Microbiota and genetic variation in reindeer calves: Analysis of data in relation to survival

Mearg Belay Brhane

Independent project • 30 credits
Swedish University of Agricultural Sciences, SLU
Department of Animal Science
Programme: Bioinformatics
Uppsala, 2026

Microbiota and genetic variation in reindeer calves: analysis of data in relation to survival

Mikrobiota och genetisk variation hos renkalvar: analys av data i relation till överlevnad

Mearg Belay Brhane

Supervisor:	**Peter Halvarsson, SLU, Department of HBIO**
Assistant supervisor:	**Anna Skarin**, SLU, Department of Applied Animal Science and Welfare
Assistant supervisor:	**Maria Cortaza**, UU, Dept. of Ecology and Genetics
Examiner:	**Erik Bongcam Rudloff,** SLU, Department of Bioinformatics

Credits:	30 credits
Level:	Second cycle, A2E)
Course title:	**Independent project**
Course code:	EX1002
Programme/education:	Bioinformatics
Course coordinating dept:	Animal Science
Place of publication:	Uppsala
Year of publication:	2026
Cover picture:	Peter Ewins

| **Keywords:** | *Rangifer tarandus*, microbiome, taxa, survival, *Alysiella* |

Swedish University of Agricultural Sciences
Department of Animal Science

ABSTRACT

Microbiota tightly linked to their hosts significantly affect vertebrates' early growth, health, and survival. However, their effects on the fitness of wild and semi-domesticated species remain poorly understood. In Reindeers husbandry, calf mortality is one of the most significant challenges. Previous studies have shown that mortality is influenced by many interconnected factors, such as environmental conditions, maternal condition, sickness, and the quality of the pasture. Both the field of ecology and the management of reindeer herds require an understanding of the biological elements that contribute to the individuals' ability to survive. Using 16S rRNA gene metabarcoding, this study examined the oral and rectal microbiota of reindeer calves and how these microbial communities were influenced by factors such as sampling location, calf sex, and survival in the early stages of life.

Sequencing was performed using a paired-end Illumina approach, and DADA2 was used to identify amplicon sequence variants (ASVs). I used R packages such as phyloseq and vegan to perform analyses that let us look closely at the data. I used Observed ASVs and the Shannon diversity index to assess microbial diversity, and I used Bray–Curtis ’ dissimilarity, ordination plots, PERMANOVA, and GLM to assess community changes. I used logistic regression models based on ordination axes to assess whether microbial community composition may predict calf survival. I used several differential abundance (ALDEx2, MaAsLin2, and ANCOM-BC) to identify genera linked to early-life survival.

Alpha diversity did not differ among body-site samples. This indicates that diversity alone may not guarantee survival. Beta diversity differed between samples collected from the mouth and the rectum. PERMANOVA shows that the location of a sample has a big effect on the diversity of microbial communities. GLM models explored that microbes may predict survival. There were only weak connections, especially in the oral microbiome. Alysiella was one of the main mouth taxa discovered in survivors' microbiome analyses. The genera identified by differential abundance methods were Roseisolibacter, Capnocytophaga, Monoglobus, and Parafilimonas.

The findings indicate that the composition of the oral microbiome is marginally associated with the survival of reindeer calves during their early life stages.

Key words: *Rangifer tarandus*, microbiome, taxa, survival, *Alysiella*

Table of Contents

List of tables

List of figures

Abbreviations

Abbreviations	Description
ALDEx2	Analysis of Differential Abundance taking compositional data into account (R package)
ANCOM-BC	Analysis of Composition of Microbiomes with Bias Correction (R package)
ASV	Amplicon Sequence Variant
ASV table	Amplicon Sequence Variant abundance table
BSA	Bovine Serum Albumin
bp	Base Pair
CLR	Centered Log Ratio
ddH2O	Double Distilled Water
dNTPs	Deoxynucleotide Triphosphates
GLM	Generalized Linear Model
MaAsLin2	Multivariate Association with Linear Models 2 (R package)
NGI	National Genomics Infrastructure
NMDS	Non-metric Multidimensional Scaling
OTU	Operational Taxonomic Unit
PCoA	Principal Coordinates Analysis
PERMANOVA	Permutational Multivariate Analysis of Variance
R	R statistical environment
rRNA	Ribosomal Ribonucleic Acid
RPM	Revolutions Per Minute
SLU	Sveriges lantbruksuniversitet

1. Introduction

The microbiota linked to vertebrate hosts is essential for host health, development, and survival. Microbial communities inhabit several bodily sites early in life, facilitating immune system development, nutrient absorption, and disease resistance (Sommer & Bäckhed, 2013). There is growing evidence that the microbial composition in early life can have long-term effects on host fitness, influencing growth, disease susceptibility, and survival (Du et al., 2023; Worsley et al., 2021).

Although host–microbiota interactions have been well investigated in humans and laboratory animals, little is known about similar relationships in wild and semi-domesticated species. In wildlife systems, host-associated microbiota may be influenced by environmental exposure, diet, and seasonal fluctuations, thereby affecting host adaptation and survival (Kamenova et al., 2023; Leclaire et al., 2023).

Reindeer (*Rangifer tarandus*) are crucial herbivores in Arctic and sub-Arctic ecosystems and are very important to the environment, culture, and economy in northern areas (Kamenova et al., 2023). One of the biggest problems in reindeer farming is calf mortality. Previous research indicates that several interrelated variables, including environmental conditions, maternal condition, illness, and pasture quality, affect mortality, although predation can be a significant source of calf loss in some places(Åhman et al., 2022). Nonetheless, the potential role of host-associated microbiota in the early-life survival of reindeer remains inadequately defined.

Research has demonstrated that the gut microbiota affects energy metabolism, immunological responses, and disease resistance in ruminants and other mammals (Ding et al., 2025; Tardiolo et al., 2025). Along with gut-associated microbial communities, the oral microbiota is an important microbial habitat that has received less attention. Oral bacteria may affect host health directly via immunological interactions or indirectly by colonizing the gastrointestinal system (Baker et al., 2024; Dewhirst et al., 2010; Ptasiewicz et al., 2022).

1.1 Problem statement

Recognition of the essential role of host-associated microbiota is increasing; however, limited knowledge exists regarding the impact of early-life microbial environments on reindeer survival. Previous research indicates that several interrelated variables, including environmental conditions, maternal condition, illness, and pasture quality, affect mortality, although predation can be a significant source of calf loss in some places (Åhman et al., 2022). Nevertheless, the influence of microorganisms on early-life survival remains insufficiently investigated (Gaillard et al., 2000; Kamenova et al., 2023).

There are modest biological changes that happen during early development that can have a huge effect on how fit someone will be in the end (Gaillard et al., 2000). How ungulate populations evolve over time is mostly due to how well they survive in their early years. Evidence suggests that early microbial composition is associated with immune development and disease resistance in diverse mammalian systems. This suggests that the microbiota may affect survival outcomes (Belkaid & Hand, 2014; Worsley et al., 2021).

On the other hand, the link between changes in the microbiota composition in the mouth or intestines and early death in reindeer calves is yet unknown. It is still not clear if the general structure of the community, select taxa, or small changes in the community's composition have an effect on microbial signals that are linked to survival. To accurately discern these signals, the application of multivariate analytical techniques is essential (Anderson, 2001; Mallick et al., 2021; Zhou et al., 2022a).

Additionally, microbiome research frequently uses a solitary statistical methodology, which may introduce bias due to variations in sensitivity and the foundational assumptions of the methodology. Thus, a thorough evaluation of survival-related microbiome patterns requires integrative analyses employing many complementary approaches (Mallick et al., 2021).

This thesis provides a comprehensive characterization of the oral and rectal microbiota in reindeer calves and evaluates their associations with sex and early-life survival using contemporary bioinformatics and statistical techniques.

2. Aim and Objectives

Aim

This study aims to describe the oral and rectal microbiota of reindeer using 16S rRNA metabarcoding and to explore how microbiome composition relates to sex and early-life survival.

Objectives: The objective of this study is:

- To analyze raw paired-end 16S rRNA sequencing data using the DADA2 bioinformatics workflow, including quality filtering, error modeling, denoising, sequence inference, chimera removal, and taxonomic assignment.
- To evaluate and compare microbial diversity across sampling locations (oral and anal), sex (female and male), and survival status by analyzing alpha diversity (Observed ASVs and Shannon diversity) and beta diversity using Bray–Curtis dissimilarity, ordination methods, multivariate statistical analysis (PERMANOVA), and GLM.
- To identify genera linked to early-life survival using different differential abundance methods, such as ALDEx2, MaAsLin2, and ANCOM-BC.
- To investigate possible "signature taxa" related to survival and find out if survival trends are due to certain species or changes in the overall community composition.

3. Literature Review

3.1 Host-associated microbiota and early-life development

The establishment of host-associated microbiota during early life is a critical determinant of long-term health and fitness. Early microbial colonization influences immune system development, metabolic processes, and resistance to infections (Sommer & Bäckhed, 2013; Belkaid & Hand, 2014). Disruptions during this initial colonization period can have lasting effects on host physiology and disease susceptibility (Worsley et al., 2021; Ding et al., 2025).

In ruminants, the gut microbiota plays a crucial role in the degradation of plant material and energy extraction. Studies in cattle and sheep have demonstrated that the composition of microbial communities influences feed efficiency, growth, and immune function (Tardiolo et al., 2025; Ding et al., 2025). However, the extent to which the early-life microbiome predicts survival in natural or semi-natural ecosystems remains uncertain.

3.2 Oral versus gut microbiota

Although gut microbiota has garnered significant attention, mouth microbiota remains relatively underexplored, especially in animal species. The mouth cavity constitutes a unique microbial habitat marked by constant environmental exposure and intimate contact with the host mucosal (Dewhirst et al., 2010; Ptasiewicz et al., 2022).

The composition of oral microbiota in humans and domestic animals has been linked to many local and systemic health effects, encompassing inflammatory and metabolic disorders (Baker et al., 2024; Ptasiewicz et al., 2022). Oral bacteria may serve as a source of colonization for subsequent gastrointestinal communities, suggesting a role in shaping gut microbiota formation (Baker et al., 2024; Du et al., 2023).

There are not many studies linking microbiota composition to wildlife survival, but this area of research is growing. Recent reports have demonstrated links between gut microbiota and survival in wild mammal populations, with these associations often involving subtle changes in the community structure rather than dramatic shifts in individual taxa (Leclaire et al., 2023; Worsley et al., 2021). These findings underscore the importance of multivariate techniques, including ordination and distance-based methods, for identifying fitness-related microbiome patterns within complex ecological systems (Anderson, 2001; Mallick et al., 2021).

3.3 Bioinformatics and statistical approaches in microbiome research

Modern microbiome research increasingly uses amplicon sequence variant (ASV)-based pipelines such as DADA2, which offer greater resolution and improved reproducibility compared with traditional operational taxonomic unit (OTU) clustering techniques (Callahan et al., 2016). Common ways to analyze data include measuring alpha diversity, using ordination methods such as principal coordinates analysis (PCoA) and non-metric multidimensional scaling (NMDS), and using permutational multivariate analysis of variance (PERMANOVA) to examine differences at the community level (Anderson, 2001).

Differential abundance approaches like ALDEx2, MaAsLin2, and ANCOM-BC have been developed to address compositional and related issues that can compromise the microbiome (Mallick et al., 2021; Nearing et al., 2022). Utilizing various analytical methodologies to discern substantial, physiologically relevant connections across bacterial taxa is becoming more prevalent (Mallick et al., 2021; Nearing et al., 2022).

Amplicon sequence variations (ASVs) are supplanting operational taxonomic units (OTUs) as the preferred representation of microbial sequences in microbiome studies. OTU-based approaches group sequences based on a specific similarity criterion, which is commonly 97% (Nguyen et al., 2016). This method puts similar sequences together; the choices for clustering could mix up sequences that are physiologically different or separate sequences that are the same. Instead, ASV-based methods use simulated sequencing errors during denoising to figure out the biological sequences in the data.

The DADA2 pipeline employs ASV methodologies to enhance taxonomic resolution and ensure study reproducibility. ASVs are exact sequence variants; you don't need to specify grouping thresholds for each dataset to compare study results. This makes studies of microbial ecology easier to understand and more reliable. ASVs also help you find microbial taxa more accurately, which is important when looking at small changes in a community.

Normalization is another important way to study the microbiome. Due to differences in sequencing depth, sequencing datasets have different numbers of reads per sample. Without normalization, higher sequencing depth may make samples look more diversified or plentiful. Relative abundance transformation or compositional data modifications (like centered log-ratio transformation) are often used to make samples comparable and stop biases that can happen when sequencing depth is uneven (Zhou et al., 2022a).

Researchers use taxonomy assignment to find microorganisms in sequencing data for microbiome research. After processing and ASV inference, each sequence needs to be compared to a reference database for taxonomy classification. This approach sorts of microorganisms into groups based on their genera, families, and phyla.

A strong reference database is essential for taxonomy assignment. RDP, SILVA, and Greengenes contain taxonomy data and validated ribosomal RNA sequences. Curated resources have fewer sequences with improper annotations and a more uniform taxonomy than GenBank (Edgar, 2018). Thus, curated databases improve microbiome analysis reliability and repeatability (Cabezas et al., 2024).

Microbiome data should be grouped by taxonomic group to determine how ecological or biological factors affect microbial populations. In host-microbiome research, a valid classification approach is needed to discover microbial taxa connected to health, sickness, or survival.

3.3.1 Shannon alpha and beta diversity in microbiome studies

When people discuss microbial diversity, they commonly employ alpha and beta diversity measurements. These two kinds of measures tell us distinct but helpful things about how microbial communities are put together (Jeon et al., 2025). Beta diversity looks at how different the microbial communities are in samples. Bray–Curti's dissimilarity and other distance measures illustrate the similarity of sample microbial communities (Baker et al., 2024). You can see these distances with PCoA or non-metric multidimensional scaling.

Diversity analysis helps microbiome studies because the level of sequencing and the number of samples taken can be different (McCoy & Matsen, 2013). Even if the sequencing coverage is different, diversity estimations let researchers compare microbial communities in different samples (McCoy & Matsen, 2013). They also show how ecosystems work and how microorganisms work (Delgado-Baquerizo et al., 2016).

From a bioinformatics point of view, core microbiome analysis can help make complicated datasets easier to understand by focusing on taxa that are always present in samples. This method can help find biologically important patterns and help find microbial taxa that may be linked to certain host features.

3.3.2 Generalized Linear Model

Researchers studying microbiomes often utilize generalized linear models (GLMs) to ascertain relationships between microbial characteristics and host outcomes, including survival or disease(Zhang et al., 2024). Logistic regression, a type of extended linear model, works best when there are just two potential outcomes. It might include things like ordination axes or diversity indexes(Xia et al., 2013). GLMs are a strong and flexible way to find links between the organization of microbial communities and biological processes(Zhang et al., 2024).

3.3.3 Differential abundance analysis

Microbial taxa with different abundances between survival groups or sampling sites are identified via differential abundance analysis. Microbial taxa connected with biological or environmental variables are often found using this type of analysis in microbiome research.

ALDEx2 accounts for compositional effects via centered log-ratio transformations and Monte Carlo sampling. MaAsLin2 finds microbial taxa-explanatory variable relationships using multivariable linear models. ANCOM-BC uses compositional microbiome-specific bias-corrected log-linear models (Zhou et al., 2022b).

Different statistical procedures make different assumptions and sensitivities; hence, many differential abundance methods are recommended. Comparing data across methodologies can boost microbial association confidence and reduce false positives (Zhang et al., 2020).

3.3.4 Core microbiome

The core microbiome comprises microbial taxa that are consistently present in a group of hosts or their surroundings. (Neu et al., 2021) These taxa are usually described by setting limits on their frequency and abundance, such as being present in a particular percentage of samples.

Finding the core microbiome can assist scientists in figuring out which microbial species are stable members of a community, and which ones change more often. People generally think that core taxa might be vital for keeping ecosystems stable or host health (Götz et al., 2020). In investigations of microbiomes connected with hosts, core microbial taxa might include microorganisms that have consistent connections with the host.

4 Materials and Methods

4.1 Sampling

Samples were collected from reindeer calves during calf marking at the end of June. A mouth microbiota sample was taken from the side of the cheek and along the back teeth of the calf using an eSwab 840CE (Copan Italia). The rectal microbiota was collected by inserting a swab in the anus. All samples were collected within a three-hour window between 22 and 01 at night. After collection, the swabs were held in a chilled cooling box until they could be frozen at -18 °C and transported to the laboratory.

4.2 Ethical permit

Ethical permit for the sampling was obtained from the Umeå djurförsöksetiska nämnd (Umeå, Sweden), permit number: Dnr 16-24.

4.3 Molecular methods

4.3.1 DNA extraction

Before DNA extraction, the eSwabs 480CE (Copan Italia) were centrifuged for 10 min at 4500RPM to create a pellet with the gathered material. After centrifuging, 750 µl was discarded. The remaining volume (max 250 µl) was used as a sample for DNA extraction.

The DNA extraction was conducted using the DNeasy PowerLyzer Powersoil kit (Qiagen AB). This kit is normally used to extract DNA from soil, but is commonly used for extracting bacterial DNA from various samples as a source of DNA. The protocol has hence been adapted to improve the yield from these samples by merging three steps in the protocol. The modification is to add reagents C1, C2, and C3 in a single step (Mattei et al., 2019). This mix was vortexed for 10 min at maximum speed (Mattei et al., 2019) and then following the extraction protocol accordingly. The final elution volume was 100 µl, and the extracted DNA was stored at -20 °C until PCR.

4.3.2 Two-step PCR

The PCR reaction for each sample was performed in duplicate to decrease the risk of PCR bias. The first PCR reaction that amplifies a part of the 16s rRNA gene was conducted in 20 µl reactions consisting of 0.4 µl 10mM dNTPs, 2 µl Phusion Buffer, 0,6 µl BSA, 0.25 µl primer 515F and 806R respectively, and 0.2µl Phusion ™ Hot

Start II DNA polymerase (2U/µl) (Thermo Fischer Scientific, Waltham, USA) and 16.6µl ddH2O. To each reaction, 1 µl of DNA was added. The PCR cycle consisted of 1 cycle of 98 °C for 30s, 24 cycles of 98 °C for 10s, 56 °C for 20 s and 72 °C for 15s, and a final elongation at 72 °C for 8 min.

In the second step of PCR, unique barcode combinations were added to the PCR products from the first PCR. For a 30µl reaction, 0.6 µl 10mM dNTPs, 3 µl Phusion Buffer, 0.3 µl Phusion ™ Hot Start II DNA polymerase (2U/µl) (Therm Fischer Scientific, Waltham, USA), and ddH2O to 27.5 µl. To each reaction, 0.5 µl PCR product and 1 µl of each uniquely barcoded forward and reverse Illumina primer were added. The PCR cycle was conducted as in the first PCR.

The duplicated PCR reactions from PCR 2 were merged into one well. After that, eight samples were pooled by adding 25 µl of each sample to an Eppendorf tube. Primers and PCR reaction reagents were removed by adding 180 µl Agencourt AMPure XP (Beckman-Colter) to the 200 µl PCR product pool, giving it a ratio of 1:0.9. After mixing, a 5-minute incubation was implemented. The magnetic bead mix was separated on a magnetic rack, and the supernatant was removed. The pelleted magnetic beads were washed twice with 400 µl 70% ethanol. After the second wash, the magnetic beads were left to dry, and then the DNA was eluted by adding 100 µl 0.05M Tris-HCl. After a 5-minute incubation time, the Eppendorf tube was placed on a magnetic rack, after which the supernatant was transferred to a new Eppendorf tube.
The DNA concentration of the pooled and eluted DNA samples was measured and pooled in equal amounts per pooled sample. The concentration of the final pool was measured on a TapeStation.

The sequencing was performed at the National Genomics Infrastructure in Stockholm, financed by Science for Life Laboratory. For sequencing, P1xLeap (2x300 bp) was used on an Illumina NextSeq 2000. After sequencing, the sequences were delivered demultiplexed into separate files.

4.4 Bioinformatic processing

I used the “plotQualityProfile” function from the DADA2 package version 1.38 (Callahan et al., 2016) to check the quality of the raw sequencing reads for the first time. Given that quality scores tended to drop near the ends of reads, trimming and filtering settings were chosen to remove low-quality bases while maintaining sufficient overlap for paired-end merging.

The “filterAndTrim” function was used to filter and trim reads, applying site-specific truncation lengths to both forward and reverse reads. I removed reads with ambiguous bases, excessive predicted errors, or lengths below the required length. Only samples that passed quality control for both forward and reverse readings were kept for later analyses.

The “learnErrors” function was used to learn sequencing error rates for both forward and reverse reads. I examined error-frequency graphs to ensure that the correct error model was fitted.

I used the main DADA2 “denoising function” dada to identify amplicon sequence variants (ASVs) after dereplication, which combined identical sequences within each sample. After that, “mergePairs” was used to combine the forward and backward reads. This required that the overlaps be identical. Using the consensus technique with the “removeBimeraDenovo” function, chimeric sequences were found and removed.
The ASV table (sequence table) generated from this showed high-resolution, error-corrected biological sequences for all samples.

The SILVA reference database version 1.38 (Quast et al., 2012) was used to assign ASVs to taxonomic groups. SILVA is a comprehensive reference database of ribosomal RNA sequences used to identify and classify microorganisms such as bacteria, archaea, and eukaryotic microbes. The “assignTaxonomy” function in DADA2 was used to assign taxonomy from kingdom to genus level. For further investigation, only ASVs linked to bacterial taxa were retained.

A phyloseq object was created by combining the final ASV abundance table, taxonomic assignments, and cleaned metadata using the phyloseq package version 1.5 (McMurdie & Holmes, 2013). I didn't include samples without metadata or with a sequencing depth of zero. The phyloseq item served as the foundation for all ecological and statistical investigations.

4.5 Relative Abundance Analysis

I calculated proportions from the count data for each sample to determine relative abundances. I used stacked bar plots and average abundance plots to visualize and summarize the relative profiles at the phylum and genus levels across sampling sites, sex, and survival groups.

4.6 Alpha and Beta Diversity Analysis

I measured alpha diversity using the Shannon diversity index and the number of observed ASVs (richness). The Shannon index is used in microbiome and ecological studies because it provides a single value that reflects both how many different taxa are present and how evenly they are represented. “Estimate_richness” from phyloseq was used to calculate diversity metrics. I compared alpha diversity between different places where samples were taken, specifically between the mouth and anus, between females and males, and between survivors and non-survivors.

I used Bray–Curtis dissimilarity, which relies on relative abundance data, to measure beta diversity. I applied Principal Coordinates Analysis (PCoA) and Non-metric Multidimensional Scaling (NMDS) for ordination. To determine if there were differences in the composition of microbial communities, I employed PERMANOVA “adonis2” from the vegan package version 2.7 with models including sample site, sex, and survival as explanatory factors. For observational data, marginal effects were assessed using by = "margin". I used “betadisper” to evaluate homogeneity of dispersion and ensure that within-group dispersion did not influence the PERMANOVA results.

4.7 Generalized Linear Model (GLM)

To evaluate whether microbial community structure was associated with survival, generalized linear models (GLMs) with a binomial distribution were fitted. Survival status (0 = non-survivor, 1 = survivor) was used as the response variable, and ordination axes (PCoA1, PCoA2, and PCoA3) were included as predictor variables. Model coefficients, standard errors, and p-values were used to assess the significance of associations.

4.8 Differential Abundance Analyses

I did differential abundance analysis at the genus level to find taxa that were linked to survival status (Nearing et al., 2022). I used Wilcoxon tests with Monte Carlo sampling to look at data that had been changed using ALDEx2 version 1.4 (Zhou et al., 2022a) centered log-ratio. I looked at both raw p-values and p-values that had been modified for false discovery rate (FDR). I used MaAsLin2 version 1.1 multivariable linear models with survival as the main explanatory variable. I also used FDR correction to adjust for multiple testing. ANCOM-BC version 2.1 was also used to fix any bias.

I employed bias-corrected log-linear models to find genera that were linked to survival, considering compositionality and structural zeros.

The tables showed the genera that one or more analytical approaches thought were important. The resilience of survival-associated taxa was assessed by evaluating the effectiveness of several combined methodologies.

Using all three methods—ALDEx2, MaAsLin2, and ANCOM-BC—gives a more complete picture of how varied things are in terms of abundance. Each method has its own strengths and assumptions. For example, ALDEx2 deals with compositional data, MaAsLin2 uses multivariable modelling, and ANCOM-BC fixes biases that come from compositional data and structural zeros. Comparing data from these many methods makes it more likely that the relationships found are not just the consequence of one analytical strategy. The overlap among approaches indicates that the findings are robust and less susceptible to methodological constraints or biases.

4.9 Core Microbiome Analysis

The microbiome R package version 1.32 (Ruan et al., 2022) helped me identify the core microbiota. Core taxa were defined as those present in at least 70% of the samples within a group and with a relative abundance of at least 0.01%. I calculated the core microbiomes for each survival group and sample site separately to find species consistently linked to survival.

N.B Statistical Software and Visualization

All analyses were conducted in R version 4.5, and figures were generated using the package ggplot2 version 4.0 (Wickham, 2016) and the package phyloseq version 1.5 (McMurdie & Holmes, 2013) version 1.5 visualization tools. Statistical significance was assessed at an alpha level of 0.05 unless otherwise stated. The R code for the analysis is located in the Appendix.

5 Results

5.1 DADA2 Bioinformatics Workflow and Data Processing

The DADA2 pipeline was used to process raw paired end 16S rRNA sequencing data in this bioinformatics study. Quality profile evaluations of both forward and backward scans showed that most spots had great base quality, which made it easy to trim and filter correctly. As expected, the error models that were generated during the workflow came together. This showed a significant connection between the errors that were seen and those that were predicted. This offered us a good starting point for removing noise and figuring out the order. After filtering and getting rid of chimera, there were 108 samples remaining. This generated a high-quality dataset that may be used for more ecological research.

Raw paired-end reads showed good overall quality based on the forward and reverse quality profile plots.

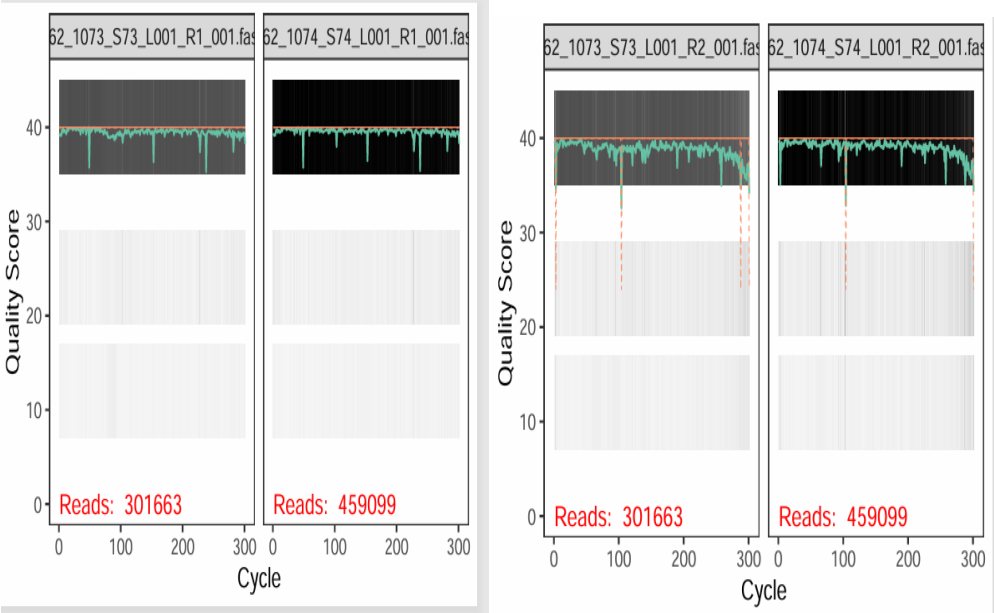

Figure 1. Quality profile forward plot and Quality profile reverse plot

The DADA2 error model (Figure 2) showed the expected relationship between observed and estimated errors, supporting successful learning of error rates before denoising.

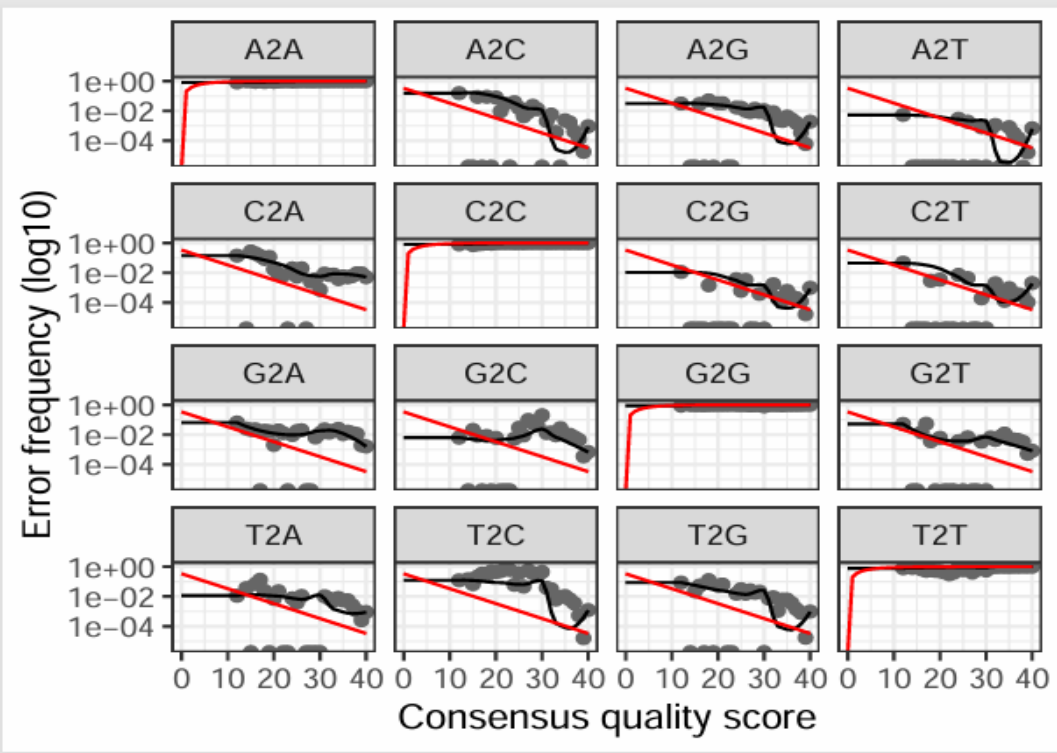

Figure 2. The error frequency plot shows the relationship between observed and estimated errors.

5.2 ASV Tables and Sample Metadata Integration

I made 77900 ASV tables and was able to successfully combine them with host sample metadata. The dataset had 778 bacterial genera spread out over 108 samples (54 from the mouth and 54 from the rectum) after deleting sequences that weren't bacterial or were contaminants. Taxonomic classification at the phylum level indicated that oral and rectal samples had persistent, different patterns. These phylum-level patterns were similar across sexes and survival groups, indicating that significant community transitions were not occurring at this high taxonomic level, and that survival signals likely lay in lower-level taxonomic changes.

5.3 Taxonomic composition (relative abundance)

5.3.1 Phylum-level composition by site and sex

Figure 3 shows the average bacterial kinds in each sample. *Fusobacteria* are the dominant phylum in the rectal sample in both female and male sexes. *Proteobacteria* are dominant in the mouth sample in both female and male sexes. The samples are categorized by sex (male and female) and by anus and mouth location. Each bar shows how abundant each bacterial type is in each sample by color.

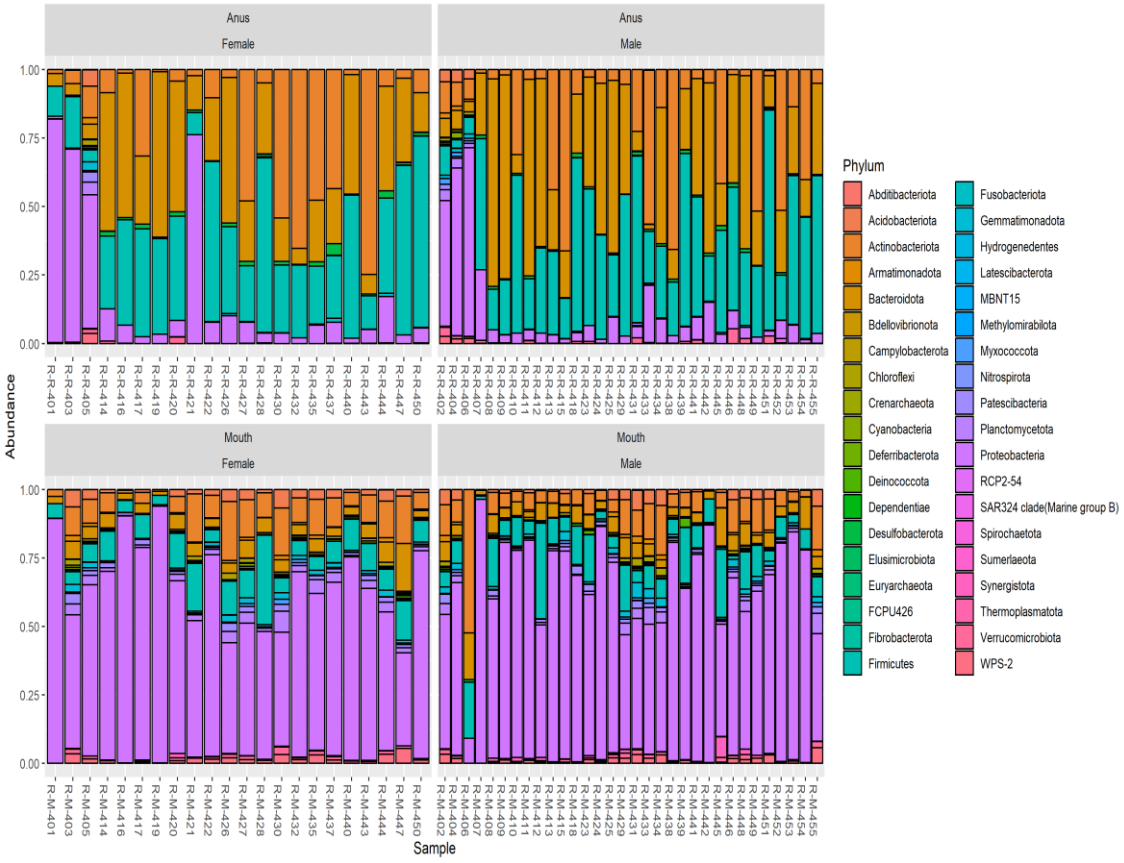

Figure 3. Phylum-level composition by site and sex

This figure shows a comparison of mouth and rectum microbial communities and sexes.

5.3.2 Phylum-level composition by site, and survival

Figure 4 shows how many bacteria are in each sample by sampling site (anus or mouth) and survival status (0 = non-survivor, 1 = survivor). *Fusobacteria* are the dominant phylum in the rectal sample in both survival statuses. *Proteobacteria* are dominant in the mouth sample in both survival statuses. Each bar represents a sample and shows bacterial phyla abundance in color..

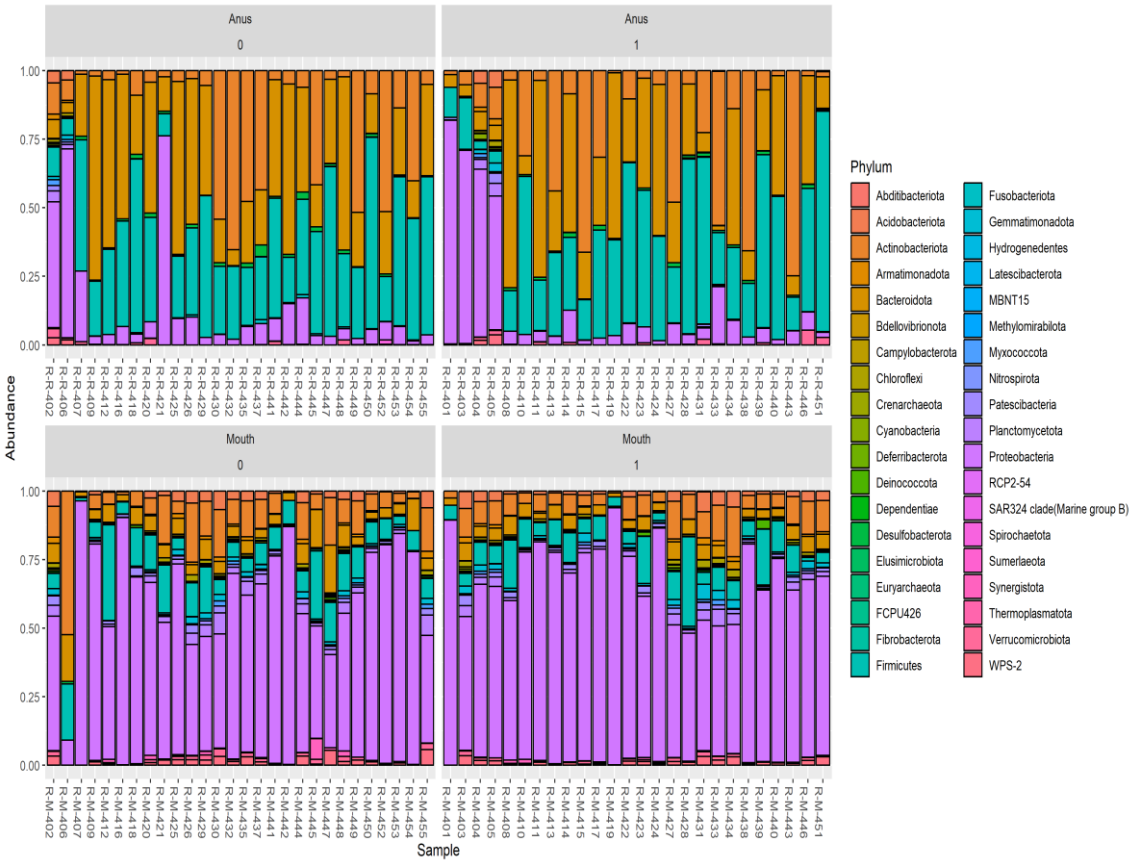

Figure 4. Phylum-level composition of microbiota by sampling site and survival status

This figure shows a comparison of mouth and rectum microbial communities and survival rates.

5.4 Alpha diversity

5.4.1 Shannon diversity (mouth vs anus)

Shannon diversity differed between sampling sites, with clear differences visible between mouth and anus samples in the site-level alpha diversity plots.

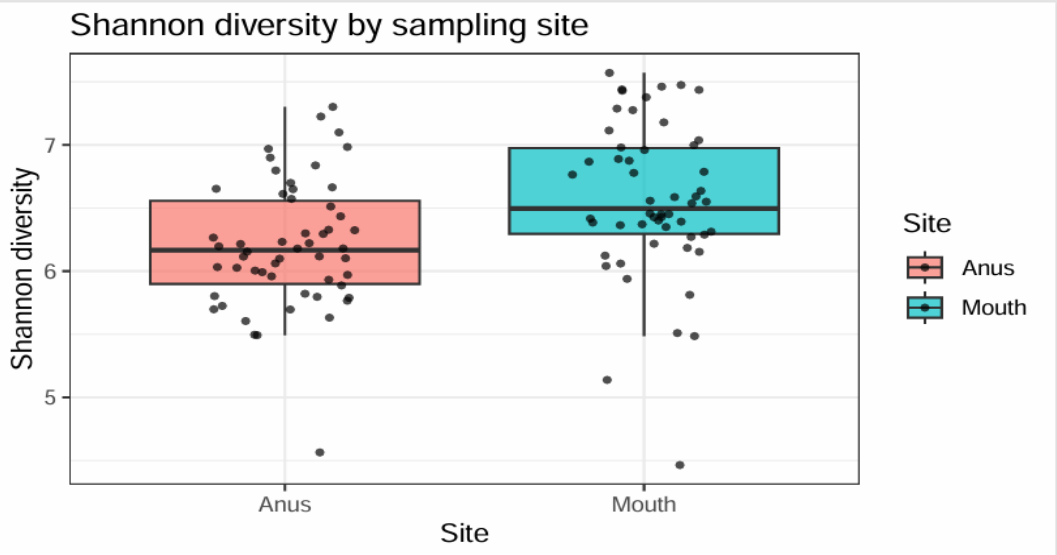

Figure 5. Alpha Shannon diversity by sampling site reflects variation in microbial diversity between oral and rectal microbiota.

Boxplot of Shannon's alpha diversity index comparing microbial diversity between mouth and anus samples. The boxplots display the median value and the upper and lower quartiles of Shannon diversity for each sampling site. The whiskers represent the range of the data, and the dots represent individual samples. Differences between sites reflect variation in microbial diversity between oral and rectal microbiota.

5.4.2 Mouth: Shannon diversity by sex and survival

Within mouth samples, Shannon diversity showed broad overlap between females and males.

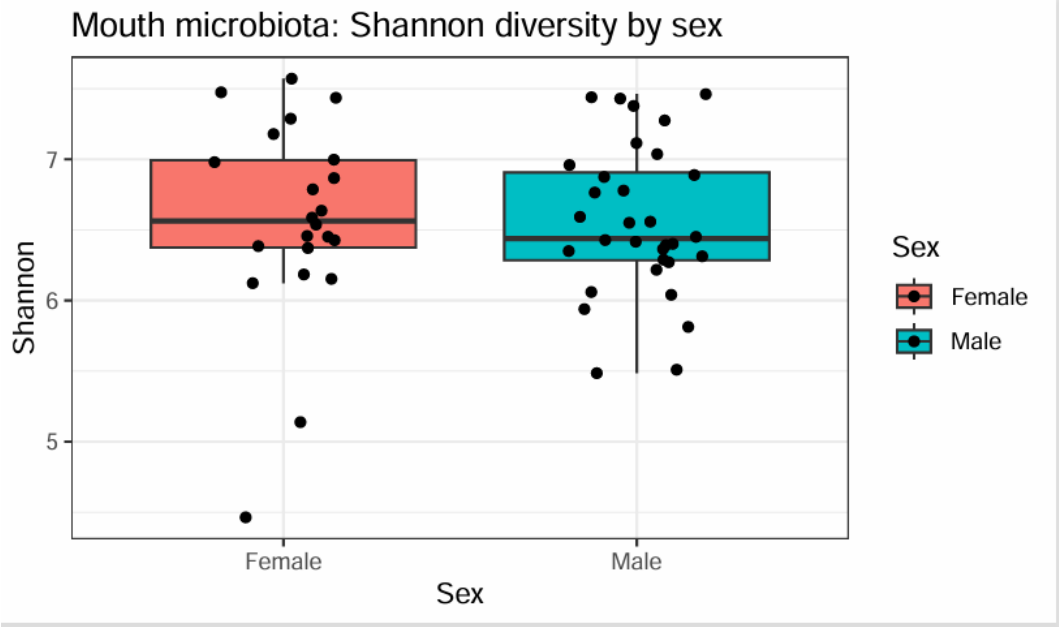

Figure 6. The box illustrates the distribution of microbial diversity in oral microbiota between female and male calves.

Boxplot of Shannon's alpha diversity index for mouth samples grouped by sex (female and male). The boxplots represent the median and the upper and lower quartiles of Shannon diversity within each group. Individual dots correspond to individual calf samples. The plot illustrates the distribution of microbial diversity in oral microbiota between female and male calves.

Similarly, Shannon diversity overlapped strongly between survival groups (0=dead vs 1=alive).

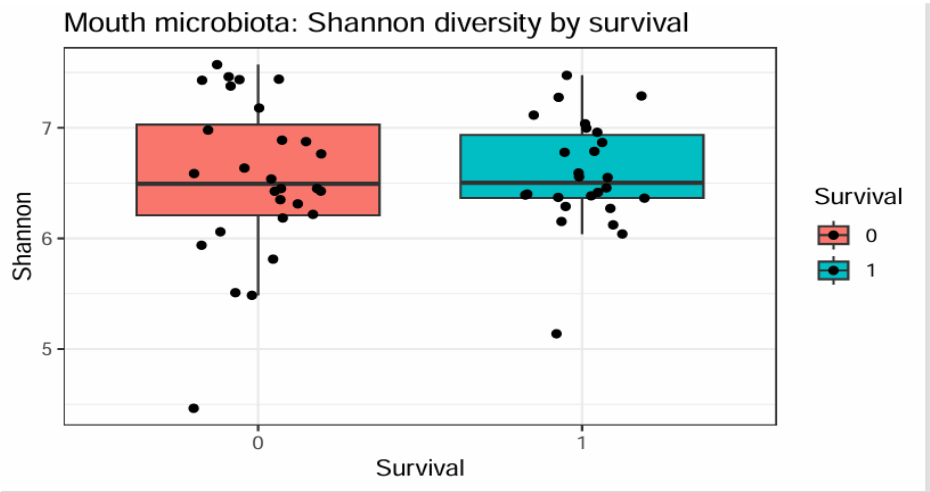

Figure 7. The figure illustrates the variation in oral microbiota diversity between calves that survived and those that did not.

Boxplot of Shannon's alpha diversity index for mouth samples grouped by survival status (survivors=1 and non-survivors=0). The boxplots show the median and the upper and lower quartiles of Shannon diversity, while the dots represent individual samples. The figure illustrates the variation in oral microbiota diversity between calves that survived and those that did not.

5.4.3 Anus: Shannon diversity by sex and survival

Within the anus samples, Shannon diversity again showed strong overlap between females and males.

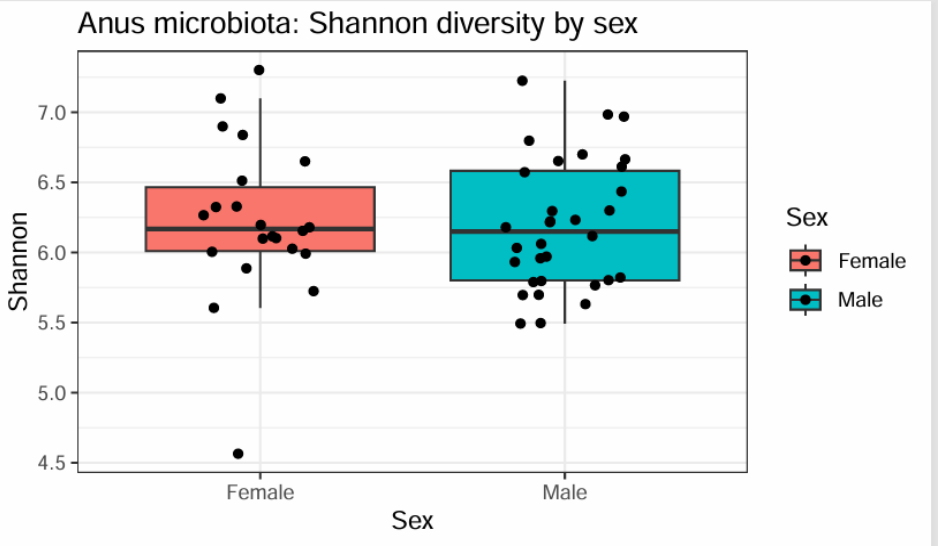

Figure 8. The boxplots represent the median anus microbiota Shannon diversity by sex

Boxplot of Shannon's alpha diversity index for rectal (anus) samples grouped by sex (female and male). The boxplots represent the median and the interquartile range of Shannon diversity, and the dots represent individual samples. This figure illustrates the distribution of microbial diversity in rectal microbiota across sexes.

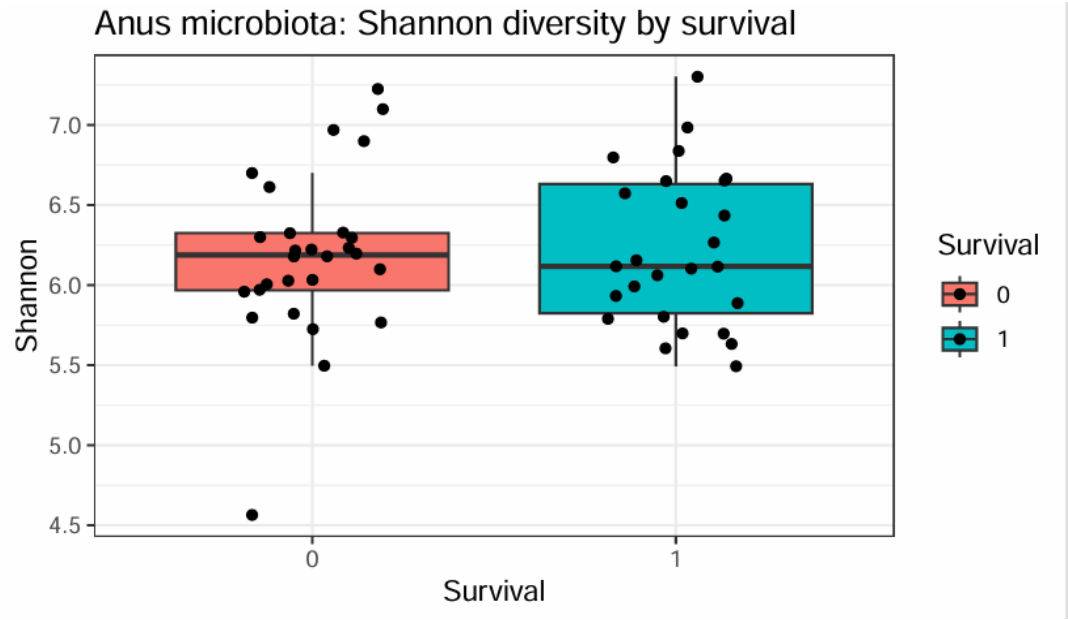

Figure 9. The boxplots indicate the median and quartiles of Shannon diversity of the Anus microbiota by survival

Boxplot of Shannon's alpha diversity index for rectal samples grouped by survival status (survivors and non-survivors). The boxplots indicate the median and quartiles of Shannon diversity, while individual dots represent samples from individual calves. The figure illustrates the variation in rectal microbiota diversity between survival groups.

5.5 Beta Diversity and Microbial Community Composition Analysis

5.5.1 Bray–Curtis Ordination

Beta diversity was assessed using Bray–Curtis ' dissimilarity, visualized by **PCoA** and **NMDS**.

Stress values (Mouth: 0.146; Anus: 0.054) indicated acceptable and excellent ordination fits, respectively.

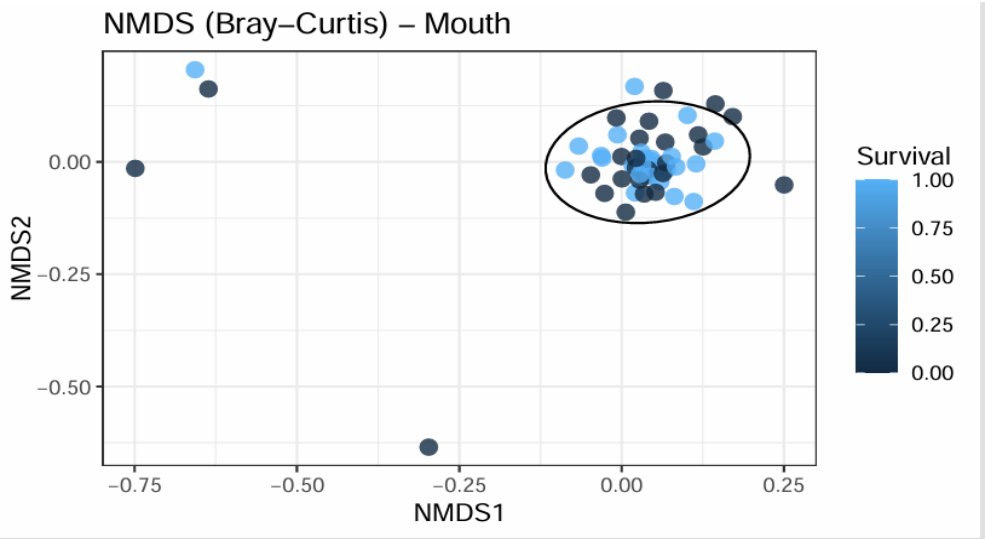

Figure 10. NMDS-Mouth shows ordination fit on the right top of the plot

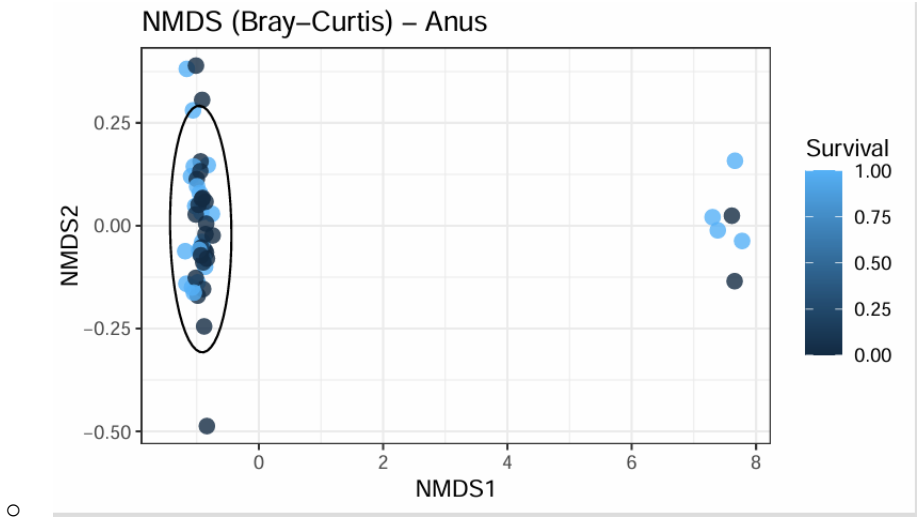

Figure 11. NMDS-Anus shows on the left top of the plot

5.5.2 PERMANOVA

A Bray–Curtis distance matrix was analyzed using PERMANOVA, including Site, Sex, and Survival. The overall model was significant ($R^2 \approx 0.112$; $p = 0.001$), indicating that microbial community composition varies across the included predictors, with site expected to be the main contributor.

5.5.3 Mouth samples: PCoA by sex and survival

In mouth samples, PCoA1 and PCoA2 explained ~6.1% and ~5.2% of the variation, respectively. Visual separation by sex was weak, and groups overlapped

substantially.

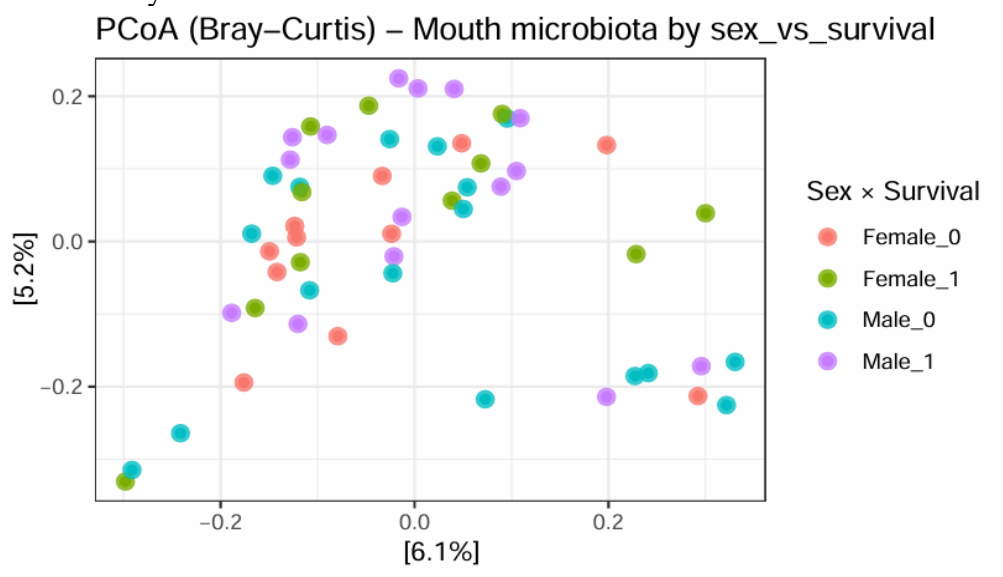

Figure 12. PCoA- Mouth Visual separation by sex was weak, and groups overlapped

PCoA colored by survival also showed substantial overlap, suggesting only weak survival-related compositional shifts at the community level.

5.5.4 Anus samples: PCoA by sex and survival

In the anus samples, PCoA1 and PCoA2 explained ~17.1% and ~10.5% of variation, respectively, reflecting stronger compositional gradients. However, separation by sex remained weak, with a large overlap between groups.

PCoA by survival similarly showed overlap, suggesting no strong survival-associated clustering in the anus microbiota.

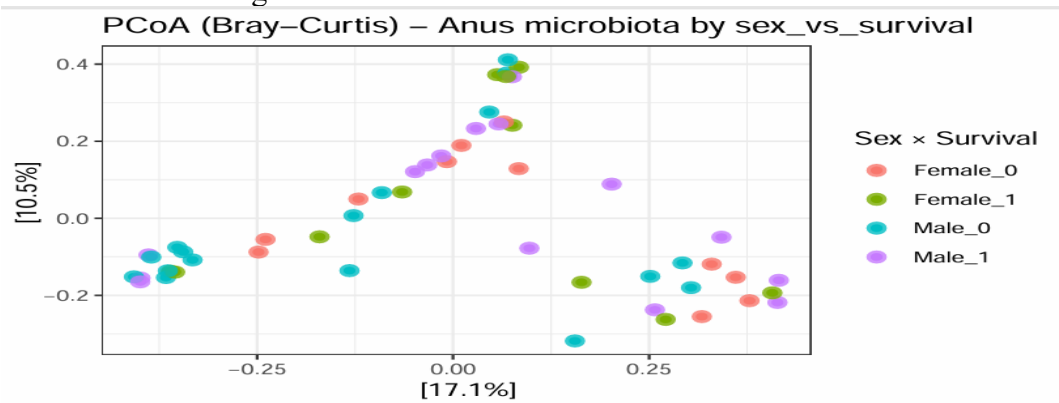

Figure 13. PCoA-Anus reflecting stronger compositional gradients

5.6 Investigation of Signature Taxa and Community Trends

The investigation of "signature taxa" and community trends utilized generalized linear models (GLMs) and core microbiome analysis.

5.6.1 GLM Survival Prediction

Rectal (Anus) microbiota

A generalized linear model with a binomial distribution was used to test whether variation in rectal microbiota composition (represented by PCoA axes) was associated with calf survival. None of the ordination axes were significantly associated with survival (PCoA1: $\beta = 0.91$, $p = 0.402$; PCoA2: $\beta = 1.31$, $p = 0.346$; PCoA3: $\beta = 0.16$, $p = 0.924$). These results indicate that microbial community composition in rectal samples did not significantly predict early-life survival in this dataset.

Oral (Mouth) Microbiota

A generalized linear model with a binomial distribution was used to test whether variation in oral microbiota composition (represented by PCoA axes) was associated with calf survival. Among the ordination axes, **PCoA2 showed a significant association with survival** ($z = 2.03$, $p = 0.042$), whereas **PCoA1 ($p = 0.981$)** and **PCoA3 ($p = 0.473$)** were not significant predictors. These results suggest that variation along the second ordination axis of the oral microbial community was weakly associated with early-life survival.

5.6.2 Core Microbiome Analysis

Using a conservative definition of the core microbiome (≥70% prevalence, ≥0.01% relative abundance): **No core taxa were detected in rectal samples**, regardless of survival status. **Two survivor-specific core ASVs were detected in mouth samples**, both classified as ***Alysiella*** (family *Neisseriaceae*).

This site-specific pattern supports the GLM results, indicating that survival-related signals were present primarily in the oral microbiota.

5.6.3 Relative Abundance of *Alysiella*

The relative abundance of *Alysiella* in mouth samples was high in both survival groups. Mean relative abundance of Non-survivors(0) is **25.0%,** and mean relative abundance of Survivors(1) is **25.4%.**

Although overall abundance was similar between groups, *Alysiella* met the criteria for survivor-specific core taxa due to its consistent prevalence among survivors.

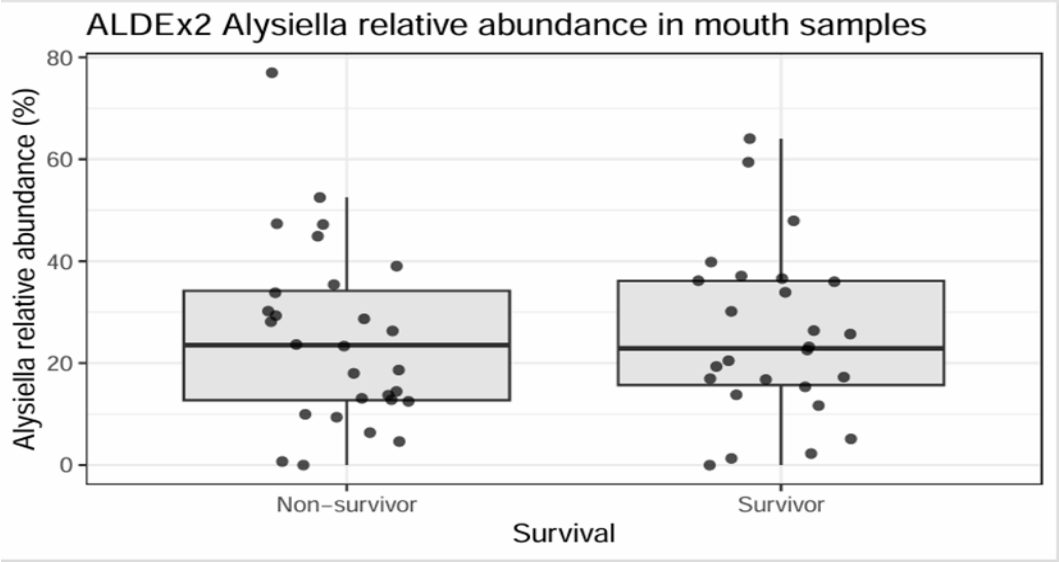

Figure 14. Relative Abundance of Alysiella in mouth samples.

5.7 Survival-associated taxa in the mouth and rectal microbiota

To investigate "signature taxa" potentially associated with survival, genus-level differential abundance (DA) analyses were run for mouth samples and rectal samples using three complementary methods.

5.7.1 Survival-associated taxa in the mouth

ALDEx2

ALDEx2 produced a volcano-style plot of effect sizes versus significance for the survival comparison. A limited number of features exceeded the raw p-value threshold (dashed line), indicating a small set of candidate survival-associated genera. **ALDEx2** identified **3** genera (*Capnocytophaga, Pseudomonas*, *Sphingomonas)*

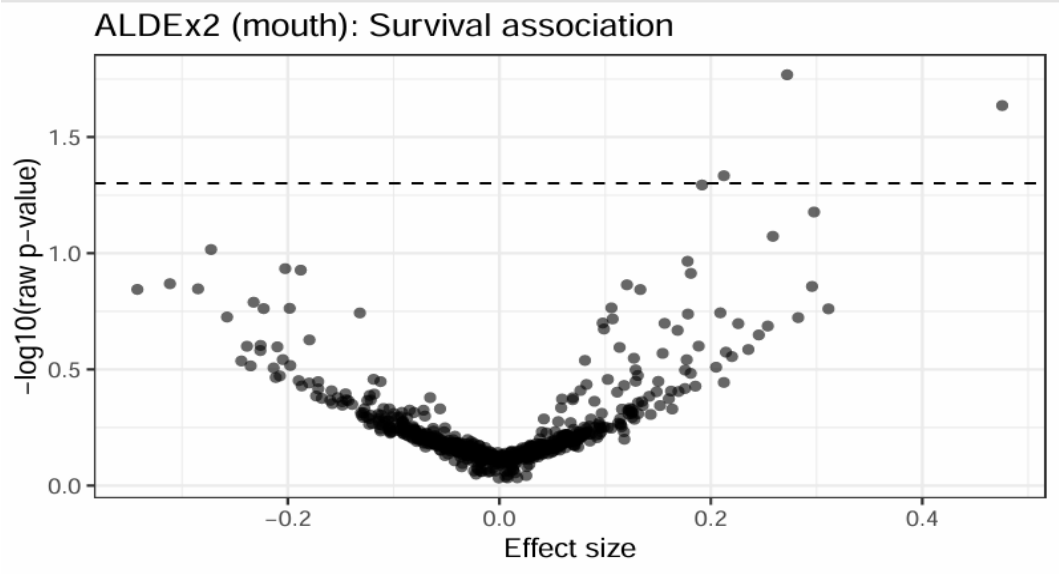

Figure 15. ALDEx2 (mouth) showed a survival association

MaAsLin2 (mouth)

MaAsLin2 also identified a set of candidate genera associated with survival, with some features exceeding the raw p-value threshold. **MaAsLin2** identified **10 genera** (*Capnocytophaga, Xylophilus*, *Olsenella*, *Monoglobus*, *Roseisolibacter*, *Prevotellaceae UCG-001*, *Stakelama*, *Parafilimonas*, *[Ruminococcus] torques group*, *Desulfovibrio)*

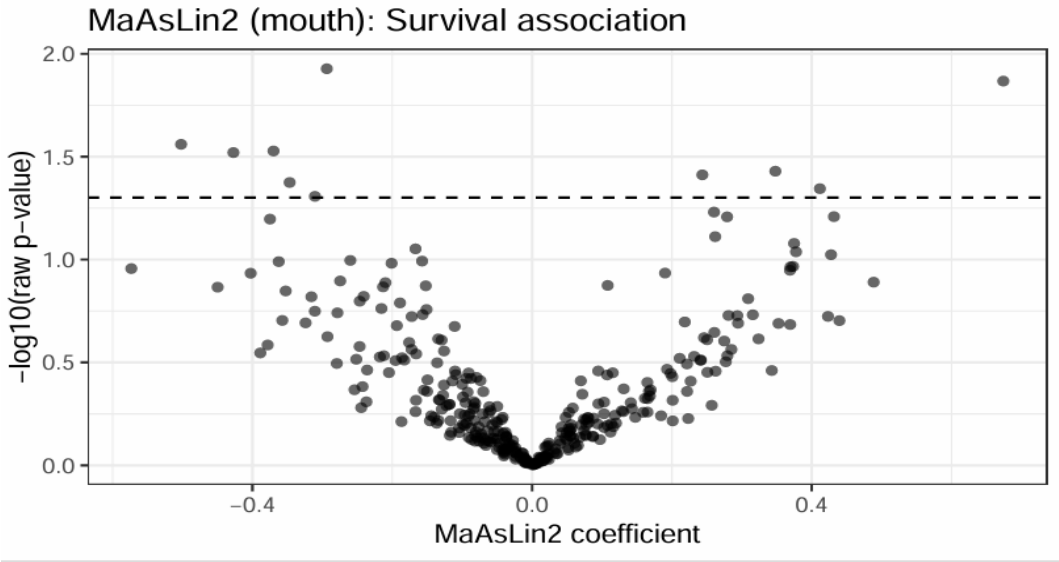

Figure 16. MaAsLin2 (mouth) showed a survival association

ANCOM-BC (mouth)

ANCOM-BC identified additional candidate genera associated with survival, visualized as log fold change versus –log10(raw p-value). **ANCOM-BC**

identified **12 genera** (*Monoglobus, Roseisolibacter, Parafilimonas, Phascolarctobacterium*, *Chitinophaga*, *Flexivirga*, *Pedococcus-Phycicoccus*, *Nocardia*, *Amnipila*, *Colidextribacter*, Alysiella, *Candidatus Berkiella)*

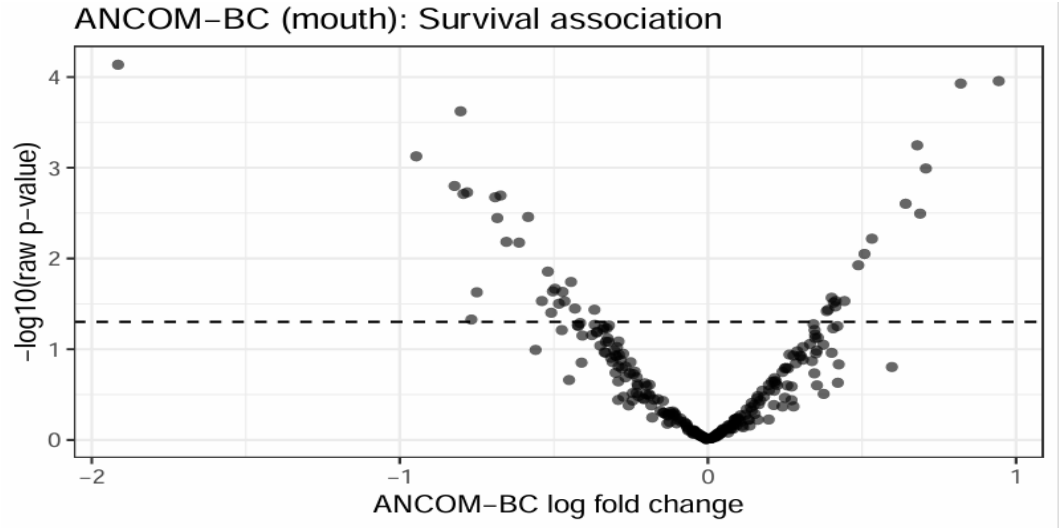

Figure 17. ANCOM-BC (mouth) showed a survival association

1. Table 1. List of genera significantly associated with survival in mouth samples.

Ge nus	**ALDEx2**	**MaAsLin2**	**ANCOM-BC**
Capnocytophaga	✓	✓	×
Pseudomonas	✓	×	×
Sphingomonas	✓	×	×
Xylophilus	×	✓	×
Olsenella	×	✓	×
Monoglobus	×	✓	✓
Roseisolibacter	×	✓	✓
Prevotellaceae UCG-001	×	✓	×
Stakelama	×	✓	×
Parafilimonas	×	✓	✓
[Ruminococcus] torques group	×	✓	×
Desulfovibrio	×	✓	×
Phascolarctobacterium	×	×	✓
Chitinophaga	×	×	✓

Flexivirga	×	×	✓
Pedococcus-Phycicoccus	×	×	✓
Nocardia	×	×	✓
Amnipila	×	×	✓
Colidextribacter	×	×	✓
Alysiella	×	×	✓
Candidatus Berkiella	×	×	✓

5.7.2 Survival-associated taxa in the rectal microbiota

Differential abundance analyses of rectal microbiota did not identify any genera significantly associated with survival after false discovery rate correction using ALDEx2, MaAsLin2, or ANCOM-BC. This result is consistent with the alpha and beta diversity analyses, which showed no clear survival-related differences in rectal samples.

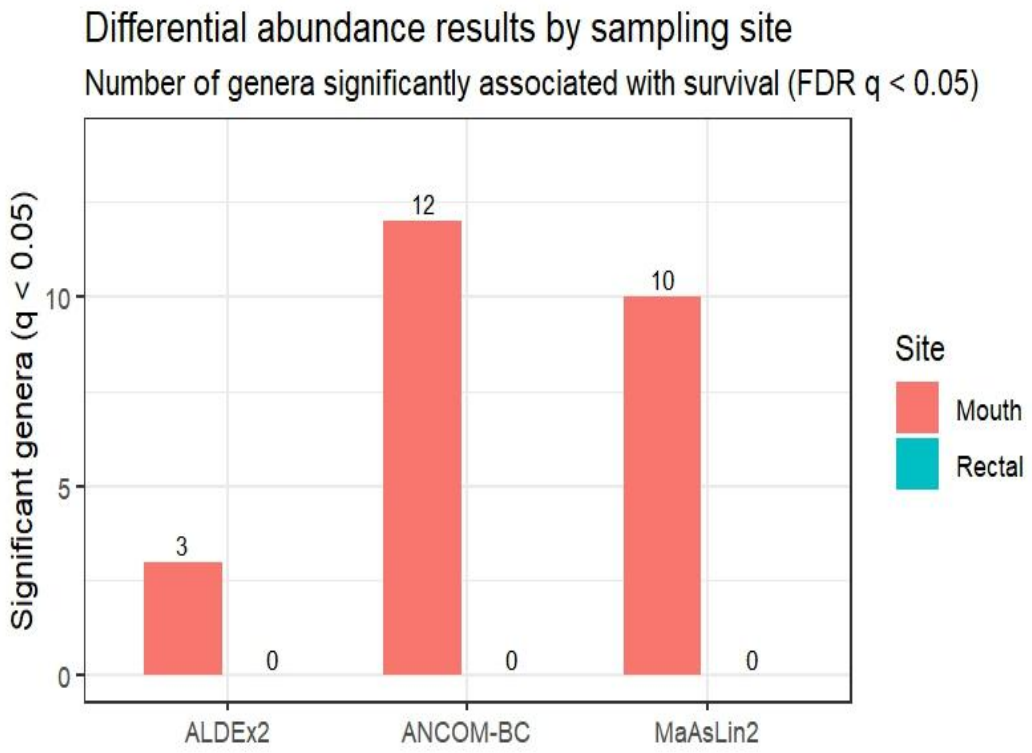

Figure 18. Comparison of the number of genera significantly associated with survival.

Agreement across methods (robustness)

Across DA approaches, a subset of genera was repeatedly detected, while others were method specific. For example, *Capnocytophaga* was detected by both ALDEx2 and MaAsLin2, while *Monoglobus, Roseisolibacter, and Parafilimonas* were detected by MaAsLin2 and ANCOM-BC. Several genera were uniquely detected by ANCOM-BC (e.g., *Chitinophaga*, *Flexivirga*, *Nocardia*, *Candidatus Berkiella*), consistent with differences in method assumptions and sensitivity.

The sign of the effect size (ALDEx2), the model coefficient (MaAsLin2), or the log fold change (ANCOM-BC) determined the direction of the link. Positive numbers indicate a higher prevalence of specific bacteria in surviving calves, whereas negative values signify a greater presence in non-surviving calves. The findings indicate that Capnocytophaga, Monoglobus, and Roseisolibacter were present at elevated levels in surviving calves, implying a correlation with survival. On the other hand, Parafilimonas was more common in those who didn't survive, which suggests a link to reduced survival.

6 Discussion

6.1 Summary of the main findings

The objective of this study was to characterize the oral and rectal microbiota of reindeer calves and to examine their associations with sex and early-life survival using 16S rRNA metabarcoding and bioinformatic analysis.

The results demonstrated pronounced differences in microbial community composition between sampling sites, while sex exerted only a minor influence on microbiota diversity and composition. Analyses of alpha diversity indicated no significant differences associated with survival status in either oral or rectal samples. Likewise, beta diversity analyses showed that survival status accounted for only a small proportion of the variation in microbial community composition. Although survival-related signals were generally weak, a modest association between oral microbiota composition and survival was observed. Logistic regression analysis based on ordination axes revealed a significant association between survival and one oral community axis (PCoA2). In contrast, no such association was identified in rectal samples. Core microbiome analysis identified two survivor-specific core amplicon sequence variants (ASVs) in oral samples classified as *Alysiella*, while no core taxa were detected in rectal microbiota. Differential abundance analyses identified several candidate genera associated with survival in the oral microbiota, although overlap among statistical methods was limited.

Collectively, these findings indicate that early-life survival in reindeer calves is linked to subtle changes in oral microbiota composition rather than substantial shifts in overall microbial diversity or gut microbiota composition.

.

6.2 Interpretation of the results

There were no big differences in alpha diversity between the groups that survived, which suggests that the total number or balance of microorganisms does not determine how many reindeer calves live through their first few months. The results show that changes in the types of microorganisms present are more likely to provide signals about survival than changes in overall diversity. Similar results have been reported in several investigations of vertebrate microbiomes, which suggest that changes in the structure of the microbial community, not alpha diversity, are associated with the host's health (Leclaire et al., 2023; Worsley et al., 2021).

The weak but consistent connection between oral microbiota and survival is probably because of biological events that start early in the calf's life. The mouth is always open to the outside world and the immune system, thus oral microorganisms can swiftly adapt to changes in food, immunity, and surroundings. Because of this,

these bacteria could provide insights about the health or immune system of the calf (Dewhirst et al., 2010; Ptasiewicz et al., 2022).

On the other hand, we did not uncover any significant trends in the rectal microbiota that were connected to survival. This could be because the gut flora in young calves is still developing and can be very different from one calf to the next. In the early years of life, the gut microbiota changes frequently, becoming more stable as calves age (Du et al., 2023; Tardiolo et al., 2025). This makes it tougher to detect survival patterns that are consistent in rectal samples. Also, rectal swabs may not reflect the whole range of gut bacteria as effectively as samples from other gastrointestinal sites (Jones et al., 2018).

The fact that *Alysiella* is a crucial bacterium in oral samples shows that some bacteria may help maintain the stability of the microbial community early in life or during host-microbe interactions. There were almost the same amounts of *Alysiella* in all of the groups that survived, which implies that its relevance may derive from its consistent presence rather than fluctuations in abundance. These data support the hypothesis that microbial signals relevant to survival emerge from a variety of population patterns, not just a single dominant taxon.

6.3 Bioinformatic and statistical methods in microbiome analysis

A thorough bioinformatic methodology and numerous complementary statistical techniques to examine microbial community patterns make this study strong. Since 16S rRNA sequencing measures relative abundances rather than absolute counts, microbiome datasets are complicated and compositional. Thus, analytical tools must handle these data's particular statistical features.

DADA2 infers ASVs from sequencing reads during sequence processing. ASV-based methods provide stronger taxonomic resolution and reproducibility than OTU-based methods. More detailed ecological and statistical analyses were possible with the ASV table.

Alpha and beta diversity measurements measure community diversity. Alpha diversity metrics like the Shannon diversity index reveal sample microbiological richness and evenness. However, beta diversity measurements like Bray–Curti's dissimilarity allow microbial community composition comparisons between samples. These compositional differences were visualized using PCoA and NMDS. PERMANOVA was used to determine if sampling site, sex, and survival status affected microbial communities.

Generalized linear models (GLMs) with ordination axes as variables were used to examine microbial community structure and survival. This method detects multivariate community patterns linked to biological consequences even when individual taxa have little effect.

Comparing differential abundance approaches like ALDEx2, MaAsLin2, and ANCOM-BC proved crucial. Differing statistical methodologies assume differing data distribution, normalization, and compositional effects, making differential abundance analysis in microbiome investigations difficult. A previous study has demonstrated that technique choice substantially affects whether taxa are relevant. Thus, numerous complementary approaches are increasingly advised to improve robustness and reproducibility.

Only a few genera were consistently recognized across methodologies, emphasizing the need for various analytical methods. Multimethod-identified genera were better survival-associated microbial signature candidates. This multi-method approach reduces false positives and boosts microbiome data biological interpretation confidence.

6.4 Comparison with previous research

The results of this study are in line with what other studies have found: the association between microbiomes and fitness in animals usually involves small changes at the community level rather than large increases in overall diversity. For instance, research on natural vertebrate populations has shown that the composition of the microbial community is more important for host survival than alpha diversity measurements (Leclaire et al., 2023; Worsley et al., 2021). These results support the idea that microbiota affect host fitness by interacting with other organisms in complex ways within microbial communities.

Previous research on mammals and livestock has shown that gut microbiota are critical for immunological function, energy metabolism, and disease resistance (Ding et al., 2025; Tardiolo et al., 2025). However, there is still limited evidence linking bacteria to survival in wildlife populations. This work contributes to this emerging field by being one of the first to examine the composition of microbiota in relation to reindeer calf survival.

This study detected several species in host-associated microbial communities that may be important in mucosal surface microbial ecology using multiple differential abundance methods. *Capnocytophaga*, interacting with host tissues and other mucosal microbes, is common in oral microbiota (Dewhirst et al., 2010). Complex plant-derived sugars are degraded by gut microbial *monoglobus*, suggesting a metabolic network role (Kowalewski & Redinbo, 2025). *Roseisolibacter* and *Parafilimonas*, found in ambient and host-associated microbiomes, have uncertain activities (Clegg & gross, 2024).

The finding that oral microbiota has a greater association with survival than rectal microbiota is also consistent with research in other systems, where microbial communities from different body sites exhibit distinct ecological patterns. Some studies have shown that the bacteria in the mouth can reflect host health and

immune status, which could explain why oral microbial populations in this dataset showed only a weak association with survival (Baker et al., 2024; Ptasiewicz et al., 2022).

Additionally, the partial agreement between the multiple abundance methodologies used in this work highlights common challenges encountered in microbiome research (Mallick et al., 2021; Nearing et al., 2022).

6.5 Methodological strengths and limitations

One of the main strengths of this study is the use of high-resolution amplicon sequence variant (ASV) analysis with the DADA2 workflow. This method provides better taxonomic resolution and reproducibility than traditional OTU-based approaches. By using a range of bioinformatic and statistical tools—including alpha diversity metrics, ordination-based beta diversity analyses, core microbiome analysis, and several differential abundance methods—the study was able to thoroughly evaluate microbiome patterns.

Collecting samples from two different body sites also made the study stronger. Comparing the oral and rectal microbiota gave useful information about how microbial communities differ by site and showed that survival-related signals were mainly found in the oral microbiota.

There are some limitations to keep in mind when looking at these results. First, the moderate sample size might make it harder to detect small biological effects. Second, since microbiota samples were only collected once, the study could not track how microbial communities change over time during early development. Because microbiota can change quickly in early life, collecting samples over a longer period would give a clearer picture of microbiome changes.

In addition, survival was measured as a simple yes-or-no outcome, which might not capture the full complexity of the biological processes that affect calf survival. The study also did not include environmental factors, maternal effects, or genetic differences, all of which could influence both the microbiota and survival.

6.6 Future research

Future research should use larger sample sizes and track how the microbiome changes during early life. This approach could reveal whether microbial community patterns differ between individuals who survive and those who do not survive.

Combining microbiome data with information about host genetics, environment, and physiology could help us better understand how microbiota relate to survival. Using metagenomic and metatranscriptomic methods may also provide deeper insights into how microorganisms and their hosts interact.

Studying microbiota in different body areas and environments could show how they affect the health and survival of wildlife populations.

6.7 Concluding statement

In summary, this work provides an extensive characterisation of the oral and rectal microbiota in reindeer calves and their potential associations with early-life survival. The findings suggest that survival-related microbial signals are nuanced and predominantly identifiable in the oral microbiota rather than the rectal microbiota. These results underscore the significance of site-specific and multivariate methodologies in wildlife microbiome research, enhancing the comprehension of host–microbiome interactions in natural and semi-domesticated animal populations.

6.8 Acknowledgment

I would like to thank my supervisor, Peter Halvarsson at SLU, from the bottom of my heart, for all the help, advice, and useful feedback he gave me during this thesis project. His knowledge of bioinformatics and microbiome research was a big assistance in both figuring out how to analyze the data and how to understand the results scientifically. I am thankful for his patience, helpful feedback, and support while I was writing. These things made this work much better. I appreciate the time he spent going over the paper and giving me helpful feedback that made the study stronger. Working under his direction has been a great way for me to learn and a big step forward in my academic career.

7 References

Åhman, B., Rasmus, S., Risvoll, C., Morten Eilertsen, S., & Norberg, H. (2022). Large predators and their impact on reindeer husbandry. In J. Moen, Ø. Holand, J. Kumpula, & T. Horstkotte, *Reindeer Husbandry and Global Environmental Change* (1st ed., pp. 118–130). Routledge. https://doi.org/10.4324/9781003118565-9

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Baker, J. L., Mark Welch, J. L., Kauffman, K. M., McLean, J. S., & He, X. (2024). The oral microbiome: Diversity, biogeography and human health. *Nature Reviews Microbiology*, *22*(2), 89–104. https://doi.org/10.1038/s41579-023-00963-6

Belkaid, Y., & Hand, T. W. (2014). Role of the Microbiota in Immunity and Inflammation. *Cell*, *157*(1), 121–141. https://doi.org/10.1016/j.cell.2014.03.011

Cabezas, M. P., Fonseca, N. A., & Muñoz-Mérida, A. (2024). MIMt: A curated 16S rRNA reference database with less redundancy and higher accuracy at species-level identification. *Environmental Microbiome*, *19*(1), 88. https://doi.org/10.1186/s40793-024-00634-w

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Clegg, T., & gross, T. (2024). *Cross-feeding Creates Tipping Points in Microbiome Diversity* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2412.05212

Delgado-Baquerizo, M., Maestre, F. T., Reich, P. B., Jeffries, T. C., Gaitan, J. J., Encinar, D., Berdugo, M., Campbell, C. D., & Singh, B. K. (2016). Microbial diversity drives multifunctionality in terrestrial ecosystems. *Nature Communications*, *7*(1), 10541. https://doi.org/10.1038/ncomms10541

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H., Lakshmanan, A., & Wade, W. G. (2010a). The Human Oral Microbiome. *Journal of Bacteriology*, *192*(19), 5002–5017. https://doi.org/10.1128/JB.00542-10

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H., Lakshmanan, A., & Wade, W. G. (2010b). The Human Oral Microbiome. *Journal of Bacteriology*, *192*(19), 5002–5017. https://doi.org/10.1128/JB.00542-10

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H., Lakshmanan, A., & Wade, W. G. (2010c). The Human Oral Microbiome. *Journal of Bacteriology*, *192*(19), 5002–5017. https://doi.org/10.1128/JB.00542-10

Dewhirst, F. E., Chen, T., Izard, J., Paster, B. J., Tanner, A. C. R., Yu, W.-H., Lakshmanan, A., & Wade, W. G. (2010d). The Human Oral Microbiome. *Journal of Bacteriology*, *192*(19), 5002–5017. https://doi.org/10.1128/JB.00542-10

Ding, G., Yang, X., Li, Y., Wang, Y., Du, Y., Wang, M., Ye, R., Wang, J., Zhang, Y., Chen, Y., & Zhang, Y. (2025). Gut microbiota regulates gut homeostasis, mucosal immunity and influences immune-related diseases. *Molecular and Cellular Biochemistry*, *480*(4), 1969–1981. https://doi.org/10.1007/s11010-024-05077-y

Doumazane, E., Scholler, P., Fabre, L., Zwier, J. M., Trinquet, E., Pin, J.-P., & Rondard, P. (2013). Illuminating the activation mechanisms and allosteric properties of metabotropic glutamate receptors. *Proceedings of the National Academy of Sciences*, *110*(15). https://doi.org/10.1073/pnas.1215615110

Du, Y., Gao, Y., Hu, M., Hou, J., Yang, L., Wang, X., Du, W., Liu, J., & Xu, Q. (2023). Colonization and development of the gut microbiome in calves. *Journal of Animal Science and Biotechnology*, *14*(1), 46. https://doi.org/10.1186/s40104-023-00856-x

Edgar, R. (2018). Taxonomy annotation and guide tree errors in 16S rRNA databases. *PeerJ*, *6*, e5030. https://doi.org/10.7717/peerj.5030

Gaillard, J.-M., Festa-Bianchet, M., Yoccoz, N. G., Loison, A., & Toïgo, C. (2000). Temporal Variation in Fitness Components and Population Dynamics of Large Herbivores. *Annual Review of Ecology and Systematics*, *31*(1), 367–393. https://doi.org/10.1146/annurev.ecolsys.31.1.367

Götz, R., Panzer, S., Trinks, N., Eilts, J., Wagener, J., Turrà, D., Di Pietro, A., Sauer, M., & Terpitz, U. (2020). Expansion Microscopy for Cell Biology Analysis in Fungi. *Frontiers in Microbiology*, *11*, 574. https://doi.org/10.3389/fmicb.2020.00574

Hansen, A. K., & Moran, N. A. (2014). The impact of microbial symbionts on host plant utilization by herbivorous insects. *Molecular Ecology*, *23*(6), 1473–1496. https://doi.org/10.1111/mec.12421

Jeon, H.-B., Won, H., Kim, H. J., Kim, S.-H., Chen, Y., Long, J., Tang, S.-L., Bathige, S. D. N. K., & Park, K.-I. (2025). Microbiome of wild oysters: Assessing diversity and detecting potential pathogens using a metabarcoding approach. *BMC Microbiology*, *25*(1), 700. https://doi.org/10.1186/s12866-025-04454-x

Jones, R. B., Zhu, X., Moan, E., Murff, H. J., Ness, R. M., Seidner, D. L., Sun, S., Yu, C., Dai, Q., Fodor, A. A., Azcarate-Peril, M. A., & Shrubsole, M. J. (2018). Inter-niche and inter-individual variation in gut microbial community assessment using stool, rectal swab, and mucosal samples. *Scientific Reports*, *8*(1), 4139. https://doi.org/10.1038/s41598-018-22408-4

Kamenova, S., de Muinck, E. J., Veiberg, V., Utsi, T. A., Steyaert, S. M. J. G., Albon, S. D., Loe, L. E., & Trosvik, P. (2023). Gut microbiome biogeography in reindeer supersedes millennia of ecological and evolutionary separation. *FEMS Microbiology Ecology*, *99*(12), fiad157. https://doi.org/10.1093/femsec/fiad157

Kowalewski, M. E., & Redinbo, M. R. (2025). Emerging gut microbial glycoside hydrolase inhibitors. *RSC Chemical Biology*, *6*(8), 1233–1251. https://doi.org/10.1039/D5CB00050E

Leclaire, S., Pineaux, M., Blanchard, P., White, J., & Hatch, S. A. (2023). Microbiota composition and diversity of multiple body sites vary according to reproductive performance in a seabird. *Molecular Ecology*, *32*(9), 2115–2133. https://doi.org/10.1111/mec.16398

Mallick, H., Rahnavard, A., McIver, L. J., Ma, S., Zhang, Y., Nguyen, L. H., Tickle, T. L., Weingart, G., Ren, B., Schwager, E. H., Chatterjee, S., Thompson, K. N., Wilkinson, J. E., Subramanian, A., Lu, Y., Waldron, L., Paulson, J. N., Franzosa, E. A., Bravo, H. C., & Huttenhower, C. (2021). Multivariable association discovery in population-scale meta-omics studies. *PLOS Computational Biology*, *17*(11), e1009442. https://doi.org/10.1371/journal.pcbi.1009442

Mattei, V., Murugesan, S., Al Hashmi, M., Mathew, R., James, N., Singh, P., Kumar, M., Lakshmanan, A. P., Terranegra, A., Al Khodor, S., & Tomei, S. (2019). Evaluation of Methods for the Extraction of Microbial DNA From Vaginal Swabs Used for Microbiome Studies. *Frontiers in Cellular and Infection Microbiology*, *9*, 197. https://doi.org/10.3389/fcimb.2019.00197

McCoy, C. O., & Matsen, F. A. (2013). *Abundance-weighted phylogenetic diversity measures distinguish microbial community states and are robust to sampling depth* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1305.0306

Nearing, J. T., Douglas, G. M., Comeau, A. M., & Langille, M. G. I. (2022). *Comparative analysis of microbiome differential abundance methods across 18 datasets. Briefings in Bioinformatics*.

Neu, A. T., Allen, E. E., & Roy, K. (2021). Defining and quantifying the core microbiome: Challenges and prospects. *Proceedings of the National Academy of Sciences*, *118*(51), e2104429118. https://doi.org/10.1073/pnas.2104429118

Nguyen, N.-P., Warnow, T., Pop, M., & White, B. (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *Npj Biofilms and Microbiomes*, *2*(1), 16004. https://doi.org/10.1038/npjbiofilms.2016.4

Ptasiewicz, M., Grywalska, E., Mertowska, P., Korona-Głowniak, I., Poniewierska-Baran, A., Niedźwiedzka-Rystwej, P., & Chałas, R. (2022). Armed to the Teeth—The Oral Mucosa Immunity System and Microbiota. *International Journal of Molecular Sciences*, *23*(2), 882. https://doi.org/10.3390/ijms23020882

Ruohan, W., Xianglilan, Z., Jianping, W., & Shuai Cheng, L. I. (2022a). DeepHost: Phage host prediction with convolutional neural network. *Briefings in Bioinformatics*, *23*(1), bbab385. https://doi.org/10.1093/bib/bbab385

Ruohan, W., Xianglilan, Z., Jianping, W., & Shuai Cheng, L. I. (2022b). DeepHost: Phage host prediction with convolutional neural network. *Briefings in Bioinformatics*, *23*(1), bbab385. https://doi.org/10.1093/bib/bbab385

Sanduleanu, S., Masclee, A. M., & Meijer, G. A. (2012). Interval cancers after colonoscopy—Insights and recommendations. *Nature Reviews Gastroenterology & Hepatology*, *9*(9), 550–554. https://doi.org/10.1038/nrgastro.2012.136

Sommer, F., & Bäckhed, F. (2013). The gut microbiota—Masters of host development and physiology. *Nature Reviews Microbiology*, *11*(4), 227–238. https://doi.org/10.1038/nrmicro2974

Tardiolo, G., La Fauci, D., Riggio, V., Daghio, M., Di Salvo, E., Zumbo, A., & Sutera, A. M. (2025). Gut Microbiota of Ruminants and Monogastric Livestock: An Overview. *Animals*, *15*(5), 758. https://doi.org/10.3390/ani15050758

Worsley, S. F., Davies, C. S., Mannarelli, M.-E., Hutchings, M. I., Komdeur, J., Burke, T., Dugdale, H. L., & Richardson, D. S. (2021). Gut microbiome composition, not alpha diversity, is associated with survival in a natural vertebrate population. *Animal Microbiome*, *3*(1), 84. https://doi.org/10.1186/s42523-021-00149-6

Xia, F., Chen, J., Fung, W. K., & Li, H. (2013). A Logistic Normal Multinomial Regression Model for Microbiome Compositional Data Analysis. *Biometrics*, *69*(4), 1053–1063. https://doi.org/10.1111/biom.12079

Zhang, L., Shi, Y., Do, K.-A., Peterson, C. B., & Jenq, R. R. (2020). *ProgPermute: Progressive permutation for a dynamic representation of the robustness of microbiome discoveries* (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2005.01169

Zhang, L., Zhang, X., & Yi, N. (2024). Bayesian compositional generalized linear models for analyzing microbiome data. *Statistics in Medicine*, *43*(1), 141–155. https://doi.org/10.1002/sim.9946

Zhang, Y., Liang, S., Choi, S., & Li, G. (2025). Effects of maternal rumen microbiota on the development of the microbial communities in the gastrointestinal tracts of neonatal sika deer. *Journal of Animal Science and Technology*, *67*(3), 619–635. https://doi.org/10.5187/jast.2024.e37

Zhou, H., He, K., Chen, J., & Zhang, X. (2022a). LinDA: Linear models for differential abundance analysis of microbiome compositional data. *Genome Biology*, *23*(1), 95. https://doi.org/10.1186/s13059-022-02655-5

Zhou, H., He, K., Chen, J., & Zhang, X. (2022b). LinDA: Linear models for differential abundance analysis of microbiome compositional data. *Genome Biology*, *23*(1), 95. https://doi.org/10.1186/s13059-022-02655-5

Zhou, H., He, K., Chen, J., & Zhang, X. (2022c). LinDA: Linear models for differential abundance analysis of microbiome compositional data. *Genome Biology*, *23*(1), 95. https://doi.org/10.1186/s13059-022-02655-5

Popular science summary

In the Arctic and sub-Arctic, reindeer are vitally essential to both the culture and the economy. But many baby reindeer die in the first few months of life. To take care of reindeer herds and learn about wildlife ecology, you need to know what makes calves live longer

There are a number of distinct kinds of bacteria in all mammals. Microbes inhabit the body, including the oral cavity and gastrointestinal tract. They help with digestion, the growth of the immune system, and keeping people from getting sick. There has been less research on the impact of microbial ecosystems on the health and survival of reindeer, despite extensive studies conducted on humans and pets.

I examined the microbiome of reindeer calves to determine the impact of microbial communities on early-life survival. Samples were taken from the mouth and the rectum. DNA sequencing revealed bacteria in every sample, while bioinformatic analysis examined the variety and community composition of the microbiomes in the calves.

The mouth and rectum each contain their own ecosystems of bacteria, which illustrates that different areas of the body have different sorts of germs. The microbial diversity was similar in both male and female calves and in both survivors and those who did not survive.

The oral microbiota showed weak relationships with survival, whereas the rectal microbiome did not. These findings indicate that minor alterations in microbial ecosystems may affect calf survival; nevertheless, more extensive datasets and more sampling are required.

This study improves our understanding of the effects of microbial communities on the health and survival of reindeer.

Appendix

Software, packages, databases, and key analysis parameters

All bioinformatic and statistical analyses were performed in the **R statistical environment**. The microbiome workflow included sequence processing, taxonomic assignment, construction of a phyloseq object, diversity analyses, core microbiome analysis, generalized linear modelling, and differential abundance analyses.

Software and packages

The following software, databases, and R packages were used in the analysis:

- **R** (version 4.5) – statistical computing environment
- **RStudio** (version 4.5) – integrated development environment for R
- **DADA2**(version 1.38) – quality filtering, denoising, ASV inference, merging of paired-end reads, chimera removal
- **phyloseq** (version 1.5)– integration of ASV table, taxonomy, and metadata; diversity analysis and visualization
- **microbiome** (version 1.32)– core microbiome analysis
- **vegan** (version 2.7)– Bray–Curtis dissimilarity, ordination support, and PERMANOVA
- **ggplot2** (version 4.0)– visualization of diversity and taxonomic composition
- **ALDEx2** (version 1.4)– differential abundance analysis
- **MaAsLin2** (version 1.1)– multivariable differential abundance analysis
- **ANCOMBC** (version 2.1)– bias-corrected differential abundance analysis
- **SILVA reference database** (version 1.38)– taxonomic assignment of ASVs

Sequence processing and ASV inference (DADA2)

Raw paired-end FASTQ files were processed using the **DADA2** pipeline. Forward and reverse reads were filtered and trimmed using the following parameters:

```
filterAndTrim(
  fnFs, filtFs,
  fnRs, filtRs,
  truncLen = c(240, 200),
  maxN = 0,
  maxEE = c(2, 2),
  truncQ = 2,
  rm.phix = TRUE,
  compress = TRUE,
  multithread = TRUE
)
```

The main settings used in the quality filtering step were:

- `truncLen = c(240, 200)` to truncate forward and reverse reads at 240 bp and 200 bp, respectively
- `maxN = 0` to discard reads containing ambiguous bases
- `maxEE = c(2, 2)` to limit expected errors in forward and reverse reads
- `truncQ = 2` to truncate reads at low-quality positions
- `rm.phix = TRUE` to remove PhiX contamination
- `multithread = TRUE` to allow parallel processing

Sequencing error rates were learned using:

```
learnErrors(filtFs, multithread = TRUE)
learnErrors(filtRs, multithread = TRUE)
```

ASVs were inferred from dereplicated reads using:

```
dada(derepFs[[i]], err = errF, multithread = FALSE)
dada(derepRs[[i]], err = errR, multithread = FALSE)
```

Paired-end reads were merged using:

```
mergePairs(
  dadaFs, derepFs,
  dadaRs, derepRs,
  verbose = TRUE
)
```

Chimeric sequences were removed using the consensus method:

```
removeBimeraDenovo(
  seqtab,
  method = "consensus",
  multithread = TRUE
)
```

The resulting chimera-filtered ASV table was used for all downstream analyses.

Taxonomic assignment (SILVA)

Taxonomic assignment was performed after ASV inference using the **SILVA reference database**, a curated ribosomal RNA database commonly used for microbial taxonomic classification. SILVA was selected because it provides validated rRNA sequences and consistent taxonomic annotation, improving the reliability and reproducibility of microbiome analyses.

Taxonomy was imported into the workflow from a saved taxonomy object:

```
taxa <- readRDS("taxonomy.rds")
taxa <- as.matrix(taxa)
```

Taxonomic information was used at several hierarchical levels, particularly the **phylum** and **genus** levels, for downstream ecological and statistical analyses.

Construction of the phyloseq object

The final ASV table, taxonomy matrix, and sample metadata were integrated into a **phyloseq** object:

```
ps <- phyloseq(
  otu_table(seqtab_final, taxa_are_rows = FALSE),
  tax_table(taxa),
  sample_data(metadata_final)
)
```

The **phyloseq** package was used to organize microbiome count data, taxonomy, and metadata in a single object suitable for ecological and statistical analyses.

Taxonomic agglomeration was performed using:

```
tax_glom(ps, taxrank = "Phylum")
tax_glom(ps, taxrank = "Genus")
```

Relative abundance transformation was performed using:

```
transform_sample_counts(ps_phylum, function(x) x / sum(x))
```

This allowed visualization and comparison of microbial composition across samples and groups.

Alpha diversity analysis

Alpha diversity was estimated using **Observed ASVs** and the **Shannon diversity index**:

```
estimate_richness(ps, measures = c("Observed", "Shannon"))
```

The Shannon index was selected because it captures both richness and evenness of microbial communities. Alpha diversity values were combined with metadata for downstream analyses and graphical visualization using boxplots and jittered sample points.

Beta diversity analysis and PERMANOVA

Beta diversity was calculated using **Bray–Curtis dissimilarity**, which is based on abundance data and is widely used in microbial ecology to compare differences in community composition between samples:

```
dist_bc <- phyloseq::distance(ps_clean, method = "bray")
```

Microbial community differences were tested using **PERMANOVA** with the `adonis2` function from the **vegan** package. The main PERMANOVA model included sampling site, sex, and survival status:

```
adonis2(
  dist_bc ~ Site + Sex + Survival,
  data = data.frame(sample_data(ps)),
  permutations = 5000
)
```

An additional PERMANOVA model was run with:

```
adonis2(
  bray_dist_clean ~ Site + Sex,
  data = meta_df_clean,
  permutations = 999
)
```

The main parameter settings used in PERMANOVA were:

- **Bray–Curtis dissimilarity** as the distance measure
- **5000 permutations** in the main model
- **999 permutations** in additional exploratory models

PERMANOVA was used to test whether microbial community composition differed significantly between predefined groups in multivariate space.

Generalized linear models (GLM)

Generalized linear models (**GLMs**) with a **binomial distribution** were used to test whether microbial diversity or microbial community structure was associated with survival status.

A GLM based on alpha diversity was fitted using:

```
glm(
  Survival ~ Shannon + Site + Sex,
  data = alpha_df,
  family = binomial
)
```

A GLM based on ordination axes was fitted using:

```
glm(
  Survival ~ Axis.1 + Axis.2 + Site,
  data = alpha_df,
  family = binomial
)
```

The key model settings were:

- `family = binomial` because survival was treated as a binary response variable
- Shannon diversity, site, sex, or ordination axes as explanatory variables, depending on the model

These GLMs were used to assess whether microbial diversity metrics or community-level compositional variation could predict survival outcomes.

Core microbiome analysis (microbiome package)

Core microbiome analysis was performed using the **microbiome** package. This analysis was used to identify taxa consistently present across samples within a defined group.

The core microbiome was defined using prevalence and detection thresholds, following the criteria described in the main Methods section:

- **Prevalence threshold:** at least 70% of samples
- **Detection threshold:** at least 0.01% relative abundance

A general implementation in the microbiome package follows the form:

```
core(x, detection = 0.0001, prevalence = 0.70)
```

Core microbiome analysis was used to identify stable taxa within oral and rectal microbial communities and to compare persistent taxa between survival groups.

Differential abundance analysis

Differential abundance analyses were performed at the **genus level** using three complementary methods: **ALDEx2**, **MaAsLin2**, and **ANCOM-BC**.

ALDEx2

ALDEx2 analysis was performed using centered log-ratio transformation with Monte Carlo sampling:

```
aldex.clr(
  genus_counts_anus,
  conds_anus,
  mc.samples = 128,
  denom = "all"
)
```

Statistical testing and effect size estimation were performed using:

```
aldex.ttest(aldex_anus)
aldex.effect(aldex_anus)
```

The important ALDEx2 parameters were:

- `mc.samples = 128`
- `denom = "all"`

Significant taxa were identified using an adjusted p-value threshold of:

```
wi.eBH < 0.05
```

MaAsLin2

MaAsLin2 was run using total sum scaling normalization and log transformation:

```
Maaslin2(
  input_data = feature_table_anus,
  input_metadata = meta_anus,
  output = "maaslin_anus_output",
  fixed_effects = c("Survival"),
  normalization = "TSS",
  transform = "LOG"
)
```

The key parameters were:

- `fixed_effects = c("Survival")`
- `normalization = "TSS"`
- `transform = "LOG"`

Significance was evaluated using:

```
qval < 0.05
```

ANCOM-BC

ANCOM-BC2 was performed using:

```
ancombc2(
  data = ps_anus_genus,
  fix_formula = "Survival",
  p_adj_method = "BH"
)
```

The key parameters were:

- `fix_formula = "Survival"`
- `p_adj_method = "BH"`

This method was used to identify taxa associated with survival while accounting for compositional bias in microbiome data.

Visualization

Figures were produced using **ggplot2** and **phyloseq**. Taxonomic bar plots were generated after agglomerating taxa at the phylum level and converting counts to relative abundance:

```
plot_bar(ps_phylum_rel, fill = "Phylum") +
  facet_wrap(Site ~ Sex, scales = "free_x")
```

Alpha diversity plots were visualized using boxplots with overlaid individual sample points, and differential abundance results were summarized using scatter plots and bar plots.

Reproducibility notes

The appendix presents the final analytical workflow and key parameters used in the study. Intermediate troubleshooting steps, repeated exploratory code, and temporary objects used during data cleaning were not included here, as the purpose of this section is to provide a concise and reproducible summary of the final bioinformatic and statistical pipeline.

Publishing and archiving

Approved students' theses at SLU can be published online. As a student, you own the copyright to your work, and in such cases, you need to approve the publication. In connection with your approval of publication, SLU will process your personal data (name) to make the work searchable on the internet. You can revoke your consent at any time by contacting the library.

Even if you choose not to publish the work or if you revoke your approval, the thesis will be archived digitally according to archive legislation.

You will find links to SLU's publication agreement and SLU's processing of personal data and your rights on this page:

- https://libanswers.slu.se/en/faq/228318

☒ YES, I, Mearg Belay Brhane, have read and agree to the agreement for publication and the personal data processing that takes place in connection with this

☐ NO, I/we do not give my/our permission to publish the full text of this work. However, the work will be uploaded for archiving, and the metadata and summary will be visible and searchable.