong correlation in taxonomic composition between the same two sample types (*Spearman correlation* = 0.86; [18]), suggesting that differences in DNA extraction methods were not the main drivers of the observed results. In fact, Pietroni et al. [39] employed identical procedures for DNA extraction from both fecal and anal/cloacal swabs, and arrived at the same conclusions as ours.

Our shotgun-based analysis shed light on several potential factors contributing to the observed variation. Firstly, we found that nearly 70% of the DNA sequences retrieved from cloacal swabs mapped to the host genome. However, this figure likely underestimates the true host fraction due to the absence of a reference genome of *S. grammicus*. Consequently, we employed the chromosome-level genome assembly of the related species *S. undulatus*, which likely resulted in reduced read mapping success. In fact, our estimations based on marker gene analysis revealed that only 5–10% of the sequences from cloacal swabs belonged to bacteria and/or archaea. With such a low fraction of prokaryotic DNA, the ability to reconstruct bacterial genomes from metagenomic mixtures is severely limited, as evidenced by the scant 3% representation of bacterial genomes derived from cloacal swabs in our MAG catalog. The low fraction of prokaryotic DNA is likely the result of a low density of microbial cells in the cloaca, combined by shedding of host epithelial cells when swabbing. In addition, cloacal microbiota of reptiles may be exposed to high selective pressure as a consequence of mucus secretion with antimicrobial properties [40], which likely contributes to the reduction in microbial diversity. The non-overlapping communities between fecal and cloacal samples may be primarily associated with a high and transient microbiota of fecal matter compared to less diverse and more stable microbiota of cloacal region.

Beyond the negligible contribution of cloacal swabs to the MAG catalog reconstruction, the rest of bacteria reconstructed from fecal samples were also largely undetected in cloacal swabs. This discrepancy further supports that the limited diversity detected in

cloacal swabs is not solely attributable to technical constraints derived from low microbial DNA fractions, but is indicative of a significantly depleted bacterial community. All but one cloacal samples were overwhelmingly dominated by a single member of the phyla Campylobacterota (unknown Helicobacteraceae) or Pseudomonadota (*Salmonella* and *Hafnia*). These taxa are known for their ability to thrive under aerobic or microaerobic conditions [2], which are more prevalent in the cloaca than the upstream sections of the intestinal tract. Furthermore, only one cloacal sample exhibited a higher diversity, encompassing multiple bacterial species also found in fecal samples. This sample likely represents remnants of fecal material from a recent defecation, further supporting the notion that biological factors rather than technical constraints drive the observed disparity in diversity between fecal and cloacal samples.

In addition to the compositional and diversity comparisons, which corroborated findings from previous 16S rRNA sequencing studies, our analyses offered novel direct insights into the functional capabilities of the bacteria associated with lizards. Community-weighted metabolic capacities of cloacal swabs were on average higher than those observed in feces, a pattern driven by the high abundances of Pseudomonadota with large genome sizes and very high metabolic capacities. In line with this observation, average capacities for synthesizing organic anions and vitamins were significantly higher in the cloaca. However, the community-level capacities of a number of metabolic functions were enriched in fecal samples. First, we found a higher polysaccharide and sugar degradation capacity in the fecal community, which aligns with the enrichment of anaerobic fermenting bacteria belonging to the Bacillota and Bacteroidota phyla. Nevertheless, this observation contrasts with a recent examination based on microbial functions predicted from 16S data [50, 51], which reported that carbohydrate metabolism was mainly associated with cloacal microbiota compared to intestinal microbiota in the red-necked keelback snake (*Rhabdophis subminiatus*). We also observed an increased capacity to degrade nitrogen compounds, suggesting the significance of intestinal microorganisms in metabolizing nitrogen waste produced by the host [44]. Finally, the increased antibiotics production capacity observed in feces is likely the result of an increased competitive pressure for resources in the intestine [43]. Overall, the functional properties of the microbiome retrieved from fecal samples exhibited a closer resemblance to those expected for a typical gut bacterial community compared to those retrieved from cloacal swabs.

## Conclusion

Our comparative analysis of sample types revealed significant disparities in the bacterial communities reconstructed from fecal samples versus cloacal swabs. The markedly low bacterial fraction observed in cloacal swabs, coupled with the distinct taxonomic and functional profiles obtained from each sample type, underscores the superiority of fecal samples as proxies for characterizing intestinal microbial communities of lizards using GRM. Beyond corroborating previous findings derived from 16S rRNA sequencing, our study provided novel insights into the underlying causes of the observed differences in composition and diversity between these contrasting sampling methods. Furthermore, given that *S. grammicus* shows a broad geographical distribution and several behavioral and physiological traits, we expect that the same pattern will occur in other spiny lizards, although further studies are required to assess how fecal and cloacal microbiota vary among different *Sceloporus* species. As the utilization of shotgun-sequencing methodologies to analyze gut microbiomes continues to grow [31], we advocate for researchers to conduct similar methodological analyses across diverse animal taxa. Only then will we be able to generate reliable data that enables us to understand the bidirectional interactions between animal and microbial ecological and evolutionary processes.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42523-025-00381-4>.

Additional file 1.  
Additional file 2.

## Acknowledgements

The authors thank La Malinche Scientific Station and Tlaxcala Center for Biology of Behavior (Autonomous University of Tlaxcala, Mexico) for access and logistical support.

## Author contributions

MH: sample collection and molecular laboratory work. MH and AA: conceptualization, formal analysis, investigation, methodology and writing-original draft. JL: bioinformatic analysis and writing-review & editing. OA: formal analysis, writing-review & editing. YEN-N: funding acquisition and writing-review & editing.

## Funding

Open access funding provided by Copenhagen University. This research was partially funded by the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT, Mexico). JL was funded by the Postdoctoral Program for the Improvement of Doctoral Research Staff of the Basque Government (grant number POS\_2022\_1\_0011). AA acknowledges the Danish National Research Foundation (grant DNRF143 A Center for Evolutionary Hologenomics).

## Availability of data and materials

Raw metagenomic reads were deposited into the NCBI Sequence Read Archive database under accession number PRJNA1106546. The R software

code used for this publication can be found on GitHub: [https://alberdilab.github.io/lizard\\_sample\\_types](https://alberdilab.github.io/lizard_sample_types).

## Declarations

### Ethics approval and consent to participate

The sampling and handling of mesquite lizards complied with ethical and legal regulations in Mexico to conduct research on wild populations (Convention on Biological Diversity), as stipulated in the *Norma Oficial Mexicana* (NOM-126-ECOL-2000). Permission for the sampling and handling of lizards was granted by the *Secretaría de Medio Ambiente y Recursos Naturales* (SEMARNAT, Mexico) under the collecting permit SGPA/DGVS/007736/20.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Departamento de Biología Celular y Genética, Escuela de Biología, Facultad de Ciencias, Universidad Nacional Autónoma de Honduras, Tegucigalpa 11101, Honduras. <sup>2</sup>Instituto de Investigaciones en Microbiología, Facultad de Ciencias, Universidad Nacional Autónoma de Honduras, Tegucigalpa 11101, Honduras. <sup>3</sup>Department of Genetics, Animal Physiology and Physical Anthropology, University of the Basque Country UPV/EHU, Barrio Sarriena S/N, 48940 Leioa, Bizkaia, Spain. <sup>4</sup>Center for Evolutionary Hologenomics, Globe Institute, University of Copenhagen, 1353 Copenhagen, Denmark. <sup>5</sup>Laboratorio de Interacciones Bióticas, Centro de Investigación en Ciencias Biológicas, Universidad Autónoma de Tlaxcala, Carretera San Martín Texmelucan Km 10.5, San Felipe Ixtacuixtla, 90120 Tlaxcala, Mexico.

Received: 18 October 2024 Accepted: 7 February 2025  
Published online: 15 February 2025

## References

- Aizpurua O, Dunn RR, Hansen LH, Gilbert MTP, Alberdi A. Field and laboratory guidelines for reliable bioinformatic and statistical analysis of bacterial shotgun metagenomic data. *Crit Rev Biotechnol*. 2023;44(6):1164–82. <https://doi.org/10.1080/07388551.2023.2254933>.
- Arai H. Regulation and function of versatile aerobic and anaerobic respiratory metabolism in *Pseudomonas aeruginosa*. *Front Microbiol*. 2011;2:103. <https://doi.org/10.3389/fmicb.2011.00103>.
- Alberdi A, Gilbert MTP. Hilldiv: an R package for the integral analysis of diversity based on Hill numbers. *BioRxiv*. 2019. <https://doi.org/10.1101/545665>.
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. Binning metagenomic contigs by coverage and composition. *Nat Methods*. 2014;11(11):1144–6. <https://doi.org/10.1038/nmeth.3103>.
- Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Soft*. 2015;67(1):1–48. <https://doi.org/10.18637/jss.v067.i01>.
- Berlow M, Kohl KD, Derryberry E. Evaluation of non-lethal gut microbiome sampling methods in a passerine bird. *Int J Avian Sci*. 2020;162(3):911–23. <https://doi.org/10.1111/ibi.12807>.
- Cao J, Hu Y, Liu F, Wang Y, Bi Y, Lv N, Li J, Zhu B, Gao GF. Metagenomic analysis reveals the microbiome and resistome in migratory birds. *Microbiome*. 2020;8:26. <https://doi.org/10.1186/s40168-019-0781-8>.
- Chao A, Chiu C-H, Jost L. Unifying species diversity, phylogenetic diversity, functional diversity, and related similarity and differentiation measures through Hill numbers. *Annu Rev Ecol Evol Syst*. 2014;45:297–324. <https://doi.org/10.1146/annurev-ecolsys-120213-091540>.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-TK: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics*. 2019;36(6):1925–7. <https://doi.org/10.1093/bioinformatics/btz848>.
- Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.

11. Chklovski A, Parks DH, Woodcroft BJ, Tyson GW. CheckM2: a rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat Methods*. 2023;20(8):1203–12. <https://doi.org/10.1038/s41592-023-01940-w>.
12. Conrad R, Brink CE, Viver T, Rodriguez-R LM, Aldeguer-Riquelme B, Hatt J, Venter S, Rossello-Mora R, Amann RL, Konstantinidis KT. Microbial species and intraspecies units exist and are maintained by ecological cohesiveness coupled to high homologous recombination. *Nat Commun*. 2024;15:9906. <https://doi.org/10.1038/s41467-024-53787-0>.
13. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):giab008. <https://doi.org/10.1093/gigascience/giab008>.
14. Durazzi F, Sala C, Castellani G, Manfreda G, Remondini D, De Cesare A. Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep*. 2021;11:3030. <https://doi.org/10.1038/s41598-021-82726-y>.
15. Eisenhofer R, Odriozola I, Alberdi A. Impact of microbial genome completeness on metagenomic functional inference. *ISME Commun*. 2023;3:12. <https://doi.org/10.1038/s43705-023-00221-z>.
16. Eisenhofer R, Alberdi A, Woodcroft B (2024) Large-scale estimation of bacterial and archaeal DNA fractions in metagenomes reveals biome-specific patterns. *bioRxiv*. Preprint.
17. Gotschlich EC, Colbert RA, Gill T. Methods in microbiome research: past, present and future. *Best Pract Res Clin Rheumatol*. 2019;33(6):101498. <https://doi.org/10.1016/j.berh.2020.101498>.
18. Hernández M, Ancona S, Hereira-Pacheco S, Díaz de la Vega-Pérez AH, Navarro-Noya YE. Comparative analysis of two nonlethal methods for the study of the gut bacterial communities in wild lizards. *Integr Zool*. 2023;18(6):1056–71. <https://doi.org/10.1111/1749-4877.12711>.
19. Hill MO. Diversity and evenness: a unifying notation and its consequences. *Ecology*. 1973;54(2):427–32. <https://doi.org/10.2307/1934352>.
20. Jiménez-Cruz E, Ramírez A, Marshall J, Lizana M, Montes de Oca A. Reproductive cycle of *Sceloporus grammicus* (Squamata: Phrynosomatidae) from Teotihuacan, state of Mexico. *Southwest Natural*. 2005;50:178–87.
21. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353–61. <https://doi.org/10.1093/nar/gkw1092>.
22. Kang DD, Froula J, Egan R, Wang Z. Metabat, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*. 2015;3:e1165. <https://doi.org/10.7717/peerj.1165>.
23. Karp PD, Riley M, Paley SM, Pellegrini-Toole A. The MetaCyc database. *Nucleic Acids Res*. 2002;30(1):59–61. <https://doi.org/10.1093/nar/30.1.59>.
24. Krijthe J, van der Maaten L, Krijthe MJ (2017) Rtsne: T-Distributed stochastic neighbor embedding using Barnes-hut implementation. R package version 0.17. <https://github.com/krijthe/Rtsne>
25. Kayani MUR, Ali Zaidi SS, Feng R, Yu K, Qiu Y, Yu X, Chen L, Huang L. Genome-resolved characterization of structure and potential functions of the zebrafish stool microbiome. *Front Cell Infect Microbiol*. 2022;12:910766. <https://doi.org/10.3389/fcimb.2022.910766>.
26. Kohl KD, Brun A, Magallanes M, Brinkerhoff J, Laspiur A, Acosta JC, Caviedes-Vidal E, Bordenstein SR. Gut microbial ecology of lizards: insights into diversity in the wild, effects of captivity, variation across gut regions and transmission. *Mol Ecol*. 2017;26(4):1175–89. <https://doi.org/10.1111/mec.13921>.
27. Köster J, Rahmann S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics*. 2012;28(19):2520–2. <https://doi.org/10.1093/bioinformatics/bts480>.
28. Koziol A, Odriozola I, Leonard A, Eisenhofer R, San José C, Aizpurua O, Alberdi A. Mammals show distinct functional gut microbiome dynamics to identical series of environmental stressors. *MBio*. 2023;14(5):e0160623. <https://doi.org/10.1128/mbio.01606-23>.
29. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol*. 2013;31(9):814–21. <https://doi.org/10.1038/nbt.2676>.
30. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
31. Leonard A, Abalos A, Adhola T, Aguirre W, Aizpurua O, Ali S, Andreone F, Aubret F, Ávila-Palma HD, Bautista Alcantara LF, Beltrán JF, Berg R, Berg TB, Bertolino S, Blumstein DT, Boldgiv B, Borowski Z, Boublil J, Büchner S. A global initiative for ecological and evolutionary hologenomics. *Trends Ecol Evol*. 2024;39(7):616–29. <https://doi.org/10.1016/j.tree.2024.03.005>.
32. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31(10):1674–6. <https://doi.org/10.1093/bioinformatics/btv033>.
33. Lin H, Das Peddada S. Analysis of compositions of microbiomes with bias correction. *Nat Commun*. 2020;11:3514. <https://doi.org/10.1038/s41467-020-17041-7>.
34. McFall-Ngai M, Hadfield MG, Bosch TCG, Carey HV, Domazet-Lošo T, Douglas AE, Dubilier N, Eberl G, Fukami T, Gilbert SF, Hentschel U, King N, Kjelleberg S, Knoll AH, Kremer N, Mazmanian SK, Metcalf JL, Nealon K, Pierce NE, Wernegreen JJ. Animals in a bacterial world, a new imperative for the life sciences. *Proceed Nation Acad Sci*. 2013;110(9):3229–36. <https://doi.org/10.1073/pnas.1218525110>.
35. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2023) Vegan: community ecology package. R package version 2.6–4. <https://CRAN.R-project.org/package=vegan>
36. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J*. 2017;11(12):2864–8. <https://doi.org/10.1038/ismej.2017.126>.
37. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. 2004;20(2):289–90. <https://doi.org/10.1093/bioinformatics/btg412>.
38. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res*. 2022;50(D1):D785–94. <https://doi.org/10.1093/nar/gkab776>.
39. Pietroni C, Gaun N, Leonard A, Lauritsen J, Martin-Bideguren G, Odriozola I, Aizpurua O, Alberdi A, Eisenhofer R. Hologenomic data generation and analysis in wild vertebrates. *Methods Ecol Evol*. 2024;16(11):97–107. <https://doi.org/10.1111/2041-210X.14456>.
40. Praja RN, Yudhana A, Haditanojo W, Oktavia V. Antimicrobial properties in cloacal fluid of olive ridley sea turtle (*Lepidochelys olivacea*). *Biodiversitas*. 2021;22(9):3671–6. <https://doi.org/10.13057/biodiv/d220909>.
41. Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol*. 2017;35(9):833–44. <https://doi.org/10.1038/nbt.3935>.
42. Radhakrishnan ST, Gallagher KI, Mullish BH, Serrano-Contreras JI, Alexander JL, Blanco JM, Danckert NP, Valdivia-García M, Hopkins BJ, Ghai A, Ayub A, Li JV, Marchesi JR, Williams HRT. Rectal swabs as a viable alternative to faecal sampling for the analysis of gut microbiota functionality and composition. *Sci Rep*. 2023;13(1):493. <https://doi.org/10.1038/s41598-022-27131-9>.
43. Raffaelli M. Learning from bacterial competition in the host to develop antimicrobials. *Nat Med*. 2018;24(8):1097–103. <https://doi.org/10.1038/s41591-018-0145-0>.
44. Ren X, Cao S, Akami M, Mansour A, Yang Y, Jian N, Wang H, Zhang G, Qi X, Xu P, Guo T, Niu C. Gut symbiotic bacteria are involved in nitrogen recycling in the tephritid fruit fly *Bactrocera dorsalis*. *BMC Biol*. 2022;20:201. <https://doi.org/10.1186/s12915-022-01399-9>.
45. Rodriguez-R LM, Conrad RE, Viver T, Feistel DJ, Lindner BG, Venter SN, Orelana LH, Amann R, Rossello-Mora R, Konstantinidis KT. An ANI gap within bacterial species that advances the definitions of intra-species units. *MBio*. 2024;15(1):e0269623. <https://doi.org/10.1128/mbio.02696-23>.
46. Rühlmann MC, Wacker EM, Ellinghaus D, Franke A. MAGScoT: a fast, lightweight and accurate bin-refinement tool. *Bioinformatics*. 2022;38(24):5430–3. <https://doi.org/10.1093/bioinformatics/btac694>.
47. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrowe AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan NB, Roux S, Wrighton KC. DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res*. 2020;48(16):8883–900. <https://doi.org/10.1093/nar/gkaa621>.
48. Shen T-CD, Daniel SG, Patel S, Kaplan E, Phung L, Lemelle-Thomas K, Chau L, Herman L, Trisolini C, Stonelake A, Toal E, Khungar V, Bittinger K, Reddy R, Wu GD. The mucosally-adherent rectal microbiota contains features

- unique to alcohol-related cirrhosis. *Gut Microbes*. 2021;13(1):1987781. <https://doi.org/10.1080/19490976.2021.1987781>.
49. Suzuki TA, Nachman MW. Spatial heterogeneity of gut microbial composition along the gastrointestinal tract in natural populations of house mice. *PLoS ONE*. 2016;11(9):e0163720. <https://doi.org/10.1371/journal.pone.0163720>.
  50. Tang K-Y, Wang Z-W, Wan Q-H, Fang S-G. Metagenomics reveals seasonal functional adaptation of the gut microbiome to host feeding and fasting in the Chinese alligator. *Front Microbiol*. 2019;10:2409. <https://doi.org/10.3389/fmicb.2019.02409>.
  51. Tang W, Zhu G, Shi Q, Yang S, Ma T, Mishra SK, Wen A, Xu H, Wang Q, Jiang Y, Wu J, Xie M, Yao Y, Li D. Characterizing the microbiota in gastrointestinal tract segments of *Rhabdophis subminiatus*: Dynamic changes and functional predictions. *MicrobiologyOpen*. 2019;8(7):e789. <https://doi.org/10.1002/mbo3.789>.
  52. Videvall E, Strandh M, Engelbrecht A, Cloete S, Cornwallis CK. Measuring the gut microbiome in birds: comparison of faecal and cloacal sampling. *Mol Ecol Resour*. 2018;18(3):424–34. <https://doi.org/10.1111/1755-0998.12744>.
  53. Westfall AK, Telemeco RS, Grizante MB, Waits DS, Clark AD, Simpson DY, Klabacka RL, Sullivan AP, Perry GH, Sears MW, Cox CL, Cox RM, Gifford ME, John-Alder HB, Langkilde T, Angilletta MJ, Leaché AD, Tollis M, Kusumi K, Schwartz TS. A chromosome-level genome assembly for the eastern fence lizard (*Sceloporus undulatus*), a reptile model for physiological and evolutionary ecology. *GigaScience*. 2021;10(10):giab066. <https://doi.org/10.1093/gigascience/giab066>.
  54. Wu Y-W, Simmons BA, Singer SW. Maxbin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2016;32(4):605–7. <https://doi.org/10.1093/bioinformatics/btv638>.
  55. Yu G. Using ggtree to visualize data on tree-like structures. *Curr Protoc Bioinform*. 2020;69(1):e96. <https://doi.org/10.1002/cpbi.96>.
  56. Zhou J, Nelson TM, Rodriguez Lopez C, Sarma RR, Zhou SJ, Rollins LA. A comparison of nonlethal sampling methods for amphibian gut microbiome analyses. *Mol Ecol Resour*. 2020;20(4):844–55. <https://doi.org/10.1111/1755-0998.13139>.

### Publisher's Note

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