

# Fungal and ciliate protozoa are the main rumen microbes associated with methane emissions in dairy cattle

Adrián López-García <sup>1</sup>, Alejandro Saborío-Montero <sup>1,2</sup>, Mónica Gutiérrez-Rivas <sup>1</sup>, Raquel Atxaerandio <sup>3</sup>, Idoia Goiri <sup>3</sup>, Aser García-Rodríguez <sup>3</sup>, Jose A. Jiménez-Montero <sup>4</sup>, Carmen González <sup>1</sup>, Javier Tamames <sup>5</sup>, Fernando Puente-Sánchez <sup>5</sup>, Magdalena Serrano <sup>1</sup>, Rafael Carrasco <sup>6</sup>, Cristina Óvilo <sup>1</sup> and Oscar González-Recio <sup>1,7,\*</sup>

<sup>1</sup>Departamento de Mejora Genética Animal, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Crta. de la Coruña km 7.5, 28040 Madrid, Spain

<sup>2</sup>Escuela de Zootecnia y Centro de Investigación en Nutrición Animal, Universidad de Costa Rica, 11501 San José, Costa Rica

<sup>3</sup>NEIKER – Instituto Vasco de Investigación y Desarrollo Agrario. Basque Research and Technology Alliance (BRTA), Campus Agroalimentario de Arkaute s/n, 01192 Arkaute, Spain

<sup>4</sup>Confederación de Asociaciones de Frisona Española (CONAFE), Ctra. de Andalucía km 23600 Valdemoro, 28340 Madrid, Spain

<sup>5</sup>Departamento de Biología de Sistemas, Centro Nacional de Biotecnología, CSIC, Madrid, 28049 Madrid, Spain

<sup>6</sup>Departamento de Periodismo y Nuevos Medios, Universidad Complutense de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

<sup>7</sup>Departamento de Producción Agraria, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain

\*Correspondence address. Óscar González-Recio. INIA-CSIC, Dpt Mejora Genética Animal, Crta. de la Coruña km 7.5, 28040 Madrid, Spain; E-mail: gonzalez.oscar@inia.es

## Abstract

**Background:** Mitigating the effects of global warming has become the main challenge for humanity in recent decades. Livestock farming contributes to greenhouse gas emissions, with an important output of methane from enteric fermentation processes, mostly in ruminants. Because ruminal microbiota is directly involved in digestive fermentation processes and methane biosynthesis, understanding the ecological relationships between rumen microorganisms and their active metabolic pathways is essential for reducing emissions. This study analysed whole rumen metagenome using long reads and considering its compositional nature in order to disentangle the role of rumen microbes in methane emissions.

**Results:** The  $\beta$ -diversity analyses suggested a subtle association between methane production and overall microbiota composition ( $0.01 < R^2 < 0.02$ ). Differential abundance analysis identified 36 genera and 279 KEGGs as significantly associated with methane production ( $P_{\text{adj}} < 0.05$ ). Those genera associated with high methane production were Eukaryota from Alveolata and Fungi clades, while Bacteria were associated with low methane emissions. The genus-level association network showed 2 clusters grouping Eukaryota and Bacteria, respectively. Regarding microbial gene functions, 41 KEGGs were found to be differentially abundant between low- and high-emission animals and were mainly involved in metabolic pathways. No KEGGs included in the methane metabolism pathway (ko00680) were detected as associated with high methane emissions. The KEGG network showed 3 clusters grouping KEGGs associated with high emissions, low emissions, and not differentially abundant in either. A deeper analysis of the differentially abundant KEGGs revealed that genes related with anaerobic respiration through nitrate degradation were more abundant in low-emission animals.

**Conclusions:** Methane emissions are largely associated with the relative abundance of ciliates and fungi. The role of nitrate electron acceptors can be particularly important because this respiration mechanism directly competes with methanogenesis. Whole metagenome sequencing is necessary to jointly consider the relative abundance of Bacteria, Archaea, and Eukaryota in the statistical analyses. Nutritional and genetic strategies to reduce CH<sub>4</sub> emissions should focus on reducing the relative abundance of Alveolata and Fungi in the rumen. This experiment has generated the largest ONT ruminal metagenomic dataset currently available.

**Keywords:** dairy cattle, microbiome, rumen, methane, Nanopore, long reads

## Introduction

Next-generation sequencing technologies have provided special relevance to microbial communities from different niches because they let their taxonomic and functional profile be identified. This has made it possible to unravel the relationships between host and microbiota, as well as the complex interactions between microbes, focusing on the special contribution of the role of digestive microbiome in complex traits both in humans [1] (e.g., Type 2 diabetes, cancer, mental diseases) and in domes-

tic animals [2,3] (e.g., feed efficiency, methane emissions, animal health).

Microbial communities are of special relevance in livestock. In ruminants, one of the main microbial communities is found in the rumen, owing to its high diversity and large microbial mass [4] and its main role in feed fermentation to provide substrate to the animal, which is then transformed into product. Additionally, enteric methane is produced in the rumen by methanogenic microorganisms during feed fermentation [5] and is the main contributor of greenhouse gases from livestock, with 2.8–3.5 gigatonnes of

Received: August 11, 2021. Revised: October 18, 2021. Accepted: November 30, 2021

© The Author(s) 2022. Published by Oxford University Press GigaScience. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

CO<sub>2</sub>-equivalent per year [6,7]. The ongoing climate emergency urgently calls for efficient strategies to mitigate the carbon footprint from all sectors, including agriculture and livestock farming. Previous studies have proven that complex traits in ruminants are usually influenced by global changes in ruminal microbial communities, more than by fluctuations in the abundance of specific microorganisms [8,9]. These global changes are usually due to the intricate interactions between different species in these communities (i.e., predation, competition of ecological niche, or co-dependency). Consequently, a better understanding of the interactions between microbial genes during methanogenesis is needed to propose strategies for reducing methane emissions. Promising strategies have been proposed to modulate the metagenome, nutrition, and genetics [10].

Classic statistical approaches do not allow the results of microbiome studies to be accurately assessed. The high sparsity of these data and their compositional nature generate multiple problems in statistical analysis, including subcompositional incoherence, increased false-positive rates in differential abundance analyses, and detection of spurious correlations [11].

As a consequence, new approaches considering both compositionality and multiple correlations are needed. It is also important to point out the advantages of whole-metagenome sequencing over metataxonomic studies because the latter cannot be used to determine functionality and because they pose some difficulties at simultaneously analysing different superkingdoms [12], which is necessary to account for the total variability of microbiomes and the interactions among their components. Different amplicons must be used to correctly classify Bacteria, Archaea, Protozoa, and Fungi, increasing the cost of the studies and involving additional bias due to PCR [13]. They pose the additional difficulty of a proper comparison between communities sequenced in different reactions with different primers. Nanopore sequencing offers a cost-efficient sequencing strategy for metagenomics studies, providing both taxonomical and functional information simultaneously and for microbes from all superkingdoms. This technology has been improved in recent years, allowing taxonomic and functional assignments to be performed with an accuracy comparable to Illumina [14].

The objective of this study was to characterize the taxonomical and functional composition of rumen microbiota using long sequence reads obtained with Nanopore technology, and their relationship with enteric methane emission.

## Results

### Taxonomy of microbial composition

After initial selection of core taxonomy, 6,394,671 reads with N50 = 4,022 bp were classified in 3,921 taxonomical features up to genus level. A filtering strategy was implemented to exclude low-abundance microbes while keeping the core microbiome relevant for methane emissions. This process removed 48,517 reads (<1%), which reduced the sparsity of the metagenome from 87% to 68%, although a large number of singleton and doubleton features remained (Supplementary Fig. S1). The final core subcomposition included a total of 6,318,344 reads, in 437 samples, classified in 1,240 taxonomical features: 967 known genera (722 bacteria, 13 archaea, and 232 eukaryotes) and 273 that only reached family rank (i.e., Unclassified denomination). Overall, 503 families, 277 orders, 158 classes, and 86 different phyla (37 bacterial phyla, 3 archaeal phyla, and 46 eukaryotic clades) were classified. The rel-

ative abundance (RA) distribution by superkingdoms and phyla is summarized below.

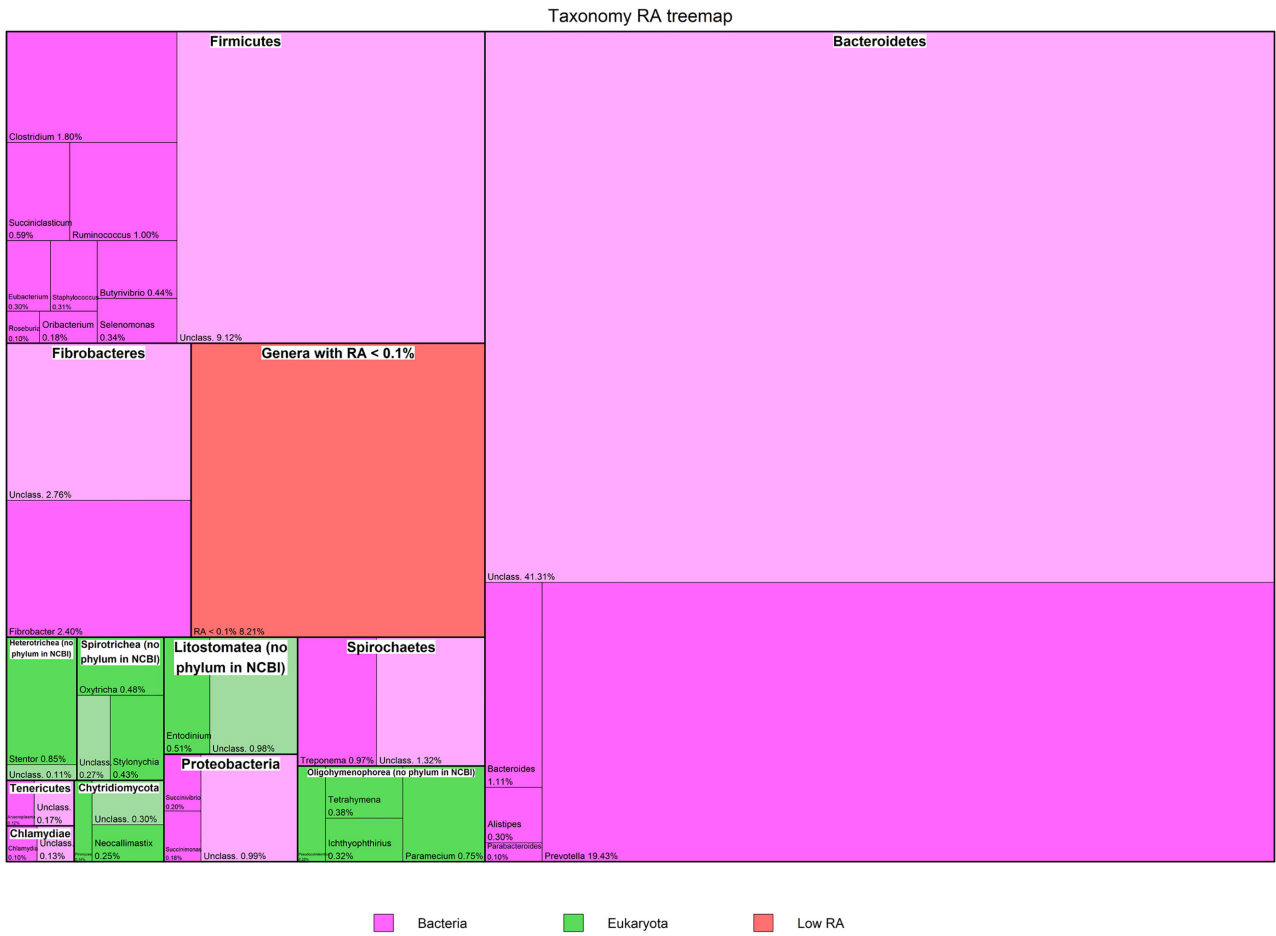
Predominant microorganisms in this core rumen subcomposition were bacteria (mean 91.6% [SD 6.93] of total mean RA) from Bacteroidetes, Firmicutes, and Fibrobacteres (Fig. 1), representing a mean RA of 63%, 16%, and 5%, respectively. The Bacteroidetes fraction was majorly composed by *Prevotella* and was the main representative genus in the total community (19.4% mean RA), along with other Prevotellaceae members. The Firmicutes group included a large number of genera. The order of Clostridiales dominated in terms of RA, with Lachnospiraceae and Ruminococcaceae families being the most representative ones. The remaining phyla (34) from the Bacteria superkingdom represented 7.6% mean RA of the core metagenome. Eukaryotes represented a total mean RA of 8.2% (SD 6.95) of the core subcomposition. Predominant eukaryotic clades were those included in the SAR supergroup (Stramenopiles-Alveolata-Rhizaria) [15], accounting for 6% of total mean RA, followed by Fungi (1.3% of total mean RA). Alveolata clade was the most abundant among the eukaryotes, with a high representation of unclassified Ophryoscolecidae, *Stentor*, and *Paramecium*. Archaea representation in the core subcomposition (mean 0.24% [SD 0.25] of total mean RA) consisted mostly of Methanomicrobia, Methanobacteria, and Thermoplasmata members. Yet, a large number of reads could not be assigned to a known genus. The relative abundance per animal of the most relevant taxonomic groups is depicted in Supplementary Fig. S2.

### Functionality of microbial composition

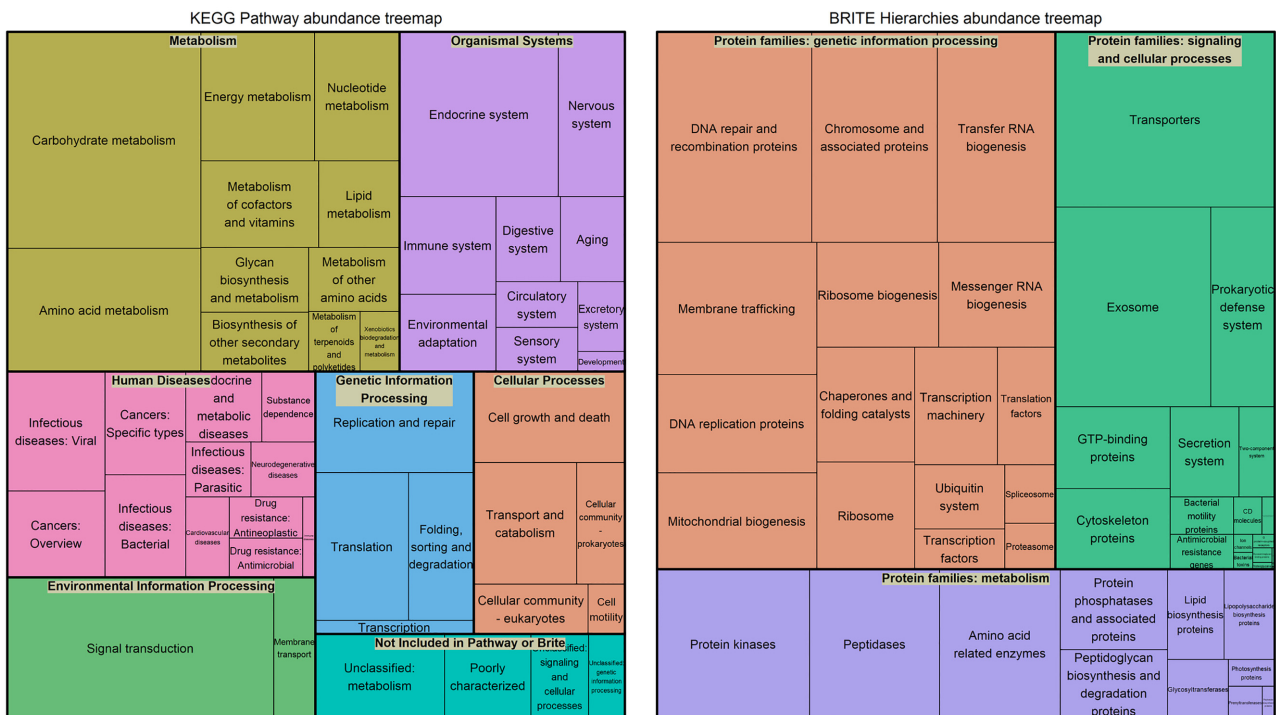
A total of 30,326,550 reads with N50 = 5,720 bp were assigned to KEGGs. After prevalence filtering, a total of 84,219 reads (0.28%) were removed and the sparsity was reduced from 72% to 39% (Supplementary Fig. S1). The final KEGG table was composed by 30,145,459 reads from 437 samples, classified in 6,644 KEGGs. These KEGG pathways and BRITE hierarchies [16–18] were represented in a Treemap according to their mean RA (Fig. 2). We found that 26% of the rumen metagenome functions were in pathways that represent the metabolism of carbohydrates, amino acids, and other biological compounds, as well as of energy metabolism. In addition, 34% of functions were involved in cellular generic processes (cell growth [3%], transport and catabolism [4%], genetic and environmental information processing [23%], and other [4%]). KEGG BRITE classification showed a high presence of proteins involved in cellular processes (36%) and metabolism (26%).

### $\beta$ -diversity and PERMANOVA analysis

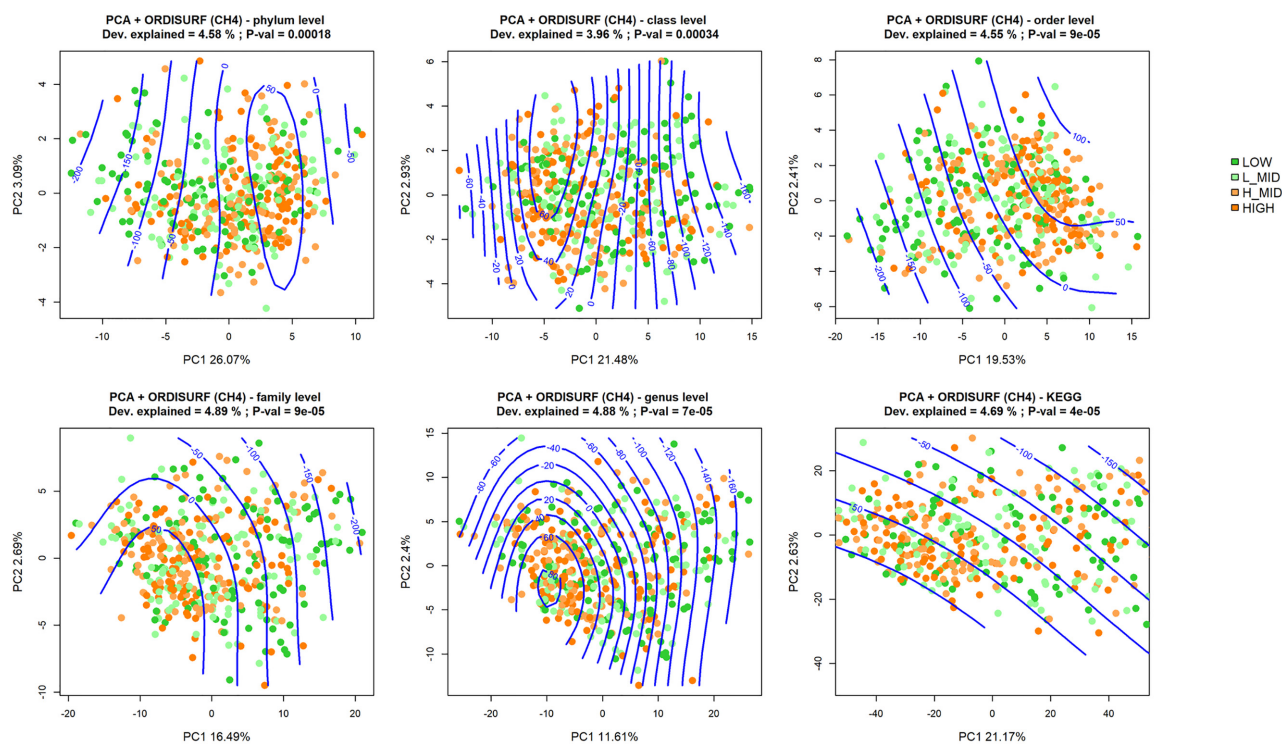
The  $\beta$ -diversity was represented in principal component analysis (PCA) between samples at 5 different taxonomic levels (phylum, class, order, family, and genus), as well as with KEGG, using centered log-ratio (CLR) transformed datasets. Then a permutational analysis of variance (PERMANOVA) was implemented [11], sequentially adding the effect of farm-batch (B), stage of lactation (SL), number of lactation (NL) and level of methane emissions (CH<sub>4</sub>) discretized in 4 groups (LOW, L-MID, H-MID, and HIGH). The visualization did not show a clear visual clustering of samples by methane emission levels (Fig. 3). However, a generalized additive model (GAM) smooth fitting allowed visualization of non-linear distribution patterns of the microbial samples according to CH<sub>4</sub> emissions inside the ordination at all taxonomic levels. The non-linear pattern was more evident at the phylum, class, and genus levels, although the proportion of methane variability explained was low ( $\approx$ 4.8% according to GAM model fitting). No relevant differences were visually appreciated using the KEGG information.



**Figure 1:** Mean relative abundance of genera. Mean relative abundance of core microbial taxa, including those classified only to family level (i.e., unclassified genera), which represent 60.2% of total abundance.



**Figure 2:** Metagenome functionality. TreeMap distribution of functionality abundances classified as KEGG pathways (left) and BRITE hierarchies (right) associated with core KEGG subcomposition.



**Figure 3:** Fitted surface representation of principal component analysis. Dots represent the samples using Euclidean distances of CLR-transformed taxa abundances, coloured by CH4 levels. CH4 emissions (ppm) corrected by number and stage of lactation are represented as smooth fitting following a generalized additive model (GAM) (green). Dev. Explained: variability explained by GAM; P-val: approximate significance of the smooth terms being 0 ( $\alpha = 0.05$ ).

**Table 1:** F statistic and P-values for stage of lactation (SL), number of lactation (NL), and methane emission (CH4) variables (added sequentially) and P-values from PERMANOVA of the entire dataset (i.e., including all superkingdoms)

Category	Variable	F statistic	R <sup>2</sup>	P-value*
Phylum	SL	6.1	0.014	<0.01
	NL	1.4	0.003	0.11
	CH4	2.8	0.019	<0.01
Class	SL	5.6	0.013	<0.01
	NL	1.5	0.003	0.07
	CH4	2.4	0.016	<0.01
Order	SL	5.4	0.012	<0.01
	NL	1.7	0.004	0.03
	CH4	2.3	0.016	<0.01
Family	SL	4.9	0.011	<0.01
	NL	1.6	0.004	0.03
	CH4	2.1	0.014	<0.01
Genus	SL	4.0	0.009	<0.01
	NL	1.4	0.003	0.03
	CH4	1.7	0.012	<0.01
KEGG	SL	5.3	0.012	<0.01
	NL	2.0	0.004	0.02
	CH4	2.4	0.016	<0.01

\*Italics indicate a statistically significant finding ( $P < 0.05$ ).

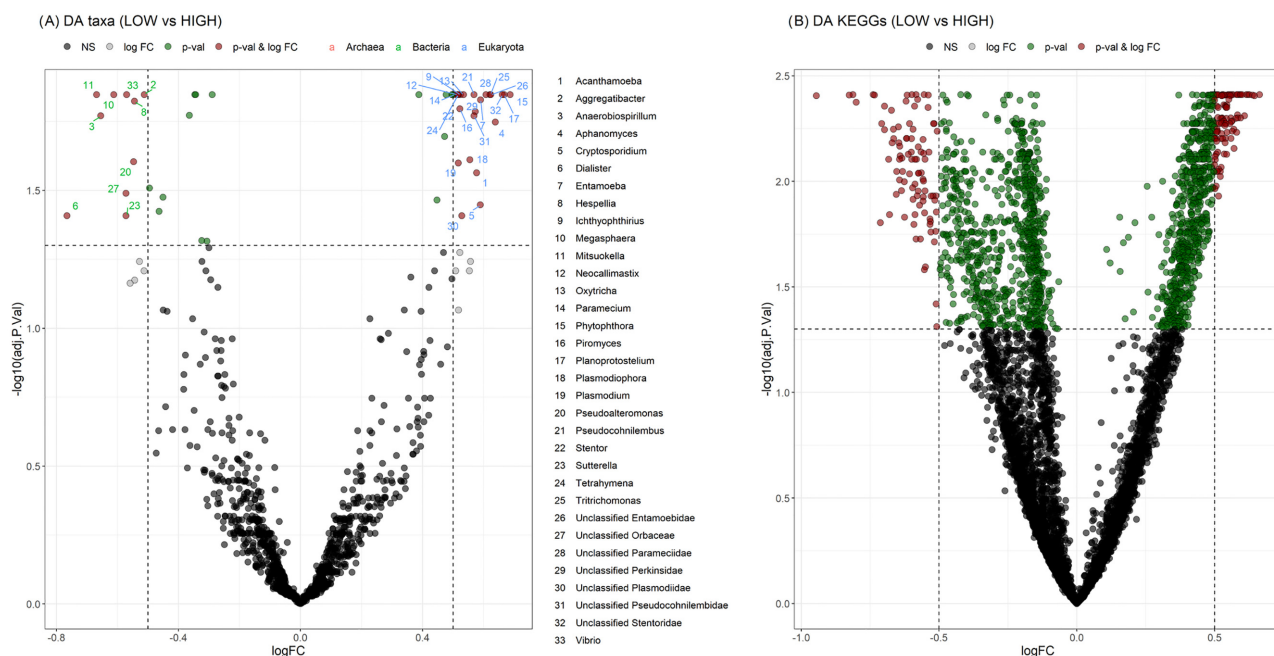
Nonetheless, some differences in the overall rumen microbiome composition between animals with different methane emissions were evidenced by the PERMANOVA analysis, both for taxonomy and functionality (Table 1). The results showed significant differences for the centroid distance between methane emission groups

at every taxonomic level and also for KEGGs ( $P < 0.01$ ), but they explained a low percentage of total variance ( $0.01 < R^2 < 0.02$ ).

### Rumen microbes associated with CH4 emissions

The effect of taxonomic features on methane emission levels was evaluated through differential abundance (DA) analysis. Thirty-three genera were found to be differentially abundant ( $P_{adj} < 0.05$ ) between LOW and HIGH emitters (Fig. 4A), while 15 genera showed DA between LOW and H-MID emitters and 1 genus between LOW and L-MID emitters (Supplementary Data S1). Note that 13 of the 15 genera showing DA ( $P_{adj} < 0.05$ ) between LOW and H-MID groups were also significant in the LOW vs HIGH contrast but not in LOW vs L-MID contrast, indicating gradual abundance change from low to high emitters. Accounting for all contrasts and duplicated genera, 36 DA genera had statistically significant results. We classified these genera according to their respective overabundance (OA) in the LOW or HIGH emissions groups. Thus, 10 of them were more abundant in the LOW group (LOW-OA) and 1 in the L-MID group. The remaining 25 genera were OA in the HIGH groups (HIGH-OA): HIGH (12), HIGH and H-MID (11), or H-MID (2). HIGH-OA genera represented an overall RA of 4.15%, whereas LOW-OA genera accounted for 0.25% of total RA. The 2 genera overabundant in H-MID were *Dictyostelium* and *Unclassified Eimeriidae*, and the one associated to L-MID was classified as *Candidatus Izimaplasma* (Tenericutes). The  $\log_2FC$  values ranged between 0.7 and  $-0.7$  in genera showing DA for methane emission levels, highlighting that the differences between groups were moderate.

Overall, DA results indicate that taxa associated with higher methane levels belong to the Eukaryota superkingdom, while those associated with lower emissions were bacteria. We found



**Figure 4:** Volcano plots. Volcano plot representing the differential abundance (DA) of genera (A) and KEGGs (B) between LOW and HIGH groups from limma. Significance thresholds were established at  $P_{adj} = 0.05$  and  $\log_2FC = \pm 0.5$ . Red points: Significant features with DA above the fold change (FC) threshold. Green points: Significant features with DA below the FC threshold. Gray points: Non-significant features with DA above the FC threshold. Black points: Non-significant features with DA below the FC threshold.

multiple Ciliophora genera associated with the HIGH group (mostly Parameciidae, Stentoridae, and Pseudocohnilembidae members) but also Amoebozoa and some Fungi or pseudo-fungi. Other bacterial genera associated with lower methane production were *Hespellia*, from Clostridiales, and *Sutterella*, an asaccharolytic genus from Betaproteobacteria.

### Microbial gene function associated with CH<sub>4</sub> emissions

DA analysis was also performed for KEGG features on methane emission levels. A total of 192 KEGGs were DA between the LOW and HIGH emissions groups (Fig. 4B). Differences were also found between the LOW and H-MID groups (Supplementary Data S1). As in the taxonomy dataset, some of the KEGGs presented significant DA in both LOW vs HIGH and LOW vs H-MID contrasts. Accounting for these duplicates and all the contrasts, 182 were overabundant in the high-emissions groups (HIGH-OA), whereas 97 KEGGs were overabundant in low-emissions groups (LOW-OA). The overall RA for HIGH-OA KEGGs was 2.31% and 0.64% for LOW-OA KEGGs. Of these, 13 HIGH-OA KEGGs and 28 LOW-OA KEGGs were assigned to metabolic pathways. No KEGGs from the ko00680 pathway were found as HIGH-OA. KEGGs related to inositol-phosphate metabolism (K00889, K01110, K18082, and K20279), starch and sucrose metabolism (K01203), or several lipid metabolism pathways were present in the HIGH-OA group. According to LOW-OA KEGGs, some of them were involved in volatile fatty acid metabolism (e.g., K00209 enoyl-[acyl-carrier protein] reductase [EC:1.3.1.9], K01902 succinyl-CoA synthetase alpha subunit [EC:6.2.1.5], and K01682 aconitate hydratase 2 [EC:4.2.1.3]) and the K09251 putrescine aminotransferase (EC:2.6.1.82) related to putrescine and cadaverine degradation to 4-amino-butanoate (GABA) or 2-oxoglutarate. Also, several KEGGs in the LOW-OA group were related to nitrogen metabolism (K00370 and K00371 nitrate reductase subunits [EC:1.7.5.1]), oxidative phosphorylation (K03885 NADH de-

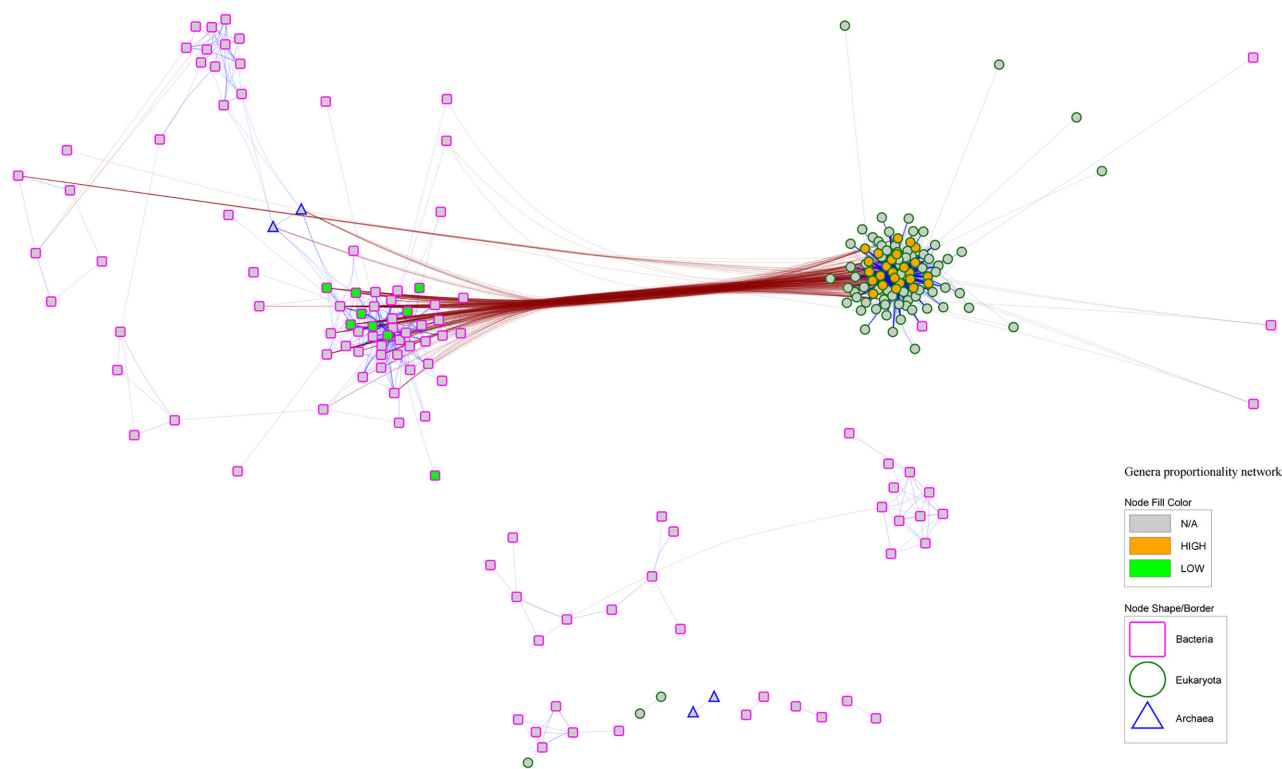
hydrogenase [EC:1.6.99.3]), and to carbohydrate, lipid, or vitamin metabolism pathways. The ko00680 KEGG K13788 was also overabundant in the LOW emissions group.

### Co-abundance of genera and KEGGs

Interaction networks were built using the previous results in order to visualize the association between taxa and genes using pairwise correlations between features. Pairwise proportionality correlation coefficients ( $\rho_p$ ) were calculated on the CLR-transformed datasets for phylum, genus, and KEGG features to mitigate the effect of spurious correlations that can potentially surge in compositional data [19].

The most relevant pairwise proportionalities between genera and between KEGGs were visualized as interaction networks, classifying features as associated with high methane emissions (HIGH), low methane emissions (LOW), or not associated with methane emissions (N/A), according to the results from the differential abundance analyses. The interaction networks for genera and KEGGs are shown in Figs 5 and 6, respectively.

Eukaryotes clustered together in the network with large representation of the SAR supergroup and showed negative proportionality to bacteria. The genera that were associated with higher methane emissions belonged to the Eukaryota superkingdom (Ciliophora and Fungi), whereas Bacteria were associated with lower CH<sub>4</sub> production. The strongest inverse proportionalities between both subpopulations connected several eukaryotes with Unclassified Veillonellaceae and *Oribacterium* ( $-0.64 < \rho_p < -0.53$ ); i.e., microbiomes with lower abundance of *Oribacterium* or Veillonellaceae tend to present larger abundances of protozoa and fungi and were therefore associated with larger emissions. Unclassified microbes from Neocallimastigaceae, Oxytrichidae, and Vibrionaceae families showed the highest centrality and a large degree of connectivity.



**Figure 5:** Taxonomy interaction network. Pairwise proportionalities between genera with  $|\rho_p| \geq 0.4$ . Node shapes and shape colour indicate superkingdom, and node fill colour, CH<sub>4</sub> association. Blue links indicate direct proportionality (>0), and brown links, inverse proportionality (<0).

The functional network showed 3 main clusters that grouped KEGGs associated with HIGH methane level (cluster H), KEGGs not related to methane emissions (cluster N), and a small one including KEGGs associated with lower emissions (cluster L). Connections between clusters were not symmetric: H cluster was connected to N cluster by inverse proportionalities between some of their components, but the L cluster seemed to be connected only to N cluster by direct proportionalities through non-clustered KEGGs. Also, most of the ko00680 KEGGs (i.e., directly involved in methanogenesis or participating in pathways leading to methanogenesis precursors) did not appear as DA between high-emission and low-emission cows.

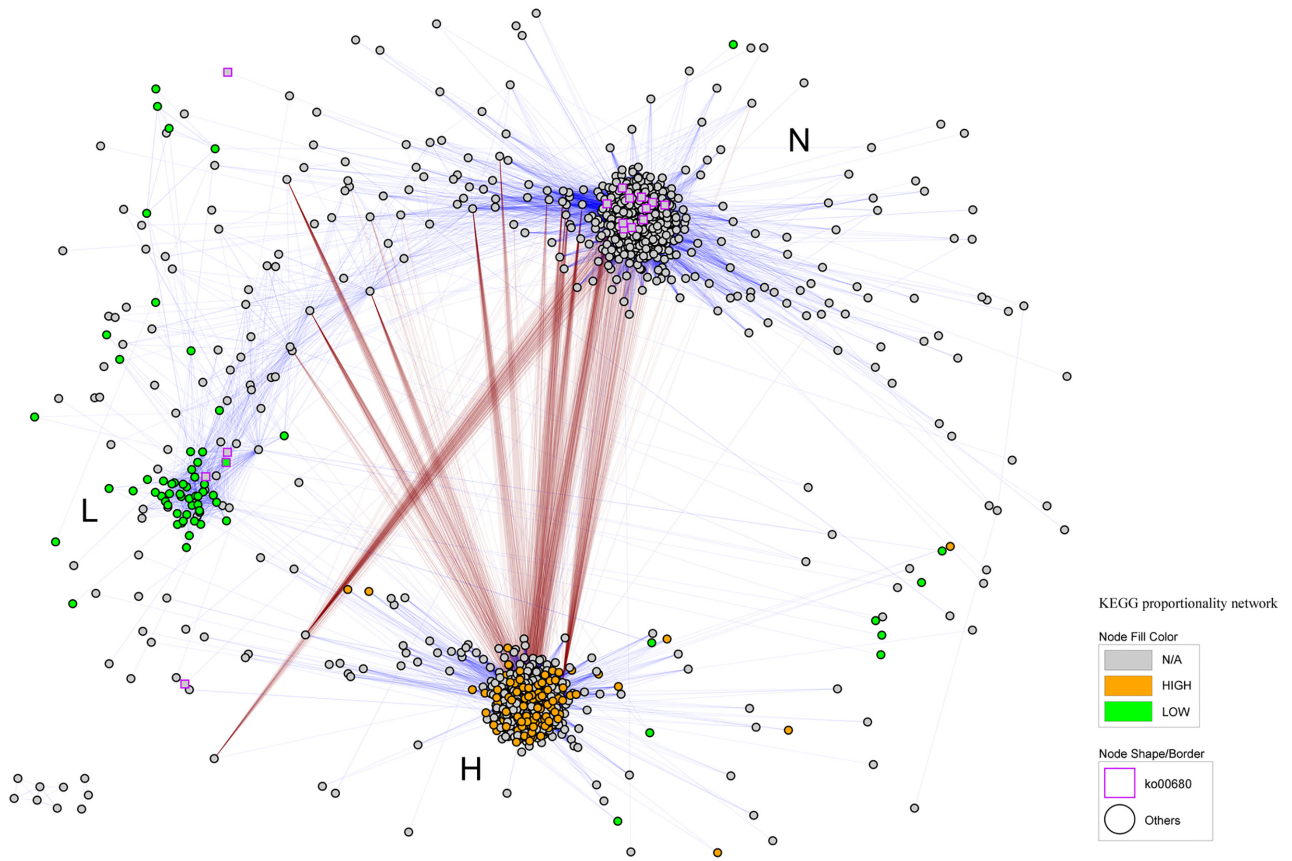
### Distribution of genes among clades

A traceback of genes' taxonomy was carried out, separately for ko00680 KEGGs and for DA KEGGs. A total of 30 of the 85 ko00680 KEGGs were predominant in Archaea groups, 1 predominated in Eukaryota (K05979), and the rest were predominant in Bacteria (Fig. 7). Although the RA distribution of these KEGGs was normally between 60% and 100% in the predominant superkingdom, 4 KEGGs were more evenly distributed between clades: K01007 and K00863 had RA < 60% in Bacteria and showed RA > 30% in Eukaryota; K05979 was the KEGG predominating in Eukaryota, but with RA near 60% (38% in Bacteria and 12% in Archaea); and K14080 had RA of 57% in Archaea and 43% in Bacteria. Regarding the DA KEGGs, those from the LOW-OA group showed larger abundance in Bacteria, mostly in genera from Proteobacteria, Bacteroidetes, and Firmicutes phyla. Different groups of bacteria also carried KEGGs from the HIGH-OA group, although these KEGGs were more abundant in eukaryotes. The HIGH-OA KEGGs were mainly mapped to unclassified eukaryotes, but those that could

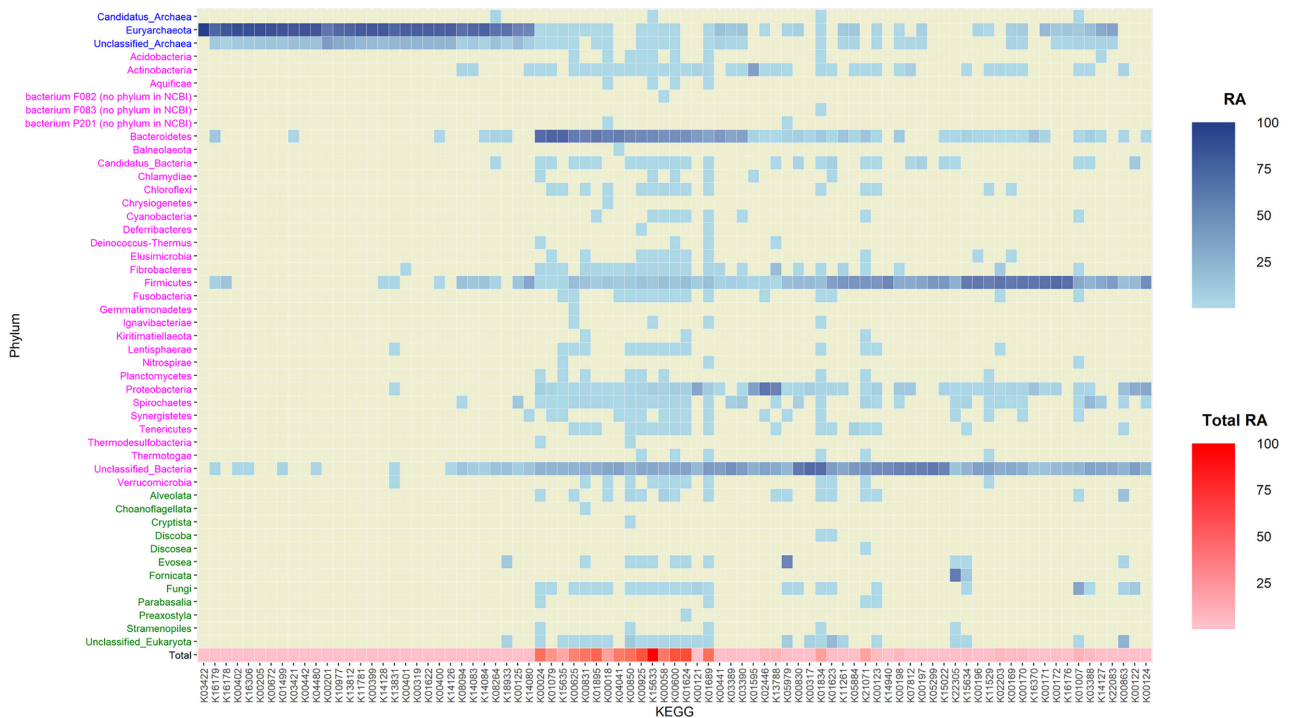
be classified belonged most often to Fungi and SAR supergroup (Fig. 8).

### Discussion

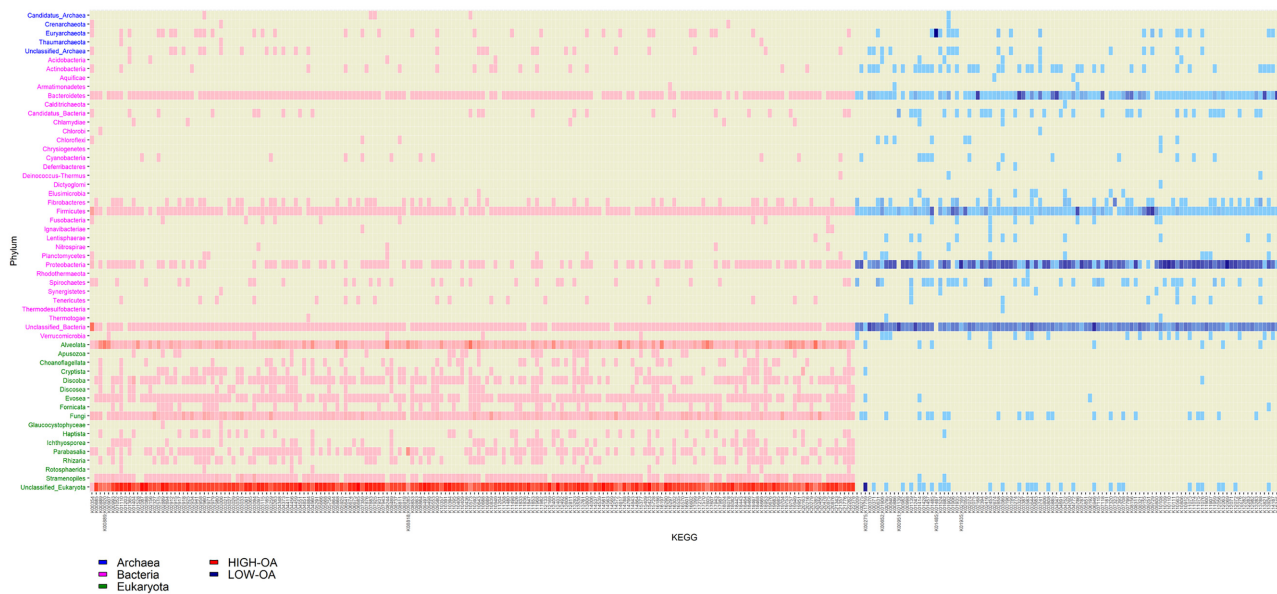
In this study we assessed the composition of the ruminal microbiota using long reads from Nanopore sequencing technology. We observed predominance of Bacteroidetes, Firmicutes, and Fibrobacteres, as reported in previous studies [8,20]. Bacteroidetes and Firmicutes are common bacteria in all kind of ecosystems, including gut microbiota of multiple animals. The fraction of Bacteroidetes was mainly composed by *Prevotella*, a group of anaerobic gram-negative bacteria involved in saccharolytic processes [21]. Their large abundance in the digestive microbiota has been previously reported in ruminant [22–26] and monogastric species [27,28]. Firmicutes were less abundant, with a more diverse distribution of genera. Fibrobacteres, a small group of cellulose-degrading bacteria usually present in ruminant digestive systems [29], was mainly represented by the *Fibrobacter* genus. Eukaryotes also represented a relevant amount of the rumen core metagenome. This group has been reported to contribute up to 50% of total ruminal biomass [30]. The SAR supergroup and Fungi were the most relevant ones, which are found in a wide variety of ruminants and pseudoruminants [15,31]. Other eukaryotes included *Stentor* and *Paramecium*; the former are aquatic free-living heterotrichs that can be particle filtrators or predators of other protozoa and live symbiotically with some algae species [32,33], whereas the latter are well-known ciliates that predate bacteria and other microorganisms, including protozoa [34]. Archaeal fraction was mostly composed of strict methanogenic organisms from Methanomicrobia and Methanobacteria clades [35]



**Figure 6:** Functionality interaction network. Presented are pairwise proportionalities between KEGGs with  $|\rho_p| \geq 0.7$ . Node shapes and shape colour indicate participation in methane metabolism (ko00680 [direct or indirect participation] or other [no participation]), and node fill colour, CH<sub>4</sub> association. Blue links indicate direct proportionality (>0), and brown links, inverse proportionality (<0). Clusters are indicated as L (KEGGs associated with LOW methane), H (KEGGs associated with HIGH methane), and N (KEGGs not related to methane emissions).



**Figure 7:** Taxonomy of ko00680 KEGGs. Relative abundance of KEGGs present in ko00680 pathway for each phylum in Archaea (blue), Bacteria (fuschia), and Eukaryota (green) superkingdoms. Relative abundance of each ko00680-KEGG with respect to the sum of reads mapped to all ko00680-KEGGs.



**Figure 8:** Taxonomic distribution of DA KEGGs. Red density scale represents KEGGs overabundant (OA) in HIGH emitters; Blue density scale represents KEGGs OA in LOW emitters. More intense colours mean a higher number of reads assigned to 1 phylum.

but also included Thermoplasmata, which are methylotrophic-methanogenic acidophilic organisms [36].

The DA analysis showed that ciliates, fungi, and pseudo-fungi were more abundant in cows with higher levels of methane emissions. Microbes associated with lower methane emissions were saccharolytic members of class Gammaproteobacteria (*Anaerobiospirillum* [37], *Vibrio* [38], or *Pseudoalteromonas* [39]), as well as Negativicutes genera from Veillonellaceae (*Dialister*, *Megasphaera*) and Selenomonadaceae (*Mitsuokella*). *Dialister* produce succinate decarboxylation, and *Megasphaera* ferment carbohydrate and lactate [40], while *Mitsuokella* are saccharolytic bacteria [41]. The low-emissions ruminotype had larger abundance of Proteobacteria and Firmicutes genera. Other authors also reported higher abundances of these bacterial phyla in low methane emissions animals [8]. Lactate and succinate producers have been reported to be more abundant in low-emitters as well [42], supporting the higher abundance of *Anaerobiospirillum* or *Megasphaera* in LOW animals.

Despite this association between methane and large taxonomic groups, it is of interest to find out which specific clades and microbial genes are participating directly or indirectly in methanogenesis. The genera co-abundance network showed a clear cluster of eukaryotes, with many of them being significantly more abundant in the high-emissions group. Other authors have already established a positive correlation between fungi abundance and methane emissions [8], as well as a close interdependence of protists and fungi. Although correlation between methane emissions and protozoa abundances is still under discussion [43,44], current meta-analyses point to a linear relationship between protozoa abundance and methane emissions ( $r = 0.96$ ) [45].

Interestingly, no taxonomic group of methanogenic archaea showed association with methane emissions. The relationship between Archaea and methane production in rumen is not consistent in the literature. Some authors reported either individual relationships between methane emissions and some archaeal species [46,47] or correlations between overall archaeal gene abundance and methane emissions level [43,48]. However, other studies showed no relationship between methanogenic Archaea

and methane [47,49]. All studies to date showed a low relative abundance of archaea in the rumen, compared to eukaryotes and bacteria [50]. However, the association between the abundance of rumen eukaryotes and methane emissions has been demonstrated through defaunation experiments, both *in vitro* [51,52] and *in vivo* [44,53], with lower emissions in defaunated animals [54]. This has been attributed to the tight link existing between methanogenic archaea abundance and some fungi and protozoa [50]. Specifically, ciliates and some Chytridiomycota (e.g., *Neocallimastix* sp.) are known to symbiotically engulf a variety of methanogenic archaea. They provide the archaea with substrate for methane production from  $H_2$  produced in their hydrogenosomes, as well as protection against oxygen toxicity [30,55,56]. Thus, free-living methanogens might represent a low fraction of microbial population [45], and  $CH_4$  biosynthesis might be more influenced by endosymbiotic methanogens [55]. Hence, a larger methanogenesis activity is expected to be correlated with a larger abundance of eukaryotes, especially ciliates, which are more abundant and better represented. Another partial explanation for the low abundance of free archaea, and thereby for the lack of association between Archaea and methane emissions in previous studies [10], is that lysis of archaea cell walls often requires specific protocols during DNA extraction, and they might be under-represented in metagenomics studies [57].

In terms of Gene Ontology, the KEGGs were associated with several metabolic functions and cellular processes (nutrient metabolism and biosynthesis, cellular transport, cell growth, or genetic information processing). Pathways related to pathogenic activity were also found, in agreement with the RA of several genera that include known pathogenic species (e.g., *Vibrio*, *Haemophilus*, *Trypanosoma*, or *Staphylococcus*), although not every species from these genera is pathogenic, but opportunistic or commensal organisms. Besides, pathogenic activity presence in our dataset might be biased owing to a larger representation of human-related diseases in the databases. The KEGGs were classified according to their presence or absence in ko00680 pathway (methane metabolism), as a way to evaluate their direct involvement in methanogenesis or an indirect involvement in



pathways leading to biosynthesis of precursor compounds. Although we found several ko00680 KEGGs, which are presumably involved in the biosynthesis of methanogenesis precursors, most of them were not associated with methane emissions (i.e., not differentially abundant between methane groups). Most of these KEGGs were mainly present in bacteria or eukaryotes and might be functioning in metabolic pathways not related to methanogenesis. For instance, some of the KEGGs inside the methane metabolism pathway can also be involved in glycine, serine, and threonine metabolism (e.g., K00058, K00831, K01079, and K00600), pyruvate and propanoate metabolism (e.g., K00625 and K13788), glycolysis (e.g., K01689, K15633, K01624, and K02446), or anaerobic carbon fixation (e.g., K00198) [16–18]. Another group of ko00680 KEGGs is exclusive from Archaea, but the under-representation of this clade in our dataset might obscure statistical significance.

Other detected KEGGs could be indirectly related to methanogenesis through biosynthesis of precursor compounds. For instance, K00209 and K13788 are involved in butyrate and propanoate biosynthesis, being essentially carried by primary fermentative bacteria [58]. Then the volatile fatty acids can be used by secondary fermenters to produce methanogenesis precursors such as H<sub>2</sub>, CO<sub>2</sub>, acetate, and formate [59,60]. In fact, K13788 is a phosphate acetyltransferase (EC:2.3.1.8) that can be involved in the biosynthesis of acetate from acetyl-CoA [61]. Also, K09251 is involved in biosynthesis of GABA and 2-oxoglutarate. GABA has been related to a volatile fatty acid concentration increment [62], while 2-oxoacid compounds can be used by Archaea to synthesize coenzyme M and coenzyme B, which are essential in methane production [63]. However, all these KEGGs were observed as overabundant in the LOW methane group, suggesting a strong presence of fermentative bacteria in these animals, not directly correlated with methane production.

Other KEGGs that were overabundant in LOW emitters might offer an explanation of the lower presence of active methanogenesis processes through competence mechanisms (e.g., LOW-OA KEGGs K01682, K01902, and K13788 are involved in citrate cycle and pyruvate metabolism, related to respiration). The K00370 and K00371 are nitrate oxidoreductase subunits playing a role in anaerobic respiration using nitrate as electron acceptor. This enzyme uses nitrate as electron acceptor, a process that has been reported as a competitive inhibitor of methanogenesis [64,65]. Nitrate supplementation has proven to be a useful strategy to mitigate methane emissions [66]. Nitrite produced by the nitrate-reductases has a known antimicrobial effect and toxicity to animal cells [67–69], which might also reduce the proportion of free archaea in LOW animals, although toxicity to archaea must be further studied [70]. However, the role of ciliates and fungi must be clarified because their abundance is also lower in LOW emitters. We hypothesize that the predatory nature of these eukaryotes might be a control mechanism for bacterial populations, and their lower relative abundance in LOW animals might allow overgrowth of related bacteria. Nevertheless, there is the possibility that a higher proportion of facultative anaerobes using nitrate as acceptor might affect ciliate populations by toxicity, thus reducing the presence of endosymbiotic methanogenic archaea.

The SqueezeMeta software [71] uses a last common ancestor (LCA) algorithm, which assigns to 1 read the lowest-level taxon common to all hits, using a stringent cut-off identity value for each taxonomic rank. On its part, functional assignments are done with the fun3 algorithm, which by default assigns the hit with the highest mean bitscore compared to the n first hits pass-

ing the e-value, identity, and coverage filters. This LCA approach ensures that reads have a large probability of being correctly classified, at the expense of a large number of reads remaining unclassified, which explains the larger number of reads assigned to a known KEGG than to taxa. Despite this strict requirement, this composition is consistent with other populations reported before [2,3,20]. Most studies to date report large abundance of Bacteroidetes and Firmicutes, with *Prevotella* spp. as the most prevalent genus. Some minor discrepancies with other studies were observed in the RA of the core subcomposition. For example, Wallace et al. [20] showed a higher presence of Proteobacteria and Euryarchaeota, although using amplicons instead of whole-metagenome sequencing.

Our statistical approach evidenced the difficulty of inferring a phenotypic association between microbiome composition and methane production, with an important role of environmental factors that mask the statistical signal. However, a meaningful relationship between the microbiome composition and methane emissions could be uncovered yet, emphasizing the role of the different phyla, with the Eukaryota superkingdom being of particular relevance. Previous studies also revealed a link between ruminal microbiota and methane production. Difford et al. [3] showed different clusters of high and low methane emitters according to their bacterial and archaeal subcomposition. Danielsson et al. [46] also found clustering for low and high methane emitters within prokaryotic rumen subcompositions. Wallace et al. [20] found that a core set of rumen microbiome was capable of explaining up to 30% of methane emissions variability, mostly formed by prokaryotes. The aforementioned studies used different methodologies, like amplicon analysis and operational taxonomic unit clustering, contrasting with our full-metagenome genus-clustering protocol, which increases the information entropy. Stewart et al. [72] used Nanopore sequencing and found significant differences between low and high methane emitter sheep, with clear clustering between groups, but using a lower number of microbial groups and animals in the same farm with similar management practices.

## Conclusions

The full metagenome compositional analysis used in this study provided novel insights in the association between the microbiota and CH<sub>4</sub> emissions through differential abundance analysis, pairwise correlation, and interaction networks. Our approach evidenced a phenotypic association between microbiome composition and methane production, regardless of the challenges posed by the microbiome complexity and the compositional nature of the data. This association is mainly driven by the relative abundance of ciliates and fungi, which carry host-specific genetic functions providing substrate to the methanogenic archaea. On the other side, we detected some bacterial groups that performed a more efficient feed digestion, leaving less hydrogen available to archaea and hence associated with lower methane emissions.

This study generated the largest ruminal metagenomic dataset sequenced using ONT and grants free access to a publicly available dataset. The complexity of the rumen microbiome and the compositional nature of their sequencing data require proper statistical methods to allow disentangling the role of microbes and their genes in host complex traits such as methane emissions. Future nutritional and genetic strategies to reduce CH<sub>4</sub> emissions should focus on reducing the relative abundance of Alveolata and Fungi in the rumen, without impairing other important metabolic processes for an efficient feed digestion in ruminants.

## Methods

### Animal housing and feeding

Our cohort included 439 Holstein lactating cows sampled at 14 different herds from northern Spain (Cantabria, Euskadi, Navarra, and Girona regions). The animals received total mixed ration diet differently formulated for each individual herd, although most of them were based on maize and grass silage plus concentrate. Cows were fed *ad libitum*, with concentrate supplementation in the automatic milking station (AMS) during milking.

### Methane measurement

Methane concentration was individually recorded through breath sampling during each cow visit to the AMS (3–7 times daily) in a period of 2–3 weeks. Eructation peaks were recorded using a non-dispersive infrared methane detector (Guardian NG infrared gas monitor, Edinburgh Sensors, Livingston, UK) as described by Rey *et al.* [73]. Each cow's peaks were then averaged to get a unique methane record per cow, as described by López-Paredes *et al.* [74]. Animals were distributed in groups according to number of lactation (NL) and stage of lactation (SL) criteria. Furthermore, quartile-based qualitative categories were created for CH<sub>4</sub> recordings (ppm), resulting in a methane factor (CH<sub>4</sub>) with 4 levels (LOW, L-MID, H-MID, and HIGH methane emissions).

### Ruminal content sampling

Ruminal fluid was sampled using an oral tube (18 mm diameter and 160 mm long) connected to a 1,000-mL Erlenmeyer flask and continued to a mechanical pump (Vacubrand ME 2SI, Wertheim, Germany), with all the material contacting the cow being carefully cleaned between cows. Each animal was moved to an individual stall for this process. The solid fraction of the ruminal content was discarded by filtering through 4 layers of sterile cheesecloth, while the outcoming liquid fraction was instantly frozen using liquid nitrogen and then stored at –80°C until DNA extraction.

### DNA extraction and sequencing

Genomic DNA was extracted from 250 µL of each thawed and homogenized ruminal content sample, using the “DNeasy Power Soil” commercial kit (Qiagen, Valencia, CA, USA). Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA) and Nanodrop ND-1000 UV/Vis spectrophotometer (Nanodrop Technologies Inc., Wilmington, DE, USA) were used to measure DNA concentration and purity. The 260/280 and 260/230 ratios were ~1.8 and ~2.0, respectively. Oxford Nanopore Technologies (ONT) SQK-LSK109 Ligation Sequencing kit was used for multiplexed sequencing in MinION automatic sequencer. The 1D Native barcoding ONT kit (EXP-NBD104 or EXP-NBD114) was used for multiplexing the samples, pooling barcoded DNA from 12 samples for each run. Pooling was done using a 1.5-mL DNA LoBind tube to perform adapter ligation and sequenced using a R9.4.1 flow cell.

### Read processing, mapping, and filtering

Guppy toolkit (ONT) was used for basecalling. A quality control was then applied removing sequences with QS <7 and length <150 bp. Sequence analysis was performed using SqueezeMeta (SQM) pipeline for long reads [71], which performs Diamond Blastx against GenBank nr taxonomic database and against COG and KEGG functional databases, then identifying and annotating open reading frames using the LCA method for taxonomy and the fun3 algorithm for functional annotation (based on e-value and iden-

tity scores). This tool is specifically designed to process long reads from ONT.

A total of 49,718,901 reads were processed in Blastx by SQM longreads pipeline. Blastx mapped 25,750,755 reads (51.79%) to taxonomy (NCBI-nr database) or function (KEGG database). All sequences mapped as non-microbial (i.e., virus, animals, and plants) were discarded. Microbial sequences were then filtered by prevalence to reduce data sparsity and sequencing errors (Supplementary Data S2). A first estimation of sample sparsity and reads distribution was assessed using R. Two animals were then withdrawn from the filtered dataset, one owing to low read coverage and another owing to lack of host information, leaving 437 animals in the final dataset.

Genera were divided into superkingdom groups (Archaea, Bacteria, or Eukaryota), and KEGGs were sorted by their involvement in methane metabolism (MP): KEGGs included in the KEGG orthology pathway ko00680 (Methane metabolism) were labeled as “ko00680,” while the rest were identified as “Other.”

### Compositional data

Considering the compositional nature of metagenomic data, a CLR method [75] was applied using the unweighted option of the CLR function from the easyCODA R package [76] as follows:

$$\mathbf{x}_{\text{clr}} = [\log(x_1/G(x)), \log(x_2/G(x)) \dots \log(x_D/G(x))],$$

with  $G(x) = \sqrt[D]{x_1 * x_2 * \dots * x_D}$ .

Being  $\mathbf{x} = [x_1, x_2, \dots, x_D]$  a vector of counted features (taxa or KEGGs) in 1 sample and  $G(x)$  the geometric mean of  $\mathbf{x}$ . Count zero values in the initial data frame were imputed through the Geometric Bayesian Multiplicative procedure, using the zCompositions R package [77] `cmultRepl` function, so that logarithms could be computed.

### $\beta$ -diversity and PERMANOVA analysis

The CLR-transformed data (at phylum, class, order, family, genus, and KEGG levels) were used to explore  $\beta$ -diversity in the samples through PCA using the `prcomp` function in R. Fitted smooth surface of methane emissions corrected by SL and NL was included for principal components 1 and 2 using the `ordisurf` function from the `vegan` R package [78]. A generalized additive model smooth fitting (GAM) was used to elucidate non-linear distribution of samples in PCA according to methane emissions. Differences between centroid distances using methane as grouping variable (CH<sub>4</sub>) were determined through PERMANOVA [79,80] following this model and using the matrix of Aitchison distances between samples (i.e., the Euclidean distance on CLR-transformed data) as input variable:

$$D_{jklmi} = \mu + B_j + SL_k + NL_l + CH_{4n} + e_{jklmi},$$

with  $B_j$  being the farm-batch effect ( $j = 24$  levels),  $SL_k$  being the stage of lactation at the day of sampling ( $k = 3$  levels),  $NL_l$  the number of lactation ( $l = 2$  levels), and  $CH_{4n}$  the methane emission level ( $n = 4$  levels: LOW, L-MID, H-MID, HIGH), and  $e_{jklmi}$  was the corresponding residual term.

### Association between microbiota and methane production

Differential abundance of genera and KEGGs between samples regarding the different methane emissions levels was addressed through linear regression using `Limma` [81]. Count normalization and log-transformation were addressed using CLR-transformed data as inputs. P-values were adjusted by the Benjamini-Hochberg

method, to control the false discovery rate (FDR). Differential abundance threshold was set to  $|\log_2FC| \geq 0.5$  and the adjusted significance threshold was set to  $\alpha = 0.05$ .

### Pairwise proportionality analysis

Pairwise correlations between phyla, genera, and KEGGs were calculated as described in the `propr` R package [82]. Proportionality coefficient  $\rho_p$  [83] under CLR data transformation was chosen. Thresholds were selected according to 2 conditions: (i) representing the maximum number of proportionalities avoiding computational issues; (ii)  $FDR < 1\%$ . The thresholds used were  $|\rho_p| \geq 0.4$  for genera proportionalities and  $|\rho_p| \geq 0.7$  for KEGG proportionalities.

### Microbial networks

Microbial networks for taxonomy (at the genus level) and functionality were built from the proportionality matrices described above. Input edges were defined from the `cytoscape` function in the `propr` package in R, which converts a `propr` object into a data frame of node connections compatible with Cytoscape software (v. 3.8.0). Results from the DA analyses were used to associate each feature (node) to high or low methane emissions levels. Significantly overabundant genera and KEGGs in the low methane emitters group (i.e., more abundant in LOW than in HIGH or H-MID groups) were designated as LOW-associated, while those on the contrary overabundant in high methane emitters were appointed as HIGH-associated. Non-DA features were classified as N/A (not associated). In addition, SK and MP factors were included as node attributes for genera and KEGGs, respectively. For graph visualization, Kamada-Kawai algorithm (edge-weighted spring embedded layout) was set [84], using  $\rho_p$  coefficient as force parameter.

### Data Availability

The raw data underlying this article are available at the ENA website and can be accessed with bioproject accession No. PRJNA789746 [85]. The filtered sequences in fastq format and their metadata are available from the *GigaScience* database [86]. Other data can be requested from the METALGEN project [87].

SqueezeMeta software is available at [88]. Guppy basecaller software was used to convert fast5 raw signals to fastq files [89]. The R environment and packages used are available from [90]. Correspondence and material requests should be addressed to O.G.R. Other data further supporting this work (including methane measurements) are openly available in the *GigaScience* repository, GigaDB [86].

### Additional Files

**Supplementary Figure S1. Feature counts distribution.** Features with zero counts, singletons, doubletons and 3 or more counts per sample. A) Count distribution in raw taxonomy table (87% sparsity); B) Count distribution in filtered taxonomy table (68% sparsity); C) Count distribution in raw KEGG table (72% sparsity); D) Count distribution in filtered KEGG table (39% sparsity). Filtering processes removed less than 1% of total reads in both datasets.

**Supplementary Figure S2. Phyla relative abundance per sample.** Samples are sorted from lowest to highest RA of Bacteroidetes.

**Supplementary Data S1. Differential Abundance of genera and KEGGs.** Tables containing significant differentially abundant (DA) genera and KEGGs for methane emissions levels. Contrast.AvsB column represents the CH<sub>4</sub> levels for which the feature is significantly DA. Grp.A and Grp.B columns represent the average num-

ber of reads per contrast group. For KEGGs, Association column shows if the specific KEGG participates in methanogenesis pathway (k00680) (Methane) or not (Others).

**Supplementary Data S2. Additional methods.** Detailed information about the taxonomy and prevalence filters applied to mapped reads.

### Abbreviations

AMS: automatic milking station; BLAST: Basic Local Alignment Search Tool; bp: base pairs; CLR: centered log-ratio; DA: differential abundance; FC: fold change; FDR: false discovery rate; GAM: generalized additive model; KEGG: Kyoto Encyclopedia of Genes and Genomes; LCA: last common ancestor; NADH: nicotinamide adenine dinucleotide; NCBI: National Center for Biotechnology Information; OA: overabundance; ONT: Oxford Nanopore Technologies; PCA: principal component analysis; PERMANOVA: permutational analysis of variance; RA: relative abundance.

### Ethics Statement

This study was conducted in accordance with Spanish Royal Decree 53/2013 for the protection of animals used for experimental and other scientific purposes and was approved by the Basque Institute for Agricultural Research and Development Ethics Committee (Neiker-OEBA-2017-004) on 28 March 2017.

### Competing Interests

The authors state that they have no competing interests.

### Funding

This research was financed by RTA2015-00022-C03-02 (METALGEN) project from the National Plan of Research, Development and Innovation 2013–2020 and the Department of Economic Development and Competitiveness (Madrid, Spain). A.L.G. was funded by FPI-INIA grant with reference FPI-SGIT2016-06.

### Authors' Contributions

A.L.G. and A.S.M. filtered and prepared the data, implemented the statistical analyses, and prepared the data visualization and the first draft of the manuscript. M.G.R. and C.G. performed the DNA extraction and sequencing. O.G.R. supervised the DNA sequencing and contributed to the statistical analyses. R.C. contributed to develop interaction networks. O.G.R., A.G.R., R.A., and I.G. conceived the study and designed the experiments. O.G.R., A.G.R., J.J.M., C.O. and M.S. contributed to the acquisition of funding and resources. J.T. and F.P.S. developed the computational pipelines for the metagenome and assisted on its analyses. A.L.G., A.S.M., and O.G.R. wrote the manuscript. All authors helped writing and configuring the last version of the manuscript.

### Acknowledgments

We thank the regional Holstein Associations and farmers collaborating in the project. Computational support from the High-Performance Computing Centre in Galicia (CESGA, Spain) is acknowledged.

## References

- Wang, B, Yao, M, Lv, L, et al. The human microbiota in health and disease. *Engineering* 2017;**3**(1):71–82. doi:10.1016/J.ENG.2017.01.008
- Delgado, B, Bach, A, Guasch, I, et al. Whole rumen metagenome sequencing allows classifying and predicting feed efficiency and intake levels in cattle. *Sci Rep* 2019;**9**(1): doi:10.1038/s41598-018-36673-w.
- Difford, GF, Plichta, DR, Løvendahl, P, et al. Host genetics and the rumen microbiome jointly associate with methane emissions in dairy cows. *PLoS Genet* 2018;**14**(10):e1007580. doi:10.1371/journal.pgen.1007580
- Weimer, PJ. Redundancy, resilience, and host specificity of the ruminal microbiota: Implications for engineering improved ruminal fermentations. *Front Microbiol* 2015;**6**: doi:10.3389/fmicb.2015.00296.
- Knapp, JR, Laur, GL, Vadas, PA, et al. Invited review: Enteric methane in dairy cattle production: Quantifying the opportunities and impact of reducing emissions. *J Dairy Sci* 2014;**97**(6):3231–61. doi:10.3168/JDS.2013-7234
- Gerber, PJ, Steinfeld, H, Mottet, O, et al. Tackling climate change through livestock. A global assessment of emissions and mitigation opportunities. In: Rome, Italy: Food and Agriculture Organization of the United Nations (FAO); <http://www.fao.org/3/a-i3437e.pdf>. 2013. Accessed 5 January 2022.
- Food and Agriculture Organization of the United Nations (FAO). *Global Livestock Environmental Assessment Model (GLEAM)*. <http://www.fao.org/gleam/results/en/>. 2019. Accessed 4 August 2021.
- Martínez-Álvarez, M, Auffret, MD, Stewart, RD, et al. Identification of complex rumen microbiome interaction within diverse functional niches as mechanisms affecting the variation of methane emissions in bovine. *Front Microbiol* 2020;**11**:doi:10.3389/fmicb.2020.00659.
- Malmuthuge, N, Guan, LL. Understanding host-microbial interactions in rumen: Searching the best opportunity for microbiota manipulation. *J Anim Sci Biotechnol* 2017;**8**(1): doi:10.1186/s40104-016-0135-3.
- Negussie, E, de Haas, Y, Dehareng, F, et al. Invited review: Large-scale indirect measurements for enteric methane emissions in dairy cattle: A review of proxies and their potential for use in management and breeding decisions. *J Dairy Sci* 2017;**100**(4):2433–53. doi:10.3168/jds.2016-12030
- Gloor, GB, Macklaim, JM, Pawlowsky-Glahn, V, et al. Microbiome datasets are compositional: And this is not optional. *Front Microbiol* 2017;**8**: doi:10.3389/fmicb.2017.02224.
- Liu, YX, Qin, Y, Chen, T, et al. A practical guide to amplicon and metagenomic analysis of microbiome data. *Protein Cell* 2021;**12**(5):315–30. doi:10.1007/s13238-020-00724-8
- Bonk, F, Popp, D, Harms, H, et al. PCR-based quantification of taxa-specific abundances in microbial communities: Quantifying and avoiding common pitfalls. *J Microbiol Methods* 2018;**153**:139–47. doi:10.1016/J.MIMET.2018.09.015
- Brandt, C, Bongcam-Rudloff, E, Müller, B. Abundance tracking by long-read nanopore sequencing of complex microbial communities in samples from 20 different biogas/wastewater plants. *Appl Sci* 2020;**10**(21):7518. doi:10.3390/app10217518
- Burki, F, Shalchian-Tabrizi, K, Minge, M, et al. Phylogenomics reshuffles the eukaryotic supergroups. *PLoS One* 2007;**2**(8):e790. doi:10.1371/journal.pone.0000790
- Kanehisa, M, Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res* 2000;**28**(1):27–30. doi:10.1093/nar/28.1.27
- Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019;**28**(11):1947–51. doi:10.1002/pro.3715
- Kanehisa, M, Sato, Y, Furumichi, M, et al. New approach for understanding genome variations in KEGG. *Nucleic Acids Res* 2019;**47**(D1):D590–5. doi:10.1093/nar/gky962
- Lovell, D, Pawlowsky-Glahn, V, Egozcue, JJ, et al. Proportionality: A valid alternative to correlation for relative data. *PLoS Comput Biol* 2015;**11**(3):e1004075. doi:10.1371/journal.pcbi.1004075
- Wallace, RJ, Sasson, G, Garnsworthy, PC, et al. A heritable subset of the core rumen microbiome dictates dairy cow productivity and emissions. *Sci Adv* 2019;**5**(7):eaav8391. doi:10.1126/sciadv.aav8391
- Shah, HN, Collins, DM. *Prevotella*, a new genus to include *Bacteroides melaninogenicus* and related species formerly classified in the genus *Bacteroides*. *Int J Syst Bacteriol* 1990;**40**(2):205–8. doi:10.1099/00207713-40-2-205
- Pitta, DW, Pinchak, WE, Dowd, SE, et al. Rumen bacterial diversity dynamics associated with changing from bermudagrass hay to grazed winter wheat diets. *Microb Ecol* 2010;**59**(3): 511–22. doi:10.1007/s00248-009-9609-6
- Lee, HJ, Jung, JY, Oh, YK, et al. Comparative survey of rumen microbial communities and metabolites across one caprine and three bovine groups, using bar-coded pyrosequencing and <sup>1</sup>H nuclear magnetic resonance spectroscopy. *Appl Environ Microbiol* 2012;**78**(17):5983–93. doi:10.1128/AEM.00104-12
- Lopes, LD, de Souza Lima, AO, Taketani, RG, et al. Exploring the sheep rumen microbiome for carbohydrate-active enzymes. *Antonie Van Leeuwenhoek* 2015;**108**(1):15–30. doi:10.1007/s10482-015-0459-6
- Li, J, Zhong, H, Ramayo-Caldas, Y, et al. A catalog of microbial genes from the bovine rumen unveils a specialized and diverse biomass-degrading environment. *Gigascience* 2020;**9**(6): doi:10.1093/gigascience/giaa057.
- Seshadri, R, Leahy, SC, Attwood, GT, et al. Cultivation and sequencing of rumen microbiome members from the Hungate1000 Collection. *Nat Biotechnol* 2018;**36**(4):359–67. doi:10.1038/nbt.4110
- Crespo-Piazuelo, D, Estellé, J, Revilla, M, et al. Characterization of bacterial microbiota compositions along the intestinal tract in pigs and their interactions and functions. *Sci Rep* 2018;**8**(1): doi:10.1038/s41598-018-30932-6.
- Han, GG, Lee, JY, Jin, GD, et al. Tracing of the fecal microbiota of commercial pigs at five growth stages from birth to shipment. *Sci Rep* 2018;**8**(1):6012. doi:10.1038/s41598-018-24508-7
- Ransom-Jones, E, Jones, DL, McCarthy, AJ, et al. The Fibrobacteres: An important phylum of cellulose-degrading bacteria. *Microb Ecol* 2012;**63**(2):267–81. doi:10.1007/s00248-011-9998-1
- Newbold, CJ, De la Fuente, G, Belanche, A, et al. The role of ciliate protozoa in the rumen. *Front Microbiol* 2015;**6**:1313. doi:10.3389/fmicb.2015.01313
- Williams, AG, Coleman, GS. The rumen microbial ecosystem. In: PN Hobson, GS Coleman, eds. *The Rumen Protozoa*. Dordrecht: Springer; 1997: doi:10.1007/978-94-009-1453-7\_3.
- Johnson, HP. A contribution to the morphology and biology of the Stentors. *J Morphol* 1893;**8**(3):467–562. doi:10.1002/jmor.1050080303
- Tartar, V. *The Biology of Stentor*. Oxford, London, New York, Paris: Pergamon; 1961. doi:10.1016/c2013-0-01654-4
- Wichterman, R. *The Biology of Paramecium*. New York: Springer; 1986. doi:10.1007/978-1-4757-0372-6
- Cavalier-Smith, T. The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial

- megaclassification. *Int J Syst Evol Microbiol* 2002;**52**(1):7–76. doi:10.1099/00207713-52-1-7
36. Poulsen, M, Schwab, C, Borg Jensen, B, et al. Methylophilic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat Commun* 2013;**4**(1): doi:10.1038/ncomms2432.
  37. Santos E de, O, Thompson, F. et al. The family Succinivibrionaceae. In: E Rosenberg, EF DeLong, S Lory, eds. *The Prokaryotes: Gammaproteobacteria*. Berlin, Heidelberg: Springer; 2014: doi:10.1007/978-3-642-38922-1\_368.
  38. Gomez-Gil, B, Thompson, CC, Matsumura, Y, et al. The family vibrionaceae. In: E Rosenberg, EF DeLong, S Lory, eds. *The Prokaryotes: Gammaproteobacteria*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-38922-1\_225
  39. Ivanova, EP, Ng, HJ, Webb, HK, et al. The family Pseudoalteromonadaceae. In: E Rosenberg, EF DeLong, S Lory, eds. *The Prokaryotes: Gammaproteobacteria*. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-38922-1\_229
  40. . In: Marchandin, H, Jumas-Bilak, E. Rosenberg, E, DeLong, EF, Lory, S et al. The family Veillonellaceae, eds. *The Prokaryotes: Firmicutes and Tenericutes*. Berlin, Heidelberg: Springer; 2014: doi:10.1007/978-3-642-30120-9\_361.
  41. Shah, HN, Collins, MD. Reclassification of *Bacteroides multiaacidus* (Mitsuoka, Terada, Watanabe and Uchida) in a new genus *Mitsuokella*, as *Mitsuokella multiaacidus* comb. nov. *Zentralblatt für Bakteriologie und Angewandte Mikrobiologie* 1982; doi:10.1016/S0721-9571(82)80006-9.
  42. Kittelmann, S, Pinares-Patiño, CS, Seedorf, H, et al. Two different bacterial community types are linked with the Low-Methane emission trait in sheep. *PLoS One* 2014;**9**(7): e103171. doi:10.1371/JOURNAL.PONE.0103171
  43. Wallace, RJ, Rooke, JA, Duthie, CA, et al. Archaeal abundance in post-mortem ruminal digesta may help predict methane emissions from beef cattle. *Sci Rep* 2014;**4**:5892. doi:10.1038/srep05892
  44. Morgavi, DP, Martin, C, Jouany, JP, et al. Rumen protozoa and methanogenesis: Not a simple cause-effect relationship. *Br J Nutr* 2012;**107**(3):388–97. doi:10.1017/S0007114511002935
  45. Tapio, I, Snelling, TJ, Strozzi, F, et al. The ruminal microbiome associated with methane emissions from ruminant livestock. *J Anim Sci Biotechnol* 2017;**8**(1): doi:10.1186/s40104-017-0141-0.
  46. Danielsson, R, Dicksved, J, Sun, L, et al. Methane production in dairy cows correlates with rumen methanogenic and bacterial community structure. *Front Microbiol* 2017;**8**: doi:10.3389/fmicb.2017.00226.
  47. Shi, W, Moon, CD, Leahy, SC, et al. Methane yield phenotypes linked to differential gene expression in the sheep rumen microbiome. *Genome Res* 2014;**24**(9):1517–25. doi:10.1101/gr.168245.113
  48. Wallace, RJ, Rooke, JA, McKain, N, et al. The rumen microbial metagenome associated with high methane production in cattle. *BMC Genomics* 2015;**16**(1): doi:10.1186/s12864-015-2032-0.
  49. Aguinaga Casañas, MA, Rangkasenee, N, Krattenmacher, N, et al. Methyl-coenzyme M reductase A as an indicator to estimate methane production from dairy cows. *J Dairy Sci* 2015;**98**(6):4074–83. doi:10.3168/jds.2015-9310
  50. Janssen, PH, Kirs, M. Structure of the archaeal community of the rumen. *Appl Environ Microbiol* 2008;**74**(12):3619–25. doi:10.1128/AEM.02812-07
  51. Newbold, CJ, Lassalas, B, Jouany, JP. The importance of methanogens associated with ciliate protozoa in ruminal methane production in vitro. *Lett Appl Microbiol* 1995;**21**(4):230–4. doi:10.1111/j.1472-765X.1995.tb01048.x
  52. Qin, WZ, Li, CY, Kim, JK, et al. Effects of defaunation on fermentation characteristics and methane production by rumen microbes in vitro when incubated with starchy feed sources. *Asian-Australas J Anim Sci* 2012;**25**(10):1381–8. doi:10.5713/ajas.2012.12240
  53. Schönhusen, U, Zitnan, R, Kuhla, S, et al. Effects of protozoa on methane production in rumen and hindgut of calves around time of weaning. *Arch Anim Nutr* 2003;**57**(4):279–95. doi:10.1080/00039420310001594423
  54. Guyader, J, Eugène, M, Nozière, P, et al. Influence of rumen protozoa on methane emission in ruminants: a meta-analysis approach. *Animal* 2014;**8**(11):1816–25. doi:10.1017/S1751731114001852
  55. Belanche, A, de la Fuente, G, Newbold, CJ. Study of methanogen communities associated with different rumen protozoal populations. *FEMS Microbiol Ecol* 2014;**90**(3):663–77. doi:10.1111/1574-6941.12423
  56. Hackstein, JHP, de Graaf, RM. Anaerobic ciliates and their methanogenic endosymbionts. In: JHP Hackstein, ed. *(Endo)symbiotic Methanogenic Archaea*. 2nd ed. Cham: Springer; 2018: doi:10.1007/978-3-319-98836-8\_2.
  57. Roopnarain, A, Mukhuba, M, Adeleke, R, et al. Biases during DNA extraction affect bacterial and archaeal community profile of anaerobic digestion samples. *3 Biotech* 2017;**7**(6):375. doi:10.1007/s13205-017-1009-x
  58. Esquivel-Elizondo, S, Ilhan, ZE, Garcia-Peña, EI, et al. Insights into butyrate production in a controlled fermentation system via gene predictions. *mSystems* 2017;**2**(4): doi:10.1128/msystems.00051-17.
  59. Crable, BR, Plugge, CM, McInerney, MJ, et al. Formate formation and formate conversion in biological fuels production. *Enzyme Res* 2011;**2011**:532536. doi:10.4061/2011/532536
  60. Hao, L, Michaelsen, TY, Singleton, CM, et al. Novel syntrophic bacteria in full-scale anaerobic digesters revealed by genome-centric metatranscriptomics. *ISME J* 2020;**14**(4):906–18. doi:10.1038/s41396-019-0571-0
  61. Brown, TDK, Jones-Mortimer, MC, Kornberg, HL. The enzymic interconversion of acetate and acetyl-coenzyme A in *Escherichia coli*. *J Gen Microbiol* 1977;**102**(2):327–36. doi:10.1099/00221287-102-2-327
  62. Xie, M, Chen, HH, Nie, SP, et al. Gamma-aminobutyric acid increases the production of short-chain fatty acids and decreases pH values in mouse colon. *Molecules* 2017;**22**(4):653. doi:10.3390/molecules22040653
  63. Graham, DE. 2-oxoacid metabolism in methanogenic CoM and CoB biosynthesis. *Methods Enzymol* 2011;**494**:301–26. doi:10.1016/B978-0-12-385112-3.00015-9
  64. Klüber, HD, Conrad, R. Effects of nitrate, nitrite, NO and N2O on methanogenesis and other redox processes in anoxic rice field soil. *FEMS Microbiol Ecol* 2006;**25**(3):301–18. doi:10.1111/j.1574-6941.1998.tb00482.x
  65. Yang, C, Rooke, JA, Cabeza, I, et al. Nitrate and inhibition of ruminal methanogenesis: Microbial ecology, obstacles, and opportunities for lowering methane emissions from ruminant livestock. *Front Microbiol* 2016;**7**: doi:10.3389/fmicb.2016.00132.
  66. Lee, C, Beauchemin, KA. Une revue de l'ajout de nitrate dans l'alimentation des ruminants : Toxicité aux nitrates, émissions de méthane et performance de production. *Can J Anim Sci* 2014;**94**(4):557–70. doi:10.4141/CJAS-2014-069
  67. Cui, X. The bacteriocidal effects of transition metal complexes containing the NO+ group on the food-spoilage bacterium *Clostridium sporogenes*. *FEMS Microbiol Lett* 1992;**98**(1-3):67–70. doi:10.1016/0378-1097(92)90133-9

68. Zumft, WG. The biological role of nitric oxide in bacteria. *Arch Microbiol* 1993;**160**(4):253–64. doi:10.1007/BF00292074
69. Reed, B, Dinn, J, McCann, S, et al. Alterations in mammalian cells induced by inactivation of vitamin B12 with nitrous oxide. In: B Zagalak, B Friedrich, eds. *Vitamin B12. Proceedings of the 3rd European Symposium on Vitamin B12 and Intrinsic Factor* Zürich: De Gruyter; 1979: doi:10.1515/9783111510828-120.
70. Klüber, HD, Conrad, R. Inhibitory effects of nitrate, nitrite, NO and N<sub>2</sub>O on methanogenesis by *Methanosarcina barkeri* and *Methanobacterium bryantii*. *FEMS Microbiol Ecol* 1998;**25**(4):331–9. doi:10.1111/j.1574-6941.1998.tb00484.x
71. Tamames, J, Puente-Sánchez, F. SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front Microbiol* 2019;**9**: doi:10.3389/fmicb.2018.03349.
72. Stewart, RD, Auffret, MD, Warr, A, et al. Compendium of 4,941 rumen metagenome-assembled genomes for rumen microbiome biology and enzyme discovery. *Nat Biotechnol* 2019;**37**(8):953–61. doi:10.1038/s41587-019-0202-3
73. Rey, J, Atxaerandio, R, Ruiz, R, et al. Comparison between non-invasive methane measurement techniques in cattle. *Animals* 2019;**9**(8):563. doi:10.3390/ani9080563
74. López-Paredes, J, Goiri, I, Atxaerandio, R, et al. Mitigation of greenhouse gases in dairy cattle via genetic selection: 1. Genetic parameters of direct methane using noninvasive methods and proxies of methane. *J Dairy Sci* 2020;**103**(8):7199–7209. doi:10.3168/jds.2019-17597
75. Aitchison, J. The statistical analysis of compositional data. *J R Stat Soc Ser B* 1982: doi:10.1111/j.2517-6161.1982.tb01195.x.
76. Greenacre, M. *Compositional Data Analysis in Practice*. Chapman & Hall/CRC; 2018. doi:10.1201/9780429455537
77. Palarea-Albaladejo, J, Martín-Fernández, JA. ZCompositions - R package for multivariate imputation of left-censored data under a compositional approach. *Chemometr Intell Lab Syst* 2015;**143**:85–96. doi:10.1016/j.chemolab.2015.02.019
78. Oksanen, J, Blanchet, FG, Friendly, M, et al. *vegan: Community Ecology Package*. R package version 2.5-6. <https://CRAN.R-project.org/package=vegan>. 2019. Accessed 5 January 2022
79. Anderson, MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;**26**(1):32–46. doi:10.1111/j.1442-9993.2001.01070.pp.x
80. McArdle, BH, Anderson, MJ. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* 2001;**82**(1):290–7. doi:10.1890/0012-9658(2001)082[0290:FMMTCD]2.0.CO;2
81. Ritchie, ME, Phipson, B, Wu, D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**(7):e47. doi:10.1093/nar/gkv007
82. Quinn, TP, Richardson, MF, Lovell, D, et al. Propr: An R-package for identifying proportionally abundant features using compositional data analysis. *Sci Rep* 2017;**7**(1): doi:10.1038/s41598-017-16520-0.
83. Erb, I, Notredame, C. How should we measure proportionality on relative gene expression data? *Theor Biosci* 2016;**135**(1-2): 21–36. doi:10.1007/s12064-015-0220-8
84. Kamada, T, Kawai, S. An algorithm for drawing general undirected graphs. *Inf Process Lett* 1989;**31**(1):7–15. doi:10.1016/0020-0190(89)90102-6
85. ENA project for metagenome data. <https://www.ebi.ac.uk/ena/browser/view/PRJNA789746>. Accessed 5 January 2022.
86. López-García, A, Saborío-Montero, A, Gutiérrez-Rivas, M, et al. Supporting data for “Fungal and ciliate protozoa are the main rumen microbes associated with methane emissions in dairy cattle.” *GigaScience Database* 2021. <http://dx.doi.org/10.5524/100950>.
87. METALGEN project. <https://www.metalgen.es>. Accessed 5 January 2022.
88. SqueezeMeta software. <https://github.com/jtamames/SqueezeMeta>. Accessed 5 January 2022.
89. Guppy Basecaller software. <https://community.nanoporetech.com/downloads>. Accessed 5 January 2022.
90. CRAN package repository. <https://cran.r-project.org/>. Accessed 5 January 2022.