**The Signature Microbiota Driving Rumen Function Shifts in Goat Kids Introduced Solid Diet Regimes**

Xiaokang Lv1\*, Jianmin Chai12\*, Qiyu Diao1, Wenqin Huang1, Yimin Zhuang1, Naifeng Zhang1**#**

1Feed Research Institute, Chinese Academy of Agricultural Sciences, Key Laboratory of Feed Biotechnology of the Ministry of Agriculture, Beijing 100081, China

2Department of Animal Science, Division of Agriculture, University of Arkansas, Fayetteville, AR, 72701, USA

Running title: signature associated with diet and rumen function

\*these authors contributed equally to this work.

**#Correspondence:**

Professor Naifeng Zhang, Email [zhangnaifeng@caas.cn](mailto:zhangnaifeng@caas.cn)

Postal address: No.12 Zhongguancun South St., Haidian District, Beijing, P.R.China 100081

Phone #: +86(10) 8210 6055

Fax #: +86(10) 6216 9105

**Detailed Materials and Methods**

***Goat kids, treatments and management***

The experimental procedure was approved by the Chinese Academy of Agricultural Sciences Animal Ethics Committee, and humane animal care and handling procedures were followed throughout the experiment. This animal trial was conducted using Haimen goat kids at a commercial farm in the Jiangsu province, China.

A total of 72 Haimen goat kids (20 days old and average body weight 4.54± 0.51kg) were separated from their dams, and randomly allotted to three groups based on their following diets: milk replacer only (**MRO**), milk replacer + concentrate (**MRC**), milk replacer + concentrate + alfalfa pellets (**MCA**). Each group had six replicates and four kids per pen were as a replicate.

Goat kids remained with their mother and received breast milk from 0 to 20 days. During 20 to 60 days of age, they were separated with their dams and the above 3 kinds of diets were provided to corresponding groups. Other feeding management including vaccination, cleaning and disinfection of pens followed farm normal policy. All animals were fed with milk replacer from 20 to 60 days. Feeding amount of milk replacer were 2% body weight. Goat kids were fed 4 times a day (0600, 1200, 1800 and 2200) at 20-30 days, and thrice daily at 30-60 days (0600, 1200 and 1800). The milk replacer was dissolved with hot water cooled to 65-70 °C after boiling, and offered to goat kids when it was cooled to 40 ± 1 °C. The ratio of milk replacer to water was 1:6 (weight (g)/ volume (ml)). The milk replacer (China patent products ZL02128844.5) used in the experiment was provided by Beijing Precision Animal Nutrition Research Center. The concentrate with ingredients of corn, soybean etc. was purchased from Cargill Feed company, Nanjing. The alfalfa pellets purchased from Baofa Agriculture and Animal Husbandry Co. Ltd, Gansu, China had same diameter (4 mm) with concentrate diet. During the animal trial, all the goat kids had *ad libitum* access to water, the MRC and MCA kids were freely to access concentrate, and the MCA goats had extra free choice of alfalfa pellets. The nutritional levels of milk replacer, concentrate and alfalfa pellets are shown in Table S1.

***Sample collection and Chemical analysis***

Daily feed intakes were recorded in animal trial. Feed samples were collected, dried in a forced-air oven at 65°C for 48 h and analyzed for crude protein (CP), non-fiber carbohydrate (NFC), and neutral detergent fiber (NDF) according to the Association of Official Analytical Chemists (1). Then, average daily intake of CP, NFC and NDF were calculated. Only data of table S1 (dietary composition) and table S2 (growth performance) were published in a Chinese journal paper (2), and other data, such as rumen fermentation parameters and microbiome analysis, were not published and used for the current draft.

Six goat kids (healthy and BW close to the average BW of the corresponding groups) from each group were selected and slaughtered for rumen samples collection. At 60 days of age, the goat kids were taken to an on-farm slaughterhouse, anesthetized using sodium pentobarbitone, and slaughtered by exsanguination from the jugular vein. Then, the rumen organs were taken out, and the ruminal content pH was measured immediately using pH electrode (PB-10; Sartorius, Goettingen, Germany). Around 10 ml rumen content were sampled from the whole mixed rumen digesta and stored at -80°C for sequencing. The rumen fluid around 10 ml filtered through four layers of gauze was placed in a 15 ml centrifuge tube immediately frozen at -20°C for measurement of rumen fermentation. Determination of rumen fluid NH3-N concentration by phenol-sodium hypochlorite colorimetric method after the liquid was thawed at 4°C. The microbial proteins were analyzed according to the method described by (3). The volatile fatty acids (VFA) in rumen fluid were quantified by gas chromatography (4) using methyl valerate as internal standard in an Agilent 6890 series GC equipped with a capillary column (HP-FFAP19095F-123, 30 m, 0.53 mm diameter and 1 mm thickness).

***DNA extraction and 16S rRNA gene sequencing***

Rumen fluid samples were thawed on ice and microbial DNA was extracted using a commercial DNA Kit (Omega Bio-tek, Norcross, GA, U.S.) according to manufacturer’s instructions. Total DNA quality analysis using a Thermo NanoDrop 2000 UV spectrophotometer and 1% agarose gel electrophoresis. The V3-V4 region of the bacteria 16S ribosomal RNA genes were amplified by PCR (95 °C for 3 min, followed by 30 cycles at 98°C for 20 s, 58 °C for 15s, and 72 °C for 20 s and a final extension at 72 °C for 5 min) using indexes and adaptors-linked universal primers (431 F:ACTCCTACGGGRSGCAGCAG; 806R: GGACTACVVGGGTATCTAATC). PCR reactions were performed in 30 μL mixture containing 15 μL of 2 × KAPA Library Amplification Ready Mix, 1 μL of each primer (10 μM), 50ng of template DNA and ddH2O. All PCR products were normalized and quantified by a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, US). Amplicon libraries were mixed using all qualified products and sequenced with Illumina HiSeq PE250 plateform at Realbio Technology Genomics Institute (Shanghai, China).

***Sequencing Data Processing***

Raw sequences were filtered through a quality control pipeline using the Quantitative Insight into Microbial Ecology (QIIME) tool kit (5). The chimeras and singletons were detected and removed by Usearch software, and the high quality sequences were clustered into operational taxonomic units (OTUs) at the 97% similarity level. Samples were normalized to 24136 sequencing reads. The representative sequence was classified based on the Ribosomal Database Project (RDP) database (6) at the default confidence threshold of 0.8, trained on the SILVA reference database (release 111) (7). The alpha diversities (Shannon Index and Observed species), and beta diversity (Unweighted and Weighted Unifrac distance) were calculated. The ANalysis Of SIMilarity (ANOSIM) test was used to examine the statistical significance of differences in beta diversity. The datasets in the current study are available in the NCBI BioProject database with the BioProject ID PRJNA544381 (https://www.ncbi.nlm.nih.gov/sra/PRJNA544381).

***Data Analysis***

Rumen fermentation parameters were shown using bar charts made in R (v3.6.0) by ‘ggplot2’ package. The Anova test was used for significance calculation after detection of homogeneity of variance. After the global test was significant, a post-hoc analysis (Tukey's HSD test) was performed to determine which group of the independent variable differ from each other group.

Alpha diversity of the rumen microbial data among three treatments was tested using Kruskal–Wallis test and a post-hoc Dunn Kruskal-Wallis multiple comparison with Bonferroni adjustment to evaluate differences between two groups, and boxplots were made in R (‘ggpubr’ packages). Beta diversity was visualized with PCoA plot through R.

RandomForest classification model was performed to identify the top microbiome signatures to differentiate 3 supplementary feeding regimes. R package ‘AUCRF’ (v.1.1) was used to process RandomForest model and select optimal variables based on the area-under-the receiver operator characteristic curve (AUC) of the RandomForest method (AUCRF) (8). The relative abundances of all the microbiota were included for predictors selection. The ‘ntree’ parameters was set at 10,000 in the model. For calculation the probability of each selected variable, a 10-fold cross validation analysis and 20 times repetitions of cross validation were performed. The model accuracy, including AUC, sensitivity and specificity of variables, was calculated using the ‘pROC’ package (v.1.13). Thus, variables importance plot was generated based on the importance scores (Mean Decrease in Accuracy, MDA) of optimal features and their boxplots of selected features were drawn in R.

RandomForest regression model was used to select the rumen microbiota that were important for average daily intake of major nutrients (i.e., CP, NDF and NFC) and rumen fermentation parameters. The model was run in R software using ‘RandomForest’ package (v 4.6-14). The percent variance explained was reported for the estimation of accuracy of regression model. The top 50 selected features were then analyzed Pearson correlation with those macro indicators respectively.

Predictive function analysis was performed using the PICRUSt algorithm based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) classification using the closed-reference OTUs (9). The Functional Shifts’ Taxonomic Contributors (FishTaco) software was used to find the rumen bacteria driving the functional shifts between supplementary regimes in this study. A taxonomic abundance at OTUs’ level and functional abundance profile at levels 3 from the PICRUSt analysis were used. In pairwise comparisons, we labeled MRO groups as control and MRC or MCA as case, and tested MRC as control vs MCA as case. Each functional shift was grouped into case-associated with driving case-enrichment or attenuating case-enrichment, and control-associated driving case-enrichment or attenuating case-enrichment. The output results visualization was performed in FishTacoPlot package in R (Version 3.6.0).

Network analysis was performed by calculating all possible Pearson rank correlation coefficients (ρ) between microbial pairs. To minimize the occurrence of spurious associations, we considered a valid co-occurrence between two different taxa if a correlation co-efficiency over 0.6 or less than 0.6 and statistically significant. The network was demonstrated by using the ‘igraph’ package in R with edges connecting nodes (bacterial taxa). The subnetworks in regimes were produced based on the betweenness cluster calculated by the Girvan-Newman algorithm (10).

**References:**

1. Aoa C.2000. Association of official analytical chemists. Official methods of analysis 12.

2. Xiaokang XIEBHWWSBIYTAOHCUIKDQZN.2019. Effects of Early Feeding on Rumen and Small Intestine Morphology of Goat Kids. ACTA VETERINARIA ET ZOOTECHNICA SINICA 50:1006-1015.

3. Makkar HPS, Sharma OP, Dawra RK, Negi SS.1982. Simple Determination of Microbial Protein in Rumen Liquor. Journal of Dairy Science 65:2170-2173.

4. Jiao JZ, Wang PP, He ZX, Tang SX, Zhou CS, Han XF, Wang M, Wu DQ, Kang JH, Tan ZL.2014. In vitro evaluation on neutral detergent fiber and cellulose digestion by post-ruminal microorganisms in goats. Journal of the Science of Food and Agriculture 94:1745-1752.

5. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Pena AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335-6.

6. Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM.2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res 37:D141-5.

7. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J, Glockner FO.2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Research 35:7188-7196.

8. Calle ML, Urrea V, Boulesteix AL, Malats N.2011. AUC-RF: a new strategy for genomic profiling with random forest. Hum Hered 72:121-32.

9. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Thurber RLV, Knight R, Beiko RG, Huttenhower C.2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31:814.

10. Girvan M, Newman ME.2002. Community structure in social and biological networks. Proc Natl Acad Sci U S A 99:7821-6.