Evaluating Microbial Community Changes in Rumen Samples Using Multiple Distance Metrics

Sachin Chandrasekara[1]

***Abstract*— This study revealed how different treatments affect the microbial communities in the rumen of cattle. Specifically, 16S rRNA sequencing and several analysis methods were used to examine the differences between treatments and their respective time periods. To validate the results, statistical technologies such as PCoA (Principal Coordinates Analysis), PCA (Principal Component Analysis), UniFrac, and PERMANOVA (Permutational Multivariate Analysis of Variance) were employed. However, we found some variations within individual microbes, suggesting the need for further study using functional or gene-based approaches.**

I. INTRODUCTION

The types and activities of microbes in the rumen can vary significantly depending on the animal's diet. Different diets can impact the abundance and diversity of microbes, as well as their functions. These changes influence not only digestion and nutrient utilization but also the amount of methane (a greenhouse gas) produced by the animal [1]. The rumen is a unique stomach compartment in ruminant animals such as cattle, consisting of a complex and diverse community of microorganisms that include bacteria, archaea, protozoa, and fungi [2]. These microbes help break down fibrous plants into tiny molecules called volatile fatty acids. These acids provide the animal with most of the energy it needs [3]. The types and activities of microbes in the rumen could change a lot depending on what the animal eats. Different diets can impcat how many microbes are there, how varied they are, and what they do. These shifts influne not only digestion and practical utilization, but also how much methane (a greenhouse gas) they produce.

In this study, effect of five dietary treatments called as Control, Tallow, Corn Oil, Tl, and COL on the were evaluted on the structure of the rumen microbbiomee across five different time periods. The goal was to whether dietary lipid supplementation, alone or in combination changed the types of microbes in the rumen over time. The 16S rRNA gene-based high thoughput sequncing was applied to profile bacterial communites. And also, which was used a combiantion of cordination methods (PCA and PCoA) and statistical test to assess treatment and time effcect in microboial community.

II. METHODOLOGY

A. Dataset

Rumen microbial samples were collected from five dairy cows over a 12-week. Rumen microbial samples were collected from five dairy cows over a 12-week period. Each cow was assigned to one of five dietary treatments: Control, Corn Oil, Tallow, COL, or TL. A total of 25 experimental samples were collected, with each cow providing one sample per treatment period. More information about the data set is depicted in Table I. In addition to the experimental samples, the plate also included the following controls, In addition to the experimental samples, the plate also included the following controls Extraction Negative Control, PCR Negative Control and PCR Positive Control.

TABLE I
ANIMAL DISTRIBUTION ACROSS TREATMENT GROUPS AND TIME POINTS

Treatment	Animals Involved	Periods
Control	7075, 7063, 7076, 7101, 7084	Periods 1–5
Corn Oil	7075, 7063, 7076, 7101, 7084	Periods 1–5
Tallow	7075, 7101, 7063, 7076, 7084	Periods 1–5
COL	7075, 7063, 7076, 7101, 7084	Periods 1–5
TL	7075, 7063, 7076, 7101, 7084	Periods 1–5

A total of 28 samples were used for downstream analysis, which included 25 true samples and 3 control samples. The sample IDs were matched with metadata that included animal ID, collection date, treatment group, time period, and sample type. The treatment groups are detailed in Table II.

TABLE II
DESCRIPTION OF DIETARY TREATMENT GROUPS

Treatment	Description
Control	No added fat
Corn Oil	3.5% corn oil
Tallow	3.5% tallow
COL	3.5% corn oil + lysoforte (a digestive aid)
TL	3.5% tallow + lysoforte (a digestive aid)

B. Ordination Analysis Using PCoA and PCA

Two ordination methods were used to explore patterns in microbial community structure across different diets and time points: Principal Component Analysis (PCA) and Principal Coordinates Analysis (PCoA).

PCoA was used to explore differences in microbial communities across samples. UniFrac distances, which consider both the phylogenetic relationships between microbes and their abundance or presence/absence in the samples, were used to perform PCoA [4].

$$D' = -\frac{1}{2} H D^2 H \tag{1}$$

Where D' is the transformed distance matrix and H is a centering matrix used to remove the mean. PCA was

Algorithm 1 Phylogenetic Tree Construction

1: **Input:** DNA sequence set $S = \{s_1, s_2, \ldots, s_n\}$
2: Align sequences: $A = \text{Align}(S)$
3: Convert alignment: $D_{\text{phy}} = \text{Convert}(A)$
4: Compute distance matrix: $D = [d(s_i, s_j)]$
5: Construct NJ tree: $T_0 = \text{NJ}(D)$
6: Define GTR model: $P(t) = \exp(Q \cdot t)$
7: Optimize tree:

$$T^*, \Theta^* = \arg\max_{T,\Theta} \log \mathcal{L}(T, \Theta \mid A)$$

8: **Output:** Optimized tree T^*

applied to normalized microbial abundance data, specifically relative abundance. It identifies new axes known as principal components (PCs) that reveal the greatest variation in the data [5].

$$Z = XW \tag{2}$$

Where, Z is the matrix of principal component scores, X is the input data matrix and W is the matrix of eigenvectors.

C. UniFrac Distances

UniFrac is a distance measure that indicates how much of the microbial DNA in each sample comes from different branches of the phylogenetic tree. This not only shows which microbes are present but also provides insight into how closely related they are [6], [7]. Two types of Unifrac methods were used. Equation 3 calculates the distance between two samples.

$$d_{\text{UniFrac}}(A, B) = \frac{\sum w_e \cdot |p_{e,A} - p_{e,B}|}{\sum w_e} \tag{3}$$

Where: w_e is the length of each branch in the phylogenetic tree, $p_{e,A}, p_{e,B}$ are the amounts of microbes in A and B that come from that branch.

- Unweighted UniFrac shows only at whether or not a microbe is present. It is good at detecting rare microbes.
- Weighted UniFrac considers how abundant each microbe is. This gives more importance to the dominant microbes.

D. Construct Phylogenetic Tree

The construction of a phylogenetic tree is described by the following Algorithm 1. This algorithm has 8 steps Input DNA sequences, Align sequences, Convert alignment to phyDat format, Calculate pairwise distances, Construct initial NJ tree, Define substitution model (GTR), Optimize tree using ML and Output final phylogenetic tree. R packages were used for all studies, including DECIPHER for sequence alignment and Phangan for distance and tree optimisation [8], [9].

III. Results

Figure 1 shows rarefaction curves for microbial samples, which helps to assess the depth of the sequencing and the richness of the species. Each curve represents a sample, and the number of observed species increases with more sequencing reads until it level. The red dashed line indicates the minimum sequencing depth across all samples.

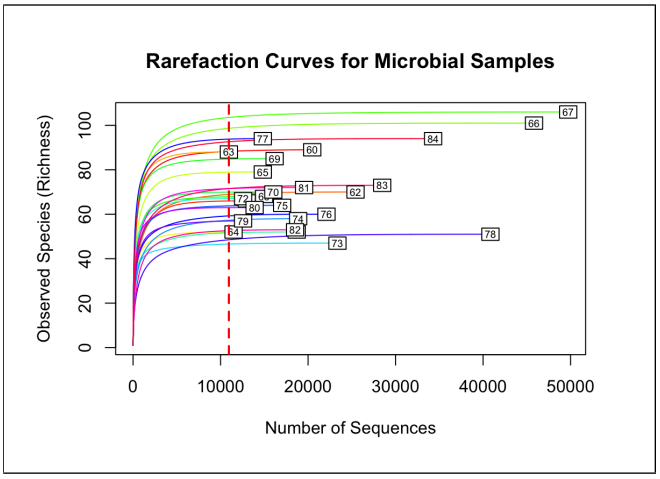

Fig. 1. Rarefaction curves for microbial samples.

This Figure 2 depicts the alpha diversity of the microbial communities in five treatment groups COL, Control, Corn Oil, Tallow, and TL. To obtain these results, we filtered ASVs with prevalence > 1 and only kept those ASVs where total abundance is > 100.

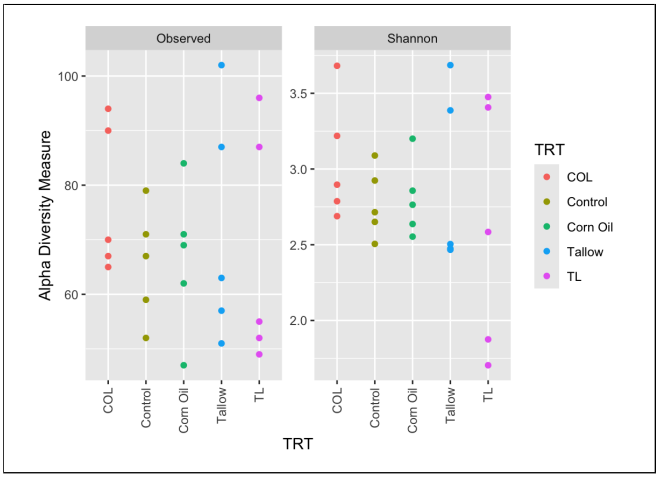

Fig. 2. Alpha diversity of microbial communities.

The observed panel shows how many different types of microbes were found in each sample. The Shannon panel displays for both the number of types and how evenly they are distributed. Tallow and TL have more variation and some of the highest observed diversity. COL and Control samples tend to have lower diversity in both measures.

A. PCoA using Bray-Curtis

Figure 3 was performed PCoA using Bray-Curtis dissimilarity and two axes explaining 21. 7% and 12. 0% of the variation, respectively. The samples showed some clustering based on treatment and sampling period, although no clear separation was observed.

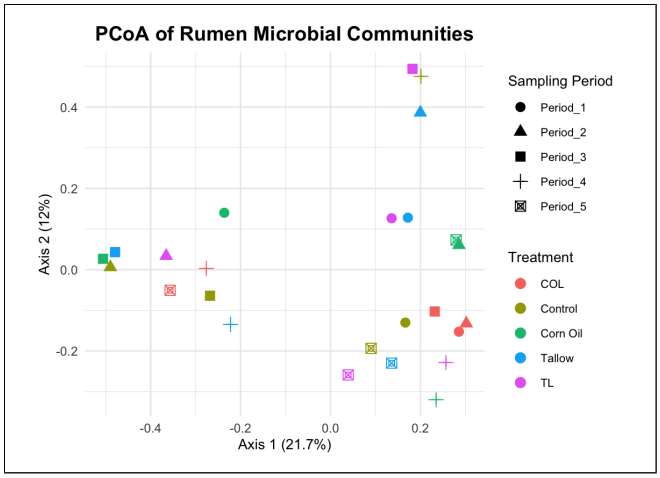

Fig. 3. PCoA using Bray-Curtis dissimilarity.

As depicted in Table III PERMANOVA (adonis2) was tested to check microbial communities differed significantly between treatment groups or across time. The treatment effect (TRT) explained 12.9% of the variation (R^2 = 0.129, p = 0.967), while the sampling period explained 16.4% (R^2 = 0.164, p = 0.605). However, neither factor was statistically significant. Most of the variation (70.7%) remained unexplained by these two factors.

TABLE III
PERMANOVA RESULTS FOR TREATMENT AND PERIOD EFFECTS ON MICROBIAL COMMUNITY COMPOSITION USING BRAY-CURTIS

Factor	Df	SumOfSqs	R^2	F	Pr(>F)
TRT	4	1.1713	0.12917	0.7310	0.967
Period	4	1.4868	0.16396	0.9278	0.605
Residual	16	6.4097	0.70686		
Total	24	9.0678	1.00000		

B. Phylogenetic Relationships of Rumen Microbial ASVs

A phylogenetic tree of amplicon sequence variants (ASVs) was constructed to visualize the evolutionary relationships among the microbial taxa identified in the rumen. This Figure 4 was built using multiple sequence alignment followed by maximum likelihood optimization under the General Time Reversible (GTR) model with gamma rate variation and invariant sites. The tips of the tree were colored on the basis of the phylum-level taxonomy of each ASV. The main phyla observed included *Bacteroidota, Bacillota, Patescibacteria, Actinomycetota, Spirochaetota, Cyanobacteriota* and *Pseudomonadota*.

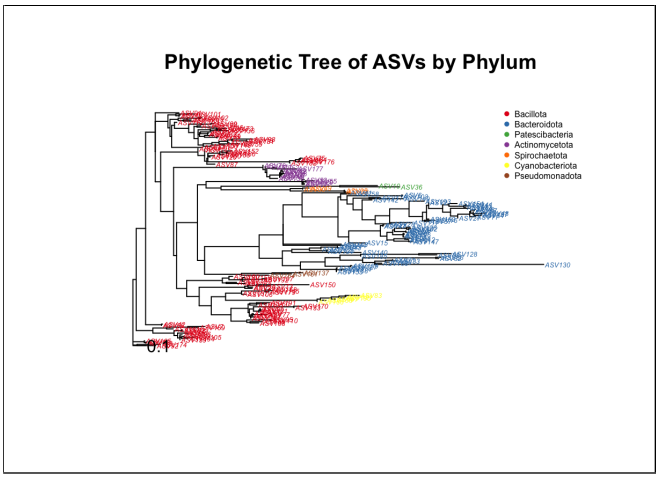

Fig. 4. Phylogenetic Relationships of Rumen Microbial ASVs.

C. Unweighted UniFrac

Figure 5 explains that PCoA-based unweighted UniFrac distances revealed clustering patterns in microbial communities according to treatment and sampling periods. The axes explained 63.0% of the total variation (45.2% by Axis 1 and 17.8% by Axis 2). Some samples seem to group together. This means that treatment and sampling time might affect the microbes present, but the results also show a lot of overlap between groups. This overlap suggests that individual differences or small changes in microbes could be responsible.

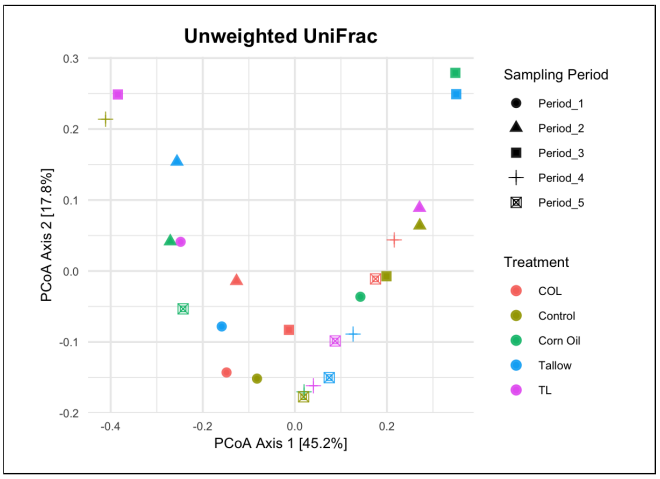

Fig. 5. PCoA using Unweighted UniFrac.

However, these trends were not statistically significant according to PERMANOVA analysis in Table IV. Treatment represented 12. 2% of the variation (p = 0.825), and the period explained 16. 9% (p = 0.520). The high residual variation (70.9%) indicates that other unmeasured factors or individual variation may be contributing more strongly to the observed differences in microbial composition.

TABLE IV

PERMANOVA RESULTS SHOWING THE EFFECT OF TREATMENT (TRT) AND SAMPLING PERIOD USING UNWEIGHTED UNIFRAC

Factor	Df	SumOfSqs	R^2	F	Pr(>F)
TRT	4	0.32949	0.12181	0.6875	0.825
Period	4	0.45837	0.16946	0.9564	0.520
Residual	16	1.91705	0.70873		
Total	24	2.70490	1.00000		

D. Weighted UniFrac

The Weighted UniFrac PCoA Figure 6 shows the differences in microbial communities between treatments and time points. The two axes explain 62.8% of the variation (42.5% on Axis 1 and 20.3% on Axis 2). Some samples from similar treatments or periods appear close to each other in the plot, but the grouping is not very strong or clear. The PERMANOVA results in Table V shows that neither treatment (p = 0.984) nor the sampling period (p = 0.932) had a statistically significant effect on the composition of the microbial community. This suggests that, although there may be minor differences, the overall microbial communities were not substantially influenced by treatment or sampling time, according to this distance measure.

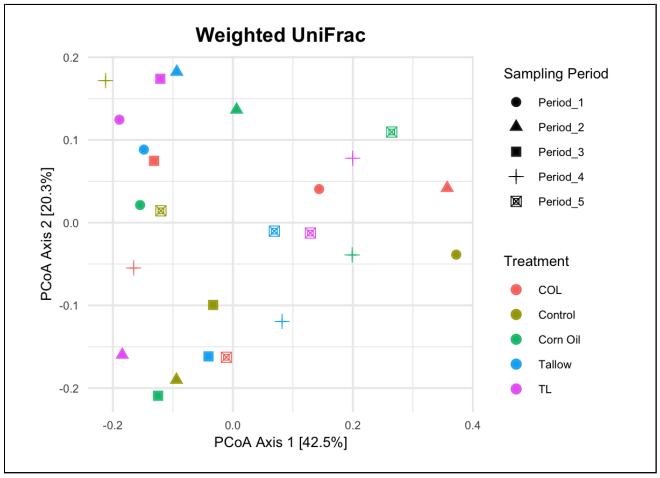

Fig. 6. PCoA using using Weighted UniFrac.

TABLE V

PERMANOVA RESULTS SHOWING THE EFFECT OF TREATMENT (TRT) AND SAMPLING PERIOD USING WEIGHTED UNIFRAC.

Source	Df	Sum of Squares	R^2	F-value	Pr(>F)
TRT	4	0.14198	0.09038	0.4530	0.984
Period	4	0.17532	0.11160	0.5594	0.932
Residual	16	1.25363	0.79802		
Total	24	1.57093	1.00000		

E. PCA on Rumen Microbial Communities

In the Figure 7 PC1 and PC2 explained 14% and 9.3% of the total variation, respectively. Although some samples from the same treatment or period appear close together, but most groups are spread across the plot, with no clear separation between treatments or time points.

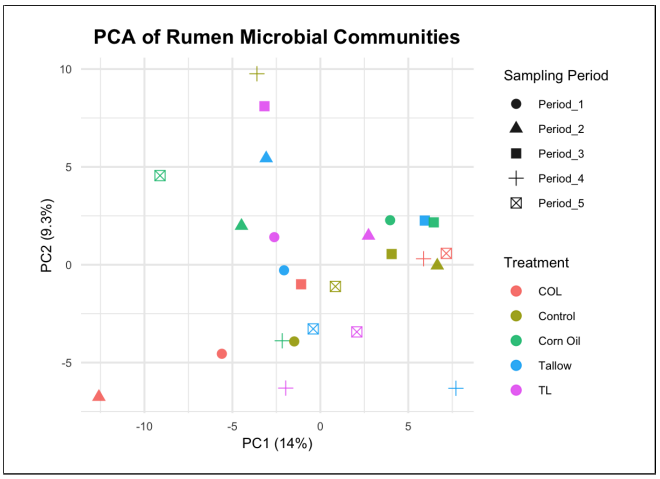

Fig. 7. PCA on Rumen Microbial Communities.

The statistical test showed in Table VI that neither treatment (p = 0.919) nor sampling period (p = 0.667) had a significant effect on the microbial community. The residual variation (69.6%) was much higher than the variation explained by treatment or time, suggesting that individual differences or other unmeasured factors contributed more to the overall variation in microbial composition.

TABLE VI

PERMANOVA RESULTS USING EUCLIDEAN DISTANCE ON PCA INPUT

Source	Df	Sum of Squares	R^2	F-value	Pr(>F)
TRT	4	0.37940	0.14283	0.8205	0.919
Period	4	0.42733	0.16087	0.9241	0.667
Residual	16	1.84963	0.69630		
Total	24	2.65635	1.00000		

F. Aitchison PCA (CLR-transformed)

In the Figure 8 PC1 and PC2 explain 24.1% and 15.2% of the variation. Describes a significant part of the differences between samples. Some treatments or sampling periods appear to form small clusters, especially on the right side of the plot, but there is still a lot of overlap between the groups.

The PERMANOVA results of Table VII show that neither treatment (p = 0.965) nor the sampling period (p = 0.540) had a statistically significant impact on the microbial composition. Most of the variation (71.2%) was due to individual differences rather than to treatment or time.

TABLE VII

PERMANOVA RESULTS BASED ON SUING ATICHISON PCA

Source	Df	Sum of Squares	R^2	F-value	Pr(>F)
TRT	4	3205.4	0.12019	0.6751	0.965
Period	4	4469.9	0.16761	0.9414	0.540
Residual	16	18993.4	0.71220		
Total	24	26668.7	1.00000		

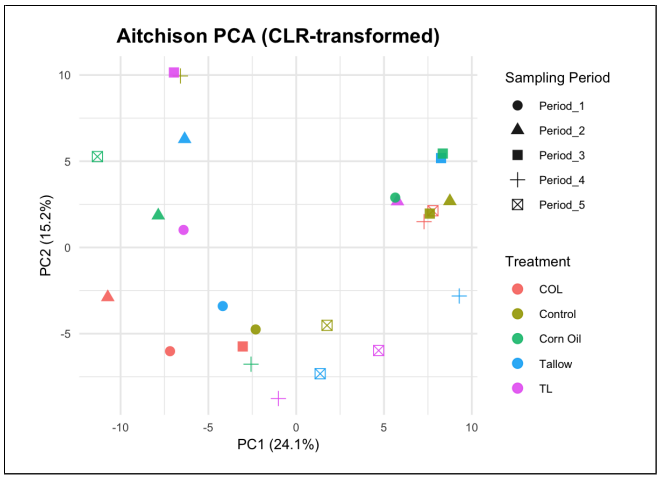

Fig. 8. PCA on Rumen Microbial Communities.

IV. Conclusion

PERMANOVA was used to check whether different treatments or sampling times changed the types of microbes found. This was tested using five different ways: Aitchison PCA, Euclidean distance, weighted UniFrac, unweighted UniFrac, and Bray-Curtis. In all tests, we coould not find any strong or significant differences between treatment groups or time points (all p-values were greater than 0.5). This suggests that treatments and sampling periods did not have a large effect on the overall makeup of the microbial communities.

Data and Code Availability

All the following codes used for data preprocessing, data analysis, and visualization in this study is publicly available in the following GitHub repository: https://github.com/sachinkavindaa/KobzaMicrobial

References

[1] Morgavi, D. P., Forano, E., Martin, C., Newbold, C. J. (2013). Microbial ecosystem and methanogenesis in ruminants: a review. Animal, 7(s1), 102-112.

[2] Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., Janssen, P. H., Global Rumen Census Collaborators. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. Scientific Reports, 5, 14567.

[3] Hook, S. E., Wright, A. D. G., McBride, B. W. (2010). Methanogens: methane producers of the rumen and mitigation strategies. Archaea, 2010, 945785.

[4] Lozupone, C., Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology, 71(12), 8228–8235. https://doi.org/10.1128/AEM.71.12.8228-8235.2005.

[5] Jolliffe, I. T. (2002). Principal Component Analysis (2nd ed.). Springer Series in Statistics. Springer. https://doi.org/10.1007/b98835.

[6] C. Lozupone and R. Knight, "UniFrac: a new phylogenetic method for comparing microbial communities," Applied and Environmental Microbiology, vol. 71, no. 12, pp. 8228–8235, 2005.

[7] C. Lozupone et al., "Quantitative and qualitative diversity measures lead to different insights into factors that structure microbial communities," Applied and Environmental Microbiology, vol. 73, no. 5, pp. 1576–1585, 2007.

[8] E. S. Wright, "DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment," BMC Bioinformatics, vol. 17, no. 1, pp. 1–14, 2016.

[9] K. P. Schliep, "phangorn: phylogenetic analysis in R," Bioinformatics, vol. 27, no. 4, pp. 592–593, 2011.