

Spotlight Selection | Ecology | Full-Length Text

# Diversity and community structure of anaerobic gut fungi in the rumen of wild and domesticated herbivores

Casey H. Meili,<sup>1</sup> Moustafa A. TagElDein,<sup>2</sup> Adrienne L. Jones,<sup>1</sup> Christina D. Moon,<sup>3</sup> Catherine Andrews,<sup>3</sup> Michelle R. Kirk,<sup>3</sup> Peter H. Janssen,<sup>3</sup> Carl J. Yeoman,<sup>4</sup> Savannah Grace,<sup>4</sup> Joanna-Lynn C. Borgogna,<sup>4</sup> Andrew P. Foote,<sup>5</sup> Yosra I. Nagy,<sup>2</sup> Mona T. Kashef,<sup>2</sup> Aymen S. Yassin,<sup>2</sup> Mostafa S. Elshahed,<sup>1</sup> Noha H. Youssef<sup>1</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 17.

**ABSTRACT** The rumen houses a diverse community that plays a major role in the digestion process in ruminants. Anaerobic gut fungi (AGF) are key contributors to plant digestion in the rumen. Here, we present a global amplicon-based survey of the rumen AGF mycobiome by examining 206 samples from 15 animal species, 15 countries, and 6 continents. The rumen AGF mycobiome was highly diverse, with 81 out of 88 currently recognized AGF genera or candidate genera identified. However, only six genera (*Neocallimastix*, *Orpinomyces*, *Caecomyces*, *Cyllamyces*, NY9, and *Piromyces*) were present at >4% relative abundance. AGF diversity was higher in members of the families *Antilocapridae* and *Cervidae* compared to *Bovidae*. Community structure analysis identified a pattern of phylosymbiosis, where host family (10% of total variance) and species (13.5%) partially explained the rumen mycobiome composition. As well, diet composition (9%–19%), domestication (11.14%), and biogeography (14.1%) also partially explained AGF community structure; although sampling limitation, geographic range restrictions, and direct association between different factors hindered accurate elucidation of the relative contribution of each factor. Pairwise comparison of rumen and fecal samples obtained from the same subject ( $n = 13$ ) demonstrated greater diversity and inter-sample variability in rumen versus fecal samples. The genera *Neocallimastix* and *Orpinomyces* were present in higher abundance in rumen samples, while *Cyllamyces* and *Caecomyces* were enriched in fecal samples. Comparative analysis of global rumen and feces data sets revealed a similar pattern. Our results provide a global view of AGF community in the rumen and identify patterns of AGF variability between rumen and feces in herbivores Gastrointestinal (GI) tract.

**IMPORTANCE** Ruminants are highly successful and economically important mammalian suborder. Ruminants are herbivores that digest plant material with the aid of microorganisms residing in their GI tract. In ruminants, the rumen compartment represents the most important location where microbially mediated plant digestion occurs, and is known to house a bewildering array of microbial diversity. An important component of the rumen microbiome is the anaerobic gut fungi (AGF), members of the phylum *Neocallimastigomycota*. So far, studies examining AGF diversity have mostly employed fecal samples, and little is currently known regarding the identity of AGF residing in the rumen compartment, factors that impact the observed patterns of diversity and community structure of AGF in the rumen, and how AGF communities in the rumen compare to AGF communities in feces. Here, we examined the rumen AGF diversity using an amplicon-based survey targeting a wide range of wild and domesticated ruminants ( $n = 206$ , 15 different animal species) obtained from 15 different countries. Our results demonstrate that while highly diverse, no new AGF genera were identified in the rumen mycobiome samples examined. Our analysis also indicate that animal host phylogeny, diet, biogeography, and domestication status could play a role in shaping

**Editor** Gemma Reguera, Michigan State University, USA

Address correspondence to Noha H. Youssef, noha@okstate.edu.

The authors declare no conflict of interest.

See the funding table on p. 17.

**Received** 29 August 2023

**Accepted** 5 January 2024

**Published** 1 February 2024

Copyright © 2024 American Society for Microbiology. All Rights Reserved.

AGF community structure. Finally, we demonstrate that a greater level of diversity and higher inter-sample variability was observed in rumen compared to fecal samples, with two genera (*Neocallimastix* and *Orpinomyces*) present in higher abundance in rumen samples, and two others (*Cyllumyces* and *Caecomyces*) enriched in fecal samples. Our results provide a global view of the identity, diversity, and community structure of AGF in ruminants, elucidate factors impacting diversity and community structure of the rumen mycobiome, and identify patterns of AGF community variability between the rumen and feces in the herbivorous GI tract.

**KEYWORDS** anaerobic fungi, rumen, phylosymbiosis

Ruminants (suborder *Ruminantia*) are one of the most diverse and prevalent groups of extant mammalian herbivores. The global population of domesticated ruminants is estimated at ~3.75 billion, and that of wild ruminants is upward of 75 million animals (1). Suborder *Ruminantia* includes six families: *Antilocapridae*, *Bovidae*, *Cervidae*, *Giraffidae*, *Moschidae*, and *Tragulidae* (2). The most species-rich family is *Bovidae* with >140 species (3), many of which are important livestock animals (e.g., cattle, goats, sheep) (1).

Ruminants are highly efficient in digesting high-fiber feeds and forages. This is primarily due to the ability to ferment feed in an anaerobic pregastric chamber (the rumen) which enables a specialized microbiome-mediated plant biomass degradation and fermentation to end products that are important energy sources for the host (4). Rumination, the process by which animals regurgitate and masticate previously swallowed plant material, allows further physical breakdown of partially digested feed, enhancing rumen microbial activity. As well, the large quantity of saliva produced by ruminants aids in buffering the rumen environment, hence stabilizing and promoting microbial activity (5).

The rumen microbial community encompasses bacteria, archaea, protozoa, viruses, and fungi (6, 7). The fungal component, the anaerobic gut fungi (AGF), belongs to the phylum *Neocallimastigomycota*. They play a key role in the plant biomass degradation process. AGF hyphae efficiently penetrate plant biomass, resulting in the efficient mechanical disruption of plant cell walls (6). As well, AGF produce a wide array of carbohydrate-active enzymes that are crucial for plant cell wall degradation (8–10). Indeed, several field studies have demonstrated the important contributions of AGF to biomass degradation in ruminants (9, 11, 12).

Surprisingly, in contrast to the wealth of information on the rumen bacterial and archaeal communities, little information is currently available on the resident rumen AGF community in the rumen of mammalian herbivores. The bulk of culture-independent and culture-based AGF characterization studies has been conducted on fecal, rather than rumen, samples due to the relative ease of sampling (13–16), and only a few studies to date have reported on AGF communities in rumen samples, all of which were limited in scope, examining few subjects and host species (17–22). AGF communities residing in the rumen could differ from those encountered in fecal samples due to possible selection and modification when passing through various partitions of the forestomach system (rumen, omasum, and abomasum) owing to the large difference in pH between these compartments (e.g., 6–6.4 in the rumen, 5.5–6.5 in the omasum, and 1.5–3 in the abomasum in cattle) before reaching the circumneutral small intestine. In addition, a fraction of fermentation occurs in the intestine [around 10% (23)]; and the potential AGF presence, origin, identity, load, and relative contribution to the AGF community encountered in fecal samples is currently unknown. Finally, distinct bacterial and archaeal communities colonize various locations in the GI tract of ruminants (24, 25), and these communities could differentially impact and modulate AGF diversity, load, and community composition through antagonistic, synergistic, or mutualistic relationships, as suggested in defined cocultures (26).

Because the rumen is the main site for feed digestion and absorption, studying the rumen AGF community provides insights into the taxa involved in active plant

biomass degradation in ruminants. A detailed understanding of the AGF diversity and community structure in the rumen is key for devising strategies for community modulation, manipulation, and augmentation to improve the host's overall health and feed efficiency. As well, the yet unexamined rumen could represent a source for novel, hitherto uncharacterized AGF taxa that are selectively lost during feed passage from the rumen to the lower GI tract. To fill this knowledge gap, we characterized the AGF communities in a global collection of rumen samples ( $n = 206$ ) belonging to 15 different species from three ruminant families using a culture-independent amplicon sequencing approach. In addition to the data set's broad host and geographic distribution, enabling biogeographic-based comparisons, it also allows comparison of domesticated ( $n = 180$ ) to wild ( $n = 26$ ) hosts, providing a unique opportunity to examine the effect of domestication on the rumen AGF community. Availability of diet metadata for a subset of hosts also allowed studying the effect of diet composition on the rumen AGF community. Furthermore, the rumen AGF community was compared to fecal AGF data sets obtained in a recent similar global survey. Finally, a direct pairwise comparison of AGF communities in the rumen and fecal samples simultaneously obtained from the same animal was conducted on a subset of animals. Our results provide a global view of the identity, diversity, and community structure patterns of AGF in the mammalian rumen and elucidate the role of phylogeny, diet, biogeography, and domestication in structuring AGF communities. Furthermore, rumen versus feces community comparisons suggest that while similar ecological and evolutionary factors impact the AGF community in both locations, distinct differences in the identity, diversity, and community structure patterns exist between both locations. We posit that such differences are driven by AGF acquisition routes, and the selection process associated with the transition of AGF from the pregastric rumen to the intestine.

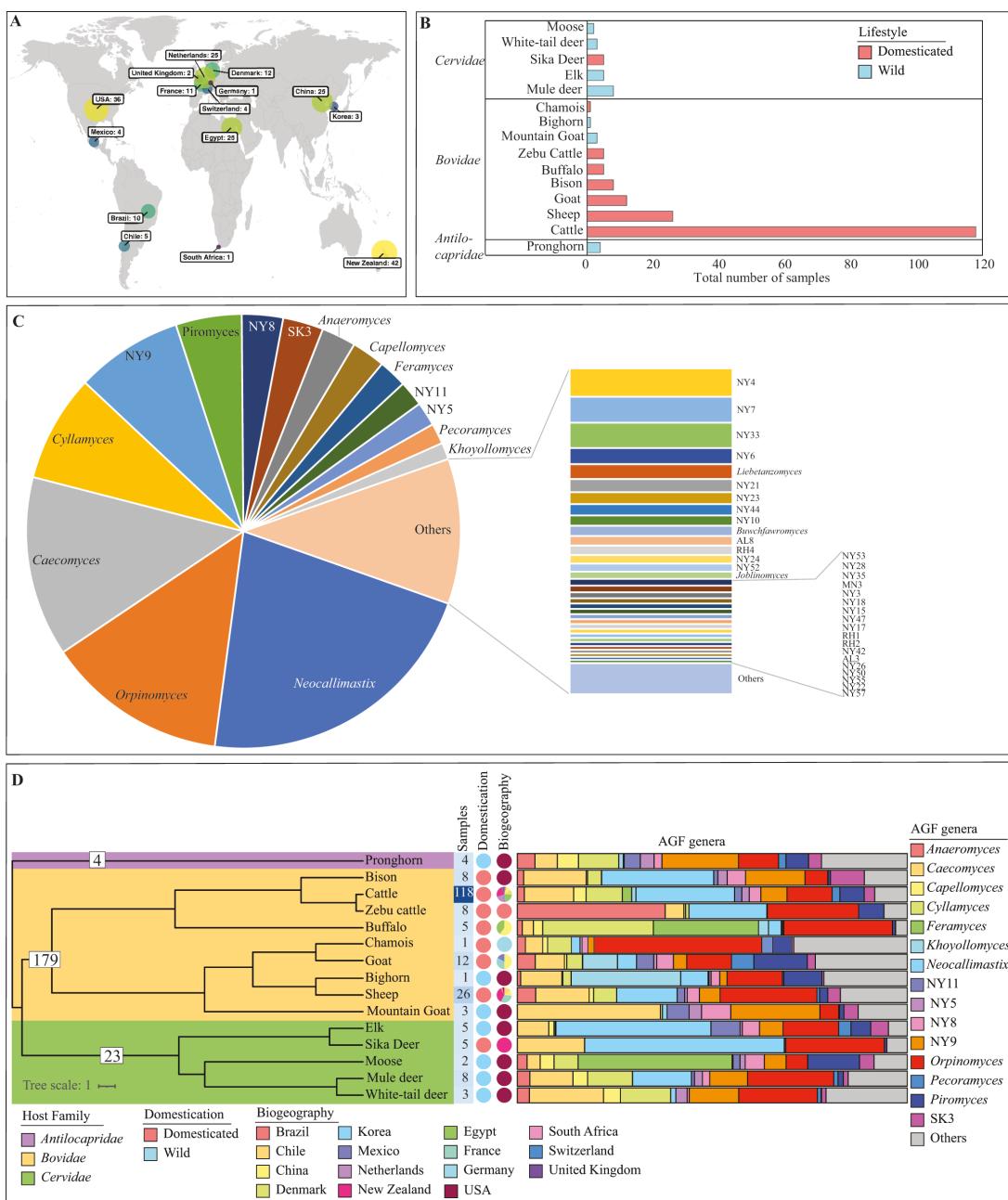
## MATERIALS AND METHODS

### Samples

A total of 206 rumen samples from 15 countries and 6 continents were obtained (Table S1; Fig. 1A). Samples include representatives from three ruminant families (*Antilocapridae*, *Bovidae*, and *Cervidae*) and 15 different domesticated and wild animal species (Fig. 1B; Table S1). Many of these samples were obtained as part of a prior global rumen census (GRC) survey of the bacterial, archaeal, and protozoal communities in the rumen (Table S1) and handled as described in the original publication (27). GRC samples used in this project were shipped from New Zealand (where they were stored long-term at  $-80^{\circ}\text{C}$ ) to Oklahoma State University. These samples were stabilized using GenTegra-DNA protectant (GenTegra, Pleasanton, CA, USA) to minimize DNA degradation during shipping from.

In addition to GRC samples, other samples were obtained from wild ruminants through collaboration with hunters at the state of Montana ( $n = 27$ ) (Table S1). Samples were stored at  $-20^{\circ}\text{C}$  and shipped to Oklahoma State University (Stillwater, OK). A third fraction of samples ( $n = 25$ ) were collected from three slaughterhouses in Egypt (Cairo, Giza, and Menya Governorates) with the help of trained technicians. The GI tracts of the slaughtered animals were separated on a clean bench and then sectioned with a knife. Rumen contents were transferred in sterile labeled 50 mL Falcon tubes. The interval between animal death and sample collection did not exceed 30 min. The Falcon tubes were stored at  $-20^{\circ}\text{C}$  until DNA extraction (Table S1). Finally, samples were also obtained from an animal housed at the Oklahoma State University Department of Animal Sciences through gastric tube insertion (Table S1) and stored at  $-20^{\circ}\text{C}$  until DNA extraction. All animal ethics approvals for rumen sampling from the GRC survey were obtained as outlined in reference (27). All hunters in Montana had the necessary hunting permits and harvested animals using legal methods.

Furthermore, for a subset of animals (12 cattle and 1 water buffalo, red text in Table S1), fecal samples were obtained at the same time as the ruminal contents were sampled.



**FIG 1** AGF diversity patterns in the rumen mycobiome. (A) A map showing the biogeographical origin of all samples ( $n = 206$ ) analyzed in this study. (The map was created in R using the packages "maps," "mapproj," and "ggplot2.") (B) Bar plots showing the number of samples belonging to each animal species. Animals are ordered by family. Domestication status is color-coded (domesticated in peach and wild in cyan). (C) Pie chart showing the overall composition of AGF genera encountered in the entire (1.86 million sequence) data set. Genera present in >1% relative abundance are named on the pie chart, genera with relative abundances 0.5%–1% are named in the bar chart to the right, and genera present in <0.5% relative abundance are collectively referred to as "Others." (D) AGF community composition in the animal species studied. The phylogenetic tree downloaded from timetree.org shows the relationship between the 15 species sampled. Species are color-coded by their family and the total number of samples belonging to each of the three families is shown at each corresponding node. "Samples" refer to the number of samples belonging to each animal species and is shown to the right of the tree as a heatmap with the exact numbers displayed. "Domestication" refers to the domestication status (domesticated in peach and wild in cyan) and is shown as a pie chart to the right of the heatmap. "Biogeography" shows the distribution of the number of samples from different geographical regions and is shown as a pie chart to the right of "Domestication." Countries are color-coded as shown in the figure key. The AGF community composition for each animal species is shown to the right as colored columns corresponding to the legend key. Genera with a total abundance of >1% are shown, while all other genera are grouped as "Others."

For one cattle subject (housed at the Oklahoma State Animal Sciences Department), rumen samples were obtained via gastric tubing, as described above, and the first fecal sample deposited after rumen collection was obtained. For the remaining animals (11 cattle and 1 water buffalo), subjects were slaughtered as part of the slaughterhouse operations, and rumen and fecal samples were directly obtained post-slaughter.

### DNA extraction, amplification, and sequencing

DNA from the GRC rumen samples were extracted from whole (i.e., solid and liquid) samples as described previously (27). For all other rumen samples, DNA was extracted from whole samples using the DNeasy Plant Pro Kit (Qiagen, Germantown, MD, USA) according to the manufacturer's instructions. PCR amplification reactions, amplicon clean-up, quantification, index and adaptor ligation, and multiplexing were conducted in a single laboratory (Oklahoma State University, Stillwater, OK, USA) to eliminate inter-laboratory variability. The procedure, previously outlined in detail in reference (15), involved amplification of a ~370 bp of the second variable region of the large ribosomal subunit (D2-LSU) using primers AGF-LSU-EnvS primer pair (AGF-LSU-EnvS For: 5'-GCGTTTRCACCASTGTTGTT-3', AGF-LSU-EnvS Rev: 5'-GTCAACATCCTAA GYGTAGGTA-3'). This primer pair selectively targets the phylum *Neocallimastigomycota* to avoid unwanted targeting of fungi ingested with feed material. The specificity and coverage of this primer pair have been thoroughly evaluated in reference (16). Indeed, no sequences associated with phyla other than the *Neocallimastigomycota* were obtained in any data set examined in this study. Pooled libraries were sequenced at the University of Oklahoma Clinical Genomics Facility (Oklahoma City, OK, USA) using an Illumina MiSeq platform as previously described (15). DNA extraction, amplification, and sequencing of fecal samples were conducted following the same procedures for rumen samples outlined above.

### Sequence processing and phylogenetic placement

Protocols for read assembly and sequence quality trimming as well as procedures for calculating thresholds for species and genus delineation and genus-level assignments were conducted as described previously (15). Assignment of sequences to AGF genera was conducted using a two-tier approach for genus-level phylogenetic placement and thresholds as described previously (13, 15, 28).

### Diversity and community structure assessment

The relationship between host identity and AGF diversity and community structure was examined across animal host families and species for animals with at least four samples at each of these levels. This included the three animal host families, and the animal species pronghorn ( $n = 4$ ), cattle ( $n = 116$ ), sheep ( $n = 26$ ), goat ( $n = 14$ ), American bison ( $n = 8$ ), water buffalo ( $n = 5$ ), zebu cattle ( $n = 5$ ), mule deer ( $n = 8$ ), sika deer ( $n = 5$ ), and elk ( $n = 5$ ).

The effect of biogeography was examined by clustering samples by country and only including countries with at least four samples. This included samples from New Zealand ( $n = 42$ ), USA ( $n = 36$ ), China ( $n = 25$ ), Egypt ( $n = 25$ ), Netherlands ( $n = 25$ ), Denmark ( $n = 12$ ), France ( $n = 11$ ), Brazil ( $n = 10$ ), Chile ( $n = 5$ ), Mexico ( $n = 4$ ), and Switzerland ( $n = 4$ ). However, due to the uneven distribution of animals across locations, we also re-analyzed the effects of biogeography across the same animal species. Only cattle and sheep were considered for this comparison as they had enough representation (at least four samples) across different countries (Brazil, China, Denmark, Egypt, Netherlands, New Zealand, and Switzerland for cattle, and Chile, France, and New Zealand for sheep).

The effect of domestication status was also examined by clustering animals into wild or domesticated categories. However, domestication status could also be conflated with host phylogeny due to unequal representation of wild animals from the families *Cervidae* and *Antilocapridae*, and domesticated animals in family *Bovidae*. Domestication status could also be conflated with biogeography, with all wild animals originating from

samples obtained in the USA. To partially alleviate this issue, we also examined the effect of domestication status within members of the same family, comparing wild bighorn and mountain goat ( $n = 4$ ) versus all other domesticated *Bovidae* species, and domesticated sika deer ( $n = 5$ ) versus all other wild *Cervidae* species. Finally, domesticated versus wild comparison of members of the same animal genus was also conducted where possible [for domesticated members of the genus *Cervus* (sika deer, *Cervus nippon*,  $n = 5$ ) versus wild (elk, *Cervus canadensis*,  $n = 5$ )].

We also investigated the effect of diet on AGF diversity and community structure. The exact diet composition of wild ruminants is difficult to obtain. However, diet metadata (forage diet details, forage diet category, forage plant detail, and forage plant category) were available for a subset of domesticated animal hosts ( $n = 135$ ; purple text in Table S1) originating from cattle ( $n = 84$ ), sheep ( $n = 18$ ), goat ( $n = 12$ ), American bison ( $n = 8$ ), zebu cattle and sika deer ( $n = 5$  each), and water buffalo ( $n = 3$ ). This reduced data set was used for examining the effect of diet composition on AGF alpha diversity and community structure. In addition, to limit the effect of the host species on the obtained results, we also performed the analyses on only the 84 cattle samples.

Alpha diversity indices (Shannon, Simpson, and Inverse Simpson) were calculated using the estimate\_richness command in the phyloseq R package (29), and the effect of factors (host family, host species, domestication status, and biogeography) on alpha diversity was calculated using analysis of variance (ANOVA) (run through the aov command in R). The TukeyHSD command in R was used for multiple comparisons of means on significant ANOVA results for all pairwise comparisons.

For community structure analysis, weighted Unifrac was calculated using the distance command in the phyloseq R package. Pairwise values were used to construct principal coordinate analysis (PCoA) ordination plots using the commands ordinate and plot\_ordination in phyloseq R package. To elucidate factors significantly impacting community structure, PERmutational Multivariate ANalysis of VAriance (PERMANOVA) tests were run using the command adonis2 in the vegan R package (30). Percentage variance explained by each factor was calculated as the percentage of the sum of squares of each factor to the total sum of squares, and F-statistics P-values were used to examine the significance of the effect.

Multiple regression of matrices (MRM), Mantel tests for matrices correlations, and Procrustes rotation were also utilized to further quantify factors that could explain the divergence in AGF communities. MRM and Mantel tests were conducted by comparing a Gower-transformed matrix of each host factor (host family, species, domestication status, and biogeography) to the weighted Unifrac beta diversity dissimilarity matrix (calculated as detailed above) using the MRM and Mantel commands in the ecodist R package (31). Gower transformation of host factor matrices was conducted using the daisy command in the cluster R package (32). The protest command in the vegan R package was utilized for Procrustes rotation calculations. P-values and coefficients ( $R^2$  regression coefficients from MRM analysis, Spearman correlation coefficients from Mantel tests, and symmetric orthogonal Procrustes statistic from Procrustes analysis) were examined to determine the significance and importance of factors, respectively, in shaping the AGF community.

To identify specific animal host-fungal associations, LIPA (Local Indicator of Phylogenetic Association) was employed using the lipaMoran command in the phylosignal R package (33). For genera with significant associations ( $P$ -value  $<0.05$ ), we calculated the average LIPA value for each animal species. Only genera with  $>1\%$  relative abundance in the entire data set ( $n = 15$ ) were examined. We considered average LIPA values in the range of 0.2–0.4 to represent weak associations, in the range of 0.4–1 to represent moderate associations, and above 1 to represent strong associations.

### AGF diversity in rumen versus fecal samples

We compared AGF diversity and community structure in a subset of animals (12 cattle and one water buffalo; red text in Table S1) in which fecal samples were obtained and processed at the same time as rumen samples as described above. As well, we sought to

evaluate whether the observed patterns from pairwise comparison of samples obtained from the same animal could be extrapolated to larger data sets where fecal and rumen samples were obtained from different animals. To this end, we compared the community structure of cattle rumen ( $n = 116$ ) and fecal ( $n = 178$ ) AGF communities using all cattle rumen samples obtained in this study and cattle fecal samples obtained in a recent global survey of the AGF mycobiome (15).

For both rumen versus feces data sets obtained from the same animal ( $n = 26$ ) and global cattle rumen versus feces ( $n = 294$ ), double principal coordinate analysis (DPCoA) plots were calculated using the plot\_ordination command in the phyloseq R package. In addition, Metastats (34) in mothur, and MaAsLin2 in R (35) were used to identify genera differentially abundant in rumen versus feces samples.

## RESULTS

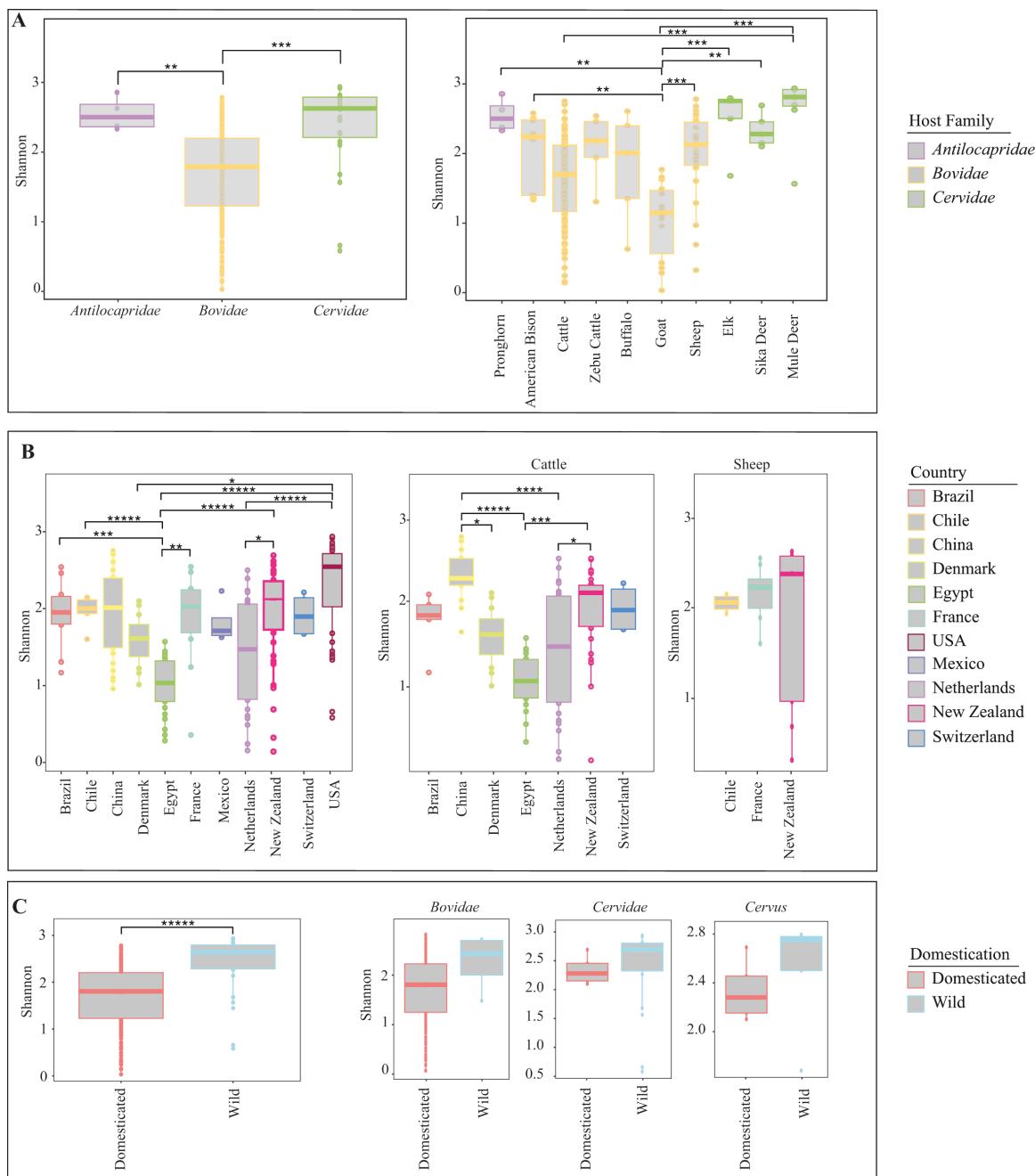
### Rumen AGF community overview

Illumina sequencing of 206 different rumen samples generated 1.86 million (average = 9,029) high-quality AGF-affiliated D2-LSU sequences (Table S2). High coverage values (average 0.996, minimum 0.92, coverage higher than 0.98 in 197/206 samples) indicated that the majority of genus-level diversity was captured in all samples (Table S2). Phylogenetic analysis identified 81 of the 88 AGF genera currently described (22/22 cultured; 59/66 uncultured) (Table S2; Fig. 1C and D), and no new AGF genera were identified in the data set. While the majority of currently recognized AGF genera and candidate genera were encountered, only 15 genera were present at >1% abundance in the entire data set (Fig. 1C), and only six genera (*Neocallimastix*, *Caecomyces*, *Orpinomyces*, *Cyllamyces*, NY9, and *Piromyces*) were present at >4% abundance (Fig. 1C). Relative abundance and occurrence for AGF genera were highly correlated ( $R^2 = 0.571$ , Fig. S1).

### Patterns of alpha diversity in the rumen mycobiome

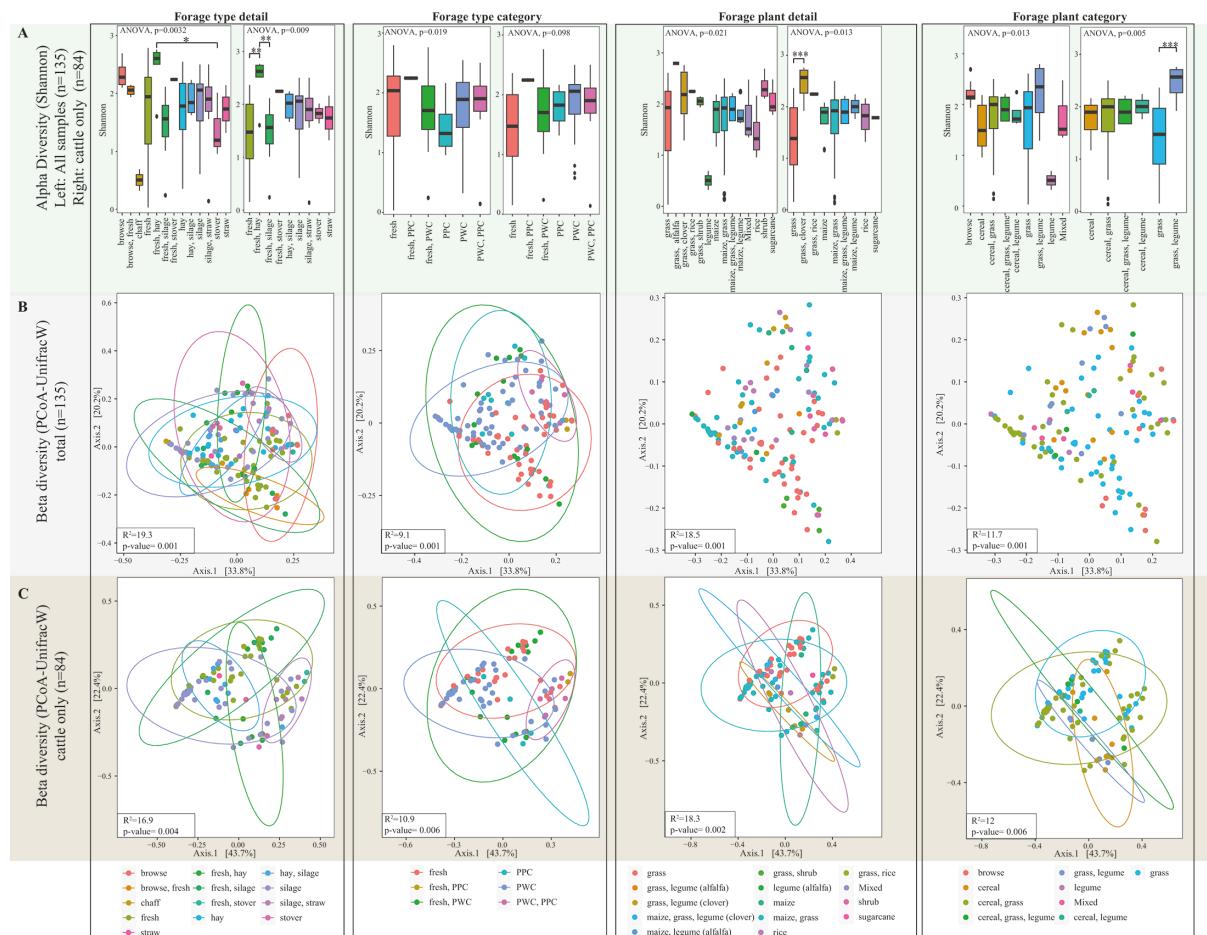
Alpha diversity patterns were assessed using three different indices: Shannon (Fig. 2), Simpson, and Inverse Simpson (Fig. S2). Collectively, samples belonging to the family *Bovidae* were significantly less diverse when compared to members of the families *Cervidae* (ANOVA  $P$ -value <0.03, Fig. 2A) and *Antilocapridae* (ANOVA  $P$ -value <0.05 (Fig. S2a and b). Within specific animal species, the AGF community in the rumen of pronghorn (family *Antilocapridae*), mule deer, and elk (family *Cervidae*) were the most diverse (ANOVA  $P$ -value <0.002), while goat (family *Bovidae*) harbored the least diverse community (ANOVA  $P$ -value <0.03). Pairwise differences in estimates of AGF community alpha diversity were significant between host species belonging to different families, e.g., pronghorn (*Antilocapridae*) versus goat (*Bovidae*) (ANOVA  $P$ -value = 0.002), mule deer (*Cervidae*) versus goat (ANOVA  $P$ -value = 9E-07) and cattle (ANOVA  $P$ -value = 0.0005) (*Bovidae*), and sika deer (*Cervidae*) versus goat (*Bovidae*) (ANOVA  $P$ -value = 0.006), as well as within few pairs of species in the family *Bovidae*, e.g., American Bison versus goat (ANOVA  $P$ -value = 0.03) and sheep versus goat (ANOVA  $P$ -value = 0.0003) (Fig. 2A; Fig. S2a and b).

In addition to host identity, multiple pairwise significant differences were observed between alpha diversity patterns of samples when grouped by the country of origin, with Egypt samples showing significantly lower diversity patterns (ANOVA  $P$ -value <0.05) (Fig. 2B; Fig. S2c and d). However, since biogeographic patterns could be a reflection of unequal distribution of hosts species across locations as described above, we also examined differences in alpha diversity between the same animal species from different countries. Only cattle and sheep had adequate samples (at least four) across different countries (Brazil, China, Denmark, Egypt, Netherlands, New Zealand, and Switzerland for cattle, and Chile, France, and New Zealand for sheep) to enable such analysis. We identified significant differences in alpha diversity based on country of origin in cattle (ANOVA  $P$ -value <0.05), but not sheep (Fig. 2B; Fig. S2c and d). Finally, a comparison of alpha diversity estimates between samples from domesticated versus wild animals



**FIG 2** AGF alpha diversity in the rumen mycobiome. (A) Box and whisker plots showing the distribution of Shannon diversity measure for different animal families (left), and animal species (right) with four or more samples. (B) Box and whisker plots showing the distribution of Shannon diversity measure for animals from different biogeographical locations (only countries with at least four samples are shown). Results are shown for the total data set (left), only cattle (middle), or only sheep (right). (C) Box and whisker plots showing the distribution of Shannon diversity measure for domesticated versus wild animals. Results are shown for the total data set (left), and for animals belonging to the families *Bovidae*, *Cervidae*, or the genus *Cervus* as depicted on top of each figure. Results for two-tailed ANOVA followed by Tukey for pairwise animal family and animal species comparisons are shown on top of the box plots only for significant comparisons. \*,  $0.01 < P < .05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ; \*\*\*\*\*,  $P < 0.00001$ .

revealed higher levels of diversity in wild compared to domesticated hosts (ANOVA  $P$ -value  $<0.006$ ) (Fig. 2C; Fig. S2e and f), although such differences were not significant when restricting the analysis to animal hosts within the same host family or genus (Fig. 2C; Fig. S2e and f).



**FIG 3** Effect of diet on rumen AGF diversity and community structure. (A) Alpha diversity (Shannon) comparisons shown as box and whisker plots for the distribution of Shannon diversity measure between different forage type details (far left), forage type category (second left), forage plant detail (second right), and forage plant category (far right). ANOVA  $P$ -value is shown for each diet composition comparison. Where applicable, results of Tukey test for pairwise comparisons are shown on top of the box plots only for significant comparisons. \*,  $0.01 < P < 0.05$ ; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ ; \*\*\*\*\*,  $P < 0.00001$ . (B–C) Effect of diet composition on AGF community structure in a data set of 135 hosts for which diet metadata were available (B) and for a subset of all cattle hosts ( $n = 84$ ) (C). PCoA ordination plots in (B) and (C) are based on AGF community structure using the phylogenetic similarity-based index weighted Unifrac. Samples are color-coded by the different diet composition categories (as shown in the color legend at the bottom of the figure) as follows: forage type details (far left), forage type category (second left), forage plant detail (second right), and forage plant category (far right). The % variance explained by the first two axes is displayed on the axes, and ellipses encompassing 95% of variance are displayed [except for the forage plant detail and forage plant category in (B) for ease of visualization]. PERMANOVA results for partitioning the dissimilarity by diet composition are shown at the bottom left corner of each ordination plot.  $R^2$  refers to percentage variance explained by each factor (calculated as the percentage of the sum of squares of each factor to the total sum of squares), while  $P$ -value refers to the F-statistics  $P$ -value.

Diet also showed a significant effect on alpha diversity measures (ANOVA  $P$ -value  $<0.03$ ) (Fig. 3A; Fig. S3a). Forage type detail (ANOVA  $P$ -value  $<0.008$ ) and forage plant detail (ANOVA  $P$ -value  $<0.04$ ) significantly affected alpha diversity. However, pairwise comparisons (using TukeyHSD) failed to identify pairs of forage types or forage plants responsible for such differences, mainly due to the unbalanced distribution of forage types or forage plants, with some categories (e.g., chaff, straw, alfalfa) only encountered a few times ( $<3$ ), while others (e.g., fresh, hay, silage, grass) were more abundant. When restricting the analysis to cattle, a similar pattern was observed, where forage type detail (ANOVA  $P$ -value  $<0.03$ ) and forage plant detail [only significant with Shannon index (ANOVA  $P$ -value = 0.013)] significantly affected alpha diversity (Fig. 3A; Fig. S3b).

## Rumen mycobiome community structure

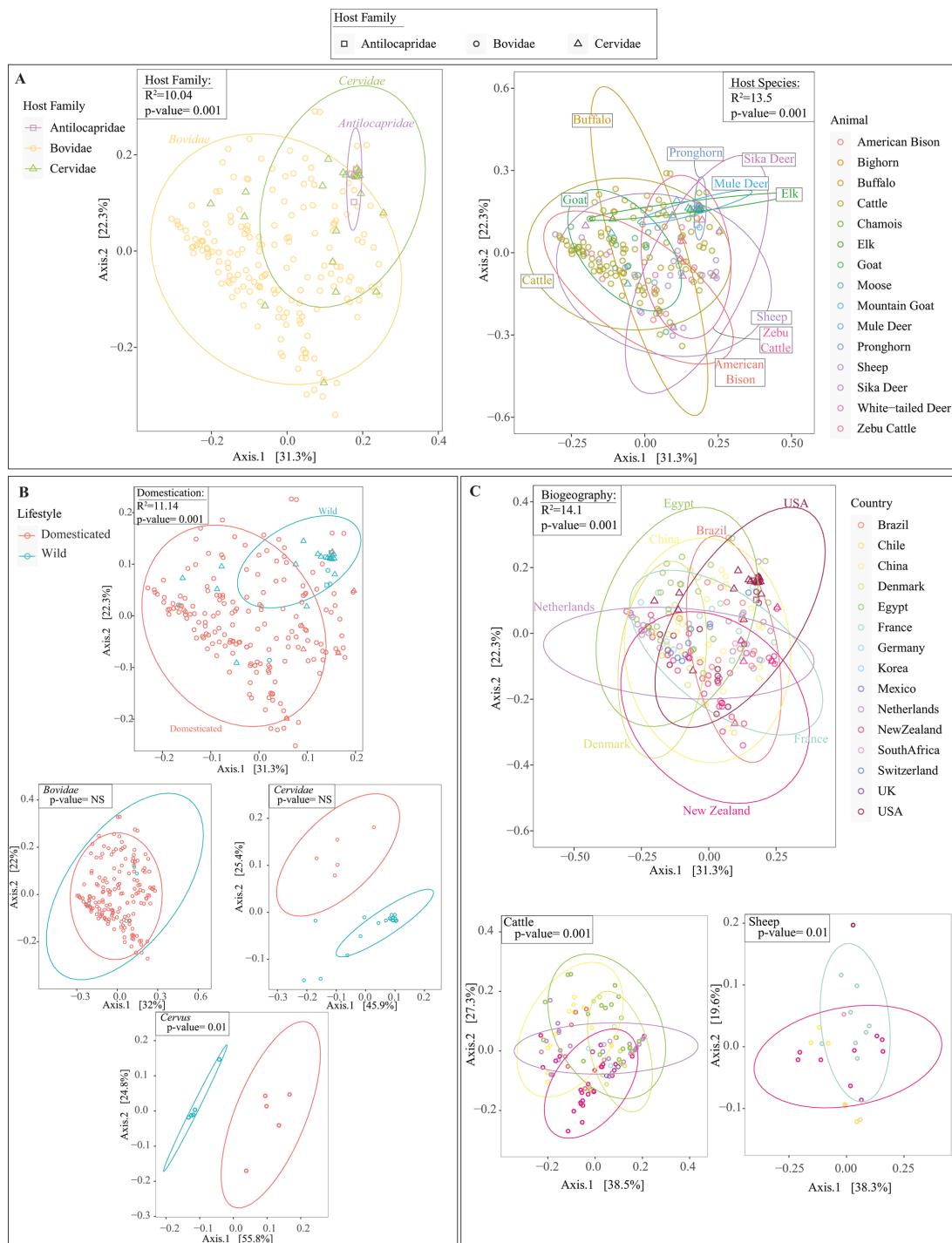
AGF community structure analysis using PERMANOVA indicated a significant role for host phylogeny in shaping AGF rumen mycobiome at the family ( $P = 0.001$ ) and the species ( $P = 0.001$ ) levels, although such factors explained only 10.0% and 13.5% of variance, respectively (Fig. 4A). Similarly, domestication status was a significant factor ( $P = 0.001$ ) in explaining 11.1% of the AGF rumen mycobiome variance when using the entire data set (Fig. 4B), as well as when subsetting samples belonging to the genus *Cervus* ( $P = 0.01$ ), but not when subsetting samples belonging to the families *Bovidae* and *Cervidae* (Fig. 4B). Finally, biogeography was also significantly associated with the AGF rumen mycobiome community, explaining 14.1% of the community variance using the entire data set level (Fig. 4C). A significant effect of biogeography on AGF community structure was also observed when restricting the analysis to a single host species (cattle and sheep, Fig. 4C). Diet (forage type detail, forage type category, forage plant detail, and forage plant category) played a significant ( $P$ -value = 0.001) role in shaping AGF community structure, explaining 9.1%–19.3% of variance using the 135 samples for which diet metadata were available (Fig. 3B). A similar significant ( $P < 0.01$ ) effect of diet was also observed when subsetting samples originating from cattle and for which diet metadata were available ( $n = 84$ ), with forage type detail, forage type category, forage plant detail, and forage plant category explaining 10.9%–18.3% of AGF community variance (Fig. 3C).

In addition to PERMANOVA, MRM, Mantel tests for matrices correlations, and Procrustes rotation were utilized to quantify factors that could explain the divergence in AGF communities. Results of matrices correlation using each of the three methods confirmed the importance of animal host species, family, biogeography, and domestication status in explaining the AGF community structure (Fig. S4). Finally, to identify whether specific AGF genera are associated with specific animal hosts and to quantify the strength of such associations, we employed LIPA analysis. For the 15 genera encountered at >1% abundance, LIPA analysis identified 17 (5 weak, 4 moderate, and 8 strong) AGF genus-animal host associations (Fig. S5).

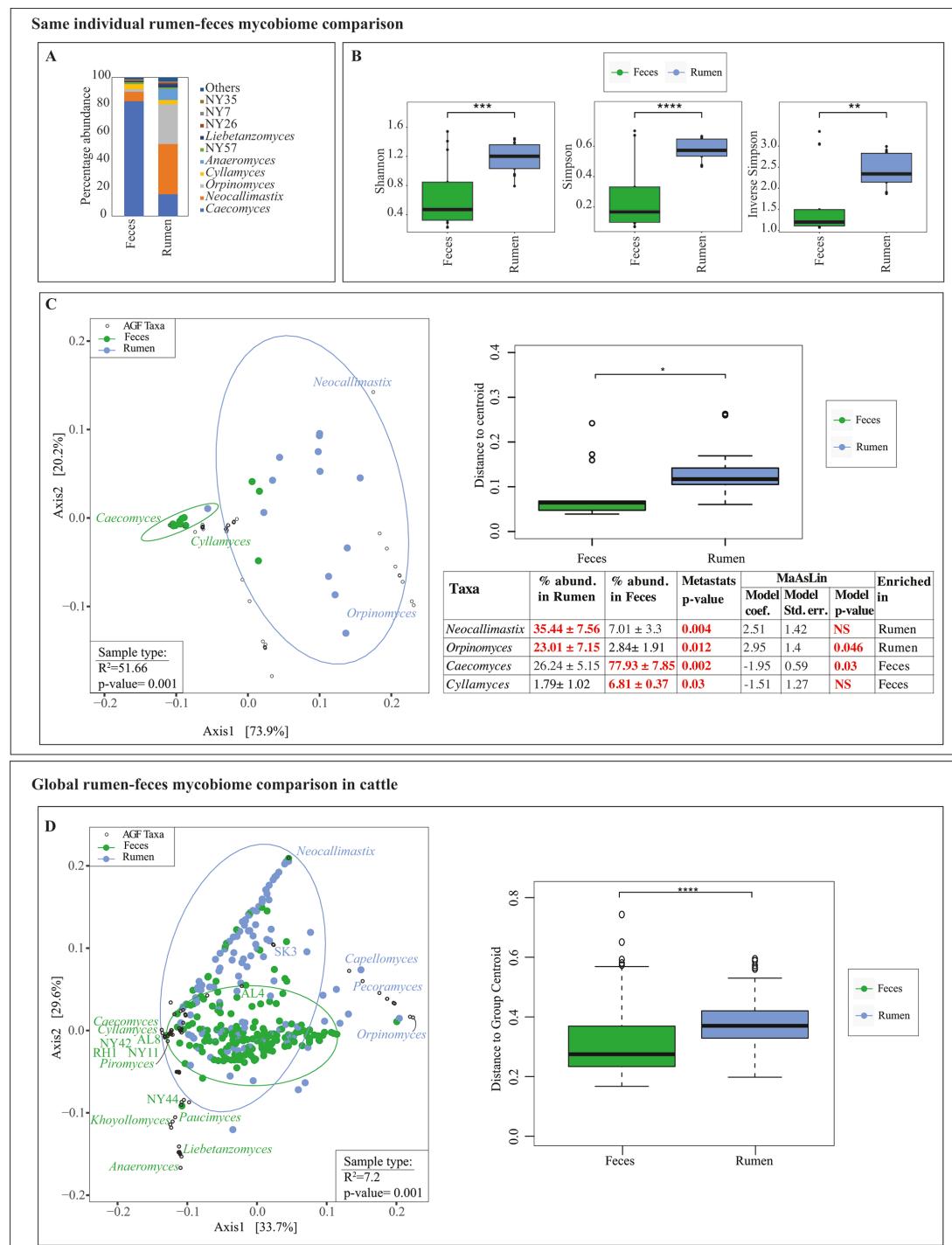
## Rumen-feces mycobiome comparison

Direct comparisons were made of rumen and fecal samples obtained simultaneously from 12 cows and 1 water buffalo sample. Within this relatively limited data set, clear differences were observed in the AGF community composition (Fig. 5A), alpha diversity (Fig. 5B), and community structure (Fig. 5C). AGF communities in fecal samples were significantly less diverse than those from rumen samples (ANOVA  $P$ -value <0.0007) (Fig. 5B). DPCoA ordination plots constructed using weighted Unifrac indices demonstrated clear clustering of rumen and fecal communities, with sampling location (rumen versus feces) explaining 51.66% of community variance. The level of variability within each sampling location was quantified by measuring the variability in Euclidean distances of samples from each sampling location to their corresponding group centroid (the centroid of the 95% ellipses shown in Fig. 5C). A significantly greater level of variability (Student's  $t$ -test  $P$ -value = 0.01) was observed in rumen versus feces samples. Furthermore, DPCoA showed selective enrichment of specific AGF taxa for each sampling location with the genera *Neocallimastix* and *Orpinomyces* selectively enriched in the rumen, and the genera *Caecomyces* and *Cyathomyces* selectively enriched in fecal samples. Metastats and MaAsLin2 confirmed significance of these specific genera selective enrichment (Fig. 5C).

We sought to evaluate whether the observed patterns of AGF genera selective enrichment in rumen versus feces could be extrapolated to a more global level on data sets where fecal and rumen samples were obtained from different animals. We, thus, compared the community structure of cattle rumen and fecal AGF communities using the cattle rumen samples analyzed in this study ( $n = 116$ ) and 178 cattle fecal samples obtained in a recent global survey of the AGF mycobiome (15). For this larger data set



**FIG 4** Patterns of AGF beta diversity in the rumen mycobiome. PCoA ordination plots based on AGF community structure in the 206 samples studied here constructed using the phylogenetic similarity-based weighted Unifrac. The shape represents the animal family as shown on top. The % variance explained by the first two axes is displayed on the axes, and ellipses encompassing 95% of variance are displayed. In (A), samples and ellipses are color-coded by animal family (left) or animal species (right). In (B), samples and ellipses are color-coded by animal domestication status when using the total data set (top), or only for animals belonging to the families *Bovidae* (middle left), *Cervidae* (middle right), or the genus *Cervus* (bottom). In (C), samples and ellipses are color-coded by animal biogeography when using the total data set (top) or only for cattle (bottom left) or sheep (bottom right). PERMANOVA results for partitioning the dissimilarity by variation sources (animal family, animal species, domestication status, and country) are shown for each plot.  $R^2$  refers to percentage variance explained by each factor (calculated as the percentage of the sum of squares of each factor to the total sum of squares), while  $P$ -value refers to the F-statistics  $P$ -value.



**FIG 5** Rumen-feces mycobiome comparison. (A–C) Same individual rumen-feces mycobiome comparison conducted on 13 animal subjects (12 cattle, and 1 buffalo). (A) Collective AGF community composition for each sampling location. Genera with >1% total abundance are color-coded as shown to the right. All other genera are grouped as “others.” (B) AGF alpha diversity patterns in the 13 rumen versus feces samples. Box and whisker plots show the distribution of Shannon (left), Simpson (middle), and Inverse Simpson (right) diversity indices in the two sampling locations. (C) DPCoA biplot based on the phylogenetic similarity-based index weighted Unifrac showing the community structure in the 13 rumen and 13 feces samples. The % variance explained by the first two axes is displayed on the axes, and ellipses encompassing 95% of variance are displayed. The samples and ellipses are color-coded by sampling location (rumen, blue; feces, green). AGF genera are shown as smaller empty black circles and the four AGF genera with selective enrichment in either sampling locations are labeled. PERMANOVA results are shown in the bottom left corner of the plot, where  $R^2$  refers to percentage variance explained by the sampling location (calculated as the percentage of the sum of squares of each factor to the total sum of squares), while  $P$ -value refers to the F-statistics  $P$ -value. To the right of the DPCoA plot, (Continued on next page)

**FIG 5 (Continued)**

the level of variability between samples from the same sampling location is shown as box and whisker plots for the distribution of DPCoA ordination distance of each sample to its group centroid, and results for two-tailed ANOVA are shown on top of the box plots: \*,  $0.01 < P < .05$ . Results of Metastats and MaAsLin2 for these four genera are shown in the table. For each taxon, the average and standard deviations of abundance are shown for the rumen versus feces, followed by Metastats  $P$ -value, MaAsLin2 model parameters (including the model coefficient with feces used as reference, standard error, and  $P$ -value), and the sampling location where the taxon was identified as significantly differentially abundant. A positive MaAsLin2 model coefficient signifies enrichment in the rumen, while a negative model coefficient signifies enrichment in the feces. NS, not significant. (D) Global cattle rumen-feces mycobiome comparison conducted on the cattle rumen samples analyzed in this study ( $n = 116$ ) and 178 cattle fecal samples obtained in a recent global survey of the AGF mycobiome (15). DPCoA biplot based on the phylogenetic similarity-based index weighted Unifrac showing the community structure in the 116 rumen and 178 feces samples. The % variance explained by the first two axes is displayed on the axes, and ellipses encompassing 95% of variance are displayed. The samples and ellipses are color-coded by sampling location (rumen, blue; feces, green). The 13 genera showing significant association with feces are displayed in green text, while the five genera showing significant association with rumen are displayed in blue text. Results of Metastats and MaAsLin2 for these 18 genera are shown in Table S3. The level of variability between samples from the same sampling location is shown as box and whisker plots for the distribution of DPCoA ordination distance of each sample to its group centroid, and results for two-tailed ANOVA are shown on top of the box plots: \*\*\*\*,  $P < 0.0001$ .

(Fig. 5D), DPCoA ordination plots constructed using weighted Unifrac indices showed significant clustering with sampling location (rumen versus feces), albeit explaining only a minor fraction of the community variance (7.2%). Similar to the smaller data set, a significantly greater degree of variability was observed in rumen versus feces samples (ANOVA  $P$ -value = 4.7E-05) (Fig. 5D). Metastats and/or MaAsLin2 (Table S3) showed significant association of 13 genera with feces (green text in the DPCoA plot in Fig. 5D), and 5 genera with rumen (blue text in the DPCoA plot in Fig. 5D). DPCoA showed a similar pattern for the four AGF taxa identified above (*Neocallimastix* and *Orpinomyces* clustering close to rumen samples, and the genera *Caecomyces* and *Cyllamyces* clustering close to fecal samples) (Fig. 5D). However, while both methods used for testing genera significant association (Metastats and MaAsLin2) confirmed significance of the selective enrichment of *Neocallimastix* in cattle rumen and *Cyllamyces* in cattle feces, only one of the two methods confirmed significance of the selective enrichment of *Orpinomyces* in cattle rumen and *Caecomyces* in cattle feces (Table S3).

## DISCUSSION

We present a detailed assessment of the rumen compartment mycobiome in 206 mammalian herbivores using a culture-independent amplicon-based survey. The current study is significant because, firstly, it provides an extensive data set (206 samples from 15 animal species, 15 countries, and 6 continents) that offers a global overview of AGF community in the rumen. While similar AGF data sets were obtained from fecal samples, studies on the AGF community in the rumen have been sparse, with prior studies being quite limited in scope (17–22). Second, the current study provides a detailed comparison of the rumen versus fecal AGF communities using samples obtained from the same subject, as well as from global rumen and feces data sets. Third, by targeting the less accessible rumen fraction, the current study provides a direct view of the community actively involved in feed degradation in the rumen compartment, and such knowledge is a prerequisite for applied efforts for targeted manipulation of AGF community to promote digestion and absorption in ruminants.

Our results highlight the high level of overall gamma diversity within the global rumen mycobiome, with 81 out of the 88 currently reported AGF genera identified (Fig. 1C). Furthermore, the seven AGF genera not identified in this study were extremely rare, identified in anywhere between 0 and 56 samples in a recent global study that examined AGF community in a data set of 661 herbivorous fecal sample (15) and always present in extremely low relative abundance values ranging between 0% and 0.043% when encountered. More importantly, the AGF rumen mycobiome community composition displayed a pattern where a relatively limited number of genera were ubiquitous (occurring in >50% of the samples) and abundant (representing a large fraction of the AGF community when encountered) (Fig. 1D; Fig. S1). Such a pattern of high diversity and predominance of few genera is consistent with prior surveys of AGF in

fecal samples of herbivores (15, 36). The rationale behind the existence and maintenance of perpetually rare genera within the herbivorous gut has previously been debated, and is potentially attributed to their superior survival capabilities or probable role played under specific conditions not adequately captured in the current sampling schema (e.g., younger age, stress, specific types of feed) (15). Significantly, our analysis failed to identify novel AGF genera beyond those previously observed in prior feces-based surveys [Fig. 1C and D (13–16, 20)], hence refuting the proposition that rumen samples could represent a significant reservoir of novel, hitherto undescribed AGF diversity. This does not preclude the possibility of encountering novel AGF diversity in wild ruminants with specialized diets or feeding behaviors, like reindeer feeding on lichen or browsing ruminants in tropical forests, but widespread novelty seems unlikely.

Beyond documenting the occurrence and relative abundance (Fig. 1) of AGF taxa, we examined patterns of their diversity and community structure in the rumen mycobiome and attempted to elucidate the role of and interplay between various factors in shaping the observed patterns. Our results document statistically significant differences in levels of diversity (Fig. 2b; Fig. S2) and community structure (Fig. 4) patterns between various families and host species, suggesting a pattern of phylosymbiosis, where host phylogenetic affiliation plays a role in shaping the AGF community. As well, LIPA analysis (Fig. S5) has shown few specific pairwise AGF genus-animal species associations (e.g., goat with *Caecomyces*, *Orpinomyces*, *Cyllamyces*, and *Neocallimastix*, buffalo with *Orpinomyces*, American bison with NY9, pronghorn, elk, and mule deer with *Khoyollomyces*). Interestingly, recent work on the fecal AGF mycobiome has also identified patterns of phylosymbiosis, with specific LIPA preferences largely concordant in most animals shared between the two data sets (e.g., buffalo with *Orpinomyces*, elk and mule deer with *Khoyollomyces*). Such correlation, as previously suggested (15), could reflect a pattern of co-evolutionary symbiosis (i.e., a deep, intimate co-evolutionary process between fungi and their preferred animal hosts). Alternatively, such preferences could be a function of post-evolutionary environmental filtering process, where prevalent *in situ* conditions in a specific animal select for specific taxa regardless of the partners' evolutionary history.

It is important to note that quantitative assessment of the role of host identity in explaining rumen AGF community structure (PERMANOVA and multivariate matrices comparisons using MRM, Mantel, and Procrustes) indicates that host species/family could explain only a relatively small fraction of the observed variance (Fig. 4; Fig. S4). This indicates that additional factors such as domestication status, biogeography, and diet could possibly play an additional role in shaping the rumen AGF community. Domesticated animals typically receive a less diverse, more frequent, and homogenous dietary regimen that is often grain-rich. This is in stark contrast to wild herbivores that browse or graze on a more heterogenous diet with a feeding regimen controlled by resource availability and predation risk. Our results indicate a higher level of AGF alpha diversity in wild animals (Fig. 2C; Fig. S2e and f), and a significant role (F-statistic  $R^2 = 11.14\%$ ,  $P$ -value = 0.001) for domestication status in shaping AGF community (Fig. 4B; Fig. S4). As such, we posit that more variable feed types and non-monotonous feeding patterns in wild herbivores could lead to enrichment and co-existence of a more diverse AGF community suited to a more stochastic feeding regimen. However, it is important to note that all animal species examined were either exclusively wild or domesticated, leading to a potential conflation of both factors (animal species and domestication status) as drivers of AGF diversity and community structure. We attempted to partially control for the conflation of both factors by re-analyzing the impact of domestication on AGF diversity and community structure on subsets of the data sets comprised of animals from the same family (families *Bovidae* and *Cervidae*, Fig. 2C; Fig. S2e and f; Fig. 4b) or species (genus *Cervus*, Fig. 2C; Fig. S2e and f; Fig. 4b). The impact of domestication on diversity and selection of taxa were not statistically significant and could be driven by conflating factors. Therefore, only a highly controlled experiment, where wild and domesticated subjects belonging to the same animal species from the same location are compared, could conclusively disentangle both factors, e.g., capturing, rearing, and

sampling white-tailed deer species in a domesticated setting and comparing their AGF community to wild deer from the same region.

Biogeography could be an additional factor impacting AGF diversity and community structure, as previously postulated for rumen bacterial and archaeal communities (27). Our results show that biogeography could play a role in shaping AGF diversity (Fig. 2B; Fig. S2c and d) and community structure (Fig. 4C). However, similar to domestication status, the result of biogeographic-based assessments could be skewed by the over-representation of specific animal species in certain locations. We attempted to partly disentangle host and biogeography by reanalyzing subsets constituting the same animal species from different locations. Our results suggest a role for biogeography in shaping AGF diversity in cattle. It is interesting to note that a similar observation was also discerned in a recent global data set of fecal samples (15). The role of biogeography in shaping the AGF community could be driven by variability in cattle breeds (with dairy cattle having faster passage rate and higher metabolic rate than beef cattle breeds), feeding regimen, and rearing conditions between two locations.

Diet could play an important role in shaping AGF community, as previously suggested from bacterial and archaeal studies, e.g., references (37–39). The stark difference between diet of wild versus domesticated herbivores, as described above, could be an important factor in shaping observed differences between wild and domesticated animals examined in this study. Unfortunately, no metadata regarding the diet of wild animals examined in this study are available, hindering further exploration on the effect of diet on AGF community within wild herbivores as well as between wild and domesticated herbivores. As such, we limited our exploration of diet to the effect of feed type in domesticated animals, mostly cattle, on the AGF community structure. While AGF community diversity and community structure appears to be impacted by diet (forage type detail, forage type category, forage plant detail, and forage plant category), the effect only explained a fraction of the observed variance in the community.

Prior studies on AGF diversity in ruminants have largely been conducted on fecal rather than rumen samples (13–16, 20). The lack of studies on AGF communities in the rumen was largely hampered by methodological limitations. Collection of rumen samples requires surgical fistulation or gastric tubing, processes that could be conducted in research settings but are largely unfeasible for a broad sampling of herds in farming and ranching settings (40, 41). In wild ruminants, such an approach is not feasible, except in extremely rare conditions, where domestication of a naturally wild host was achieved (42). Theoretically, differences in AGF community between rumen and feces could be driven by selection for or against specific AGF taxa when passing through various regions within the animal's alimentary tract, enrichment of specific AGF genera involved in intestinal fermentation (23), or interaction between AGF and the distinct bacterial and archaeal communities colonizing various locations in the animal's alimentary tract. Using a pairwise sampling scheme in 13 animal subjects, we sought to assess differences between AGF communities in rumen versus feces samples. We acknowledge the relatively limited number of replicates and restriction to mostly one species (*Bos taurus*) and hence the patterns obtained should be regarded as preliminary. Our analysis clearly demonstrated that the AGF community in rumen samples is significantly more diverse than feces (Fig. 5B). As well, distinct differences in community structure were observed between rumen and feces samples, with a selective enrichment of the genera *Neocalymastix* and *Orpinomyces* in rumen samples and *Caecomyces* and *Cyllamyces* in fecal samples. The underlying reasons for the observed inhibition and enrichment trends are presently unclear, given our current rudimentary knowledge regarding fine differences in metabolic and physiological preferences between various AGF genera. Nevertheless, it is notable that the genera *Caecomyces* and *Cyllamyces* are the only known AGF exhibiting a bulbous rhizoidal growth pattern and appear to have a unique attachment/pressing on plants compared to filamentous rhizoids. This growth pattern, with a higher proportion of the fungal thallus protected within the plant biomass, compared to the more superficial external hyphal attachment pattern in filamentous genera could offer a

better protection during rumen contents passage through the highly acidic abomasum to the intestine. As well, while all AGF appear to grow readily and specialize in attacking intact plant biomass, a differential preference or efficiency of some genera in attacking, penetrating, and colonizing intact plant biomass would confer a competitive advantage in the rumen, where intact plants are first acted upon by the animal's microbiome. On the other hand, a greater affinity for oligomers, dimers, and monomers uptake could enrich specific genera in the colon, where available substrates are mostly soluble sugars rather than intact plant material. It is interesting to note that distinct differences in rumen versus fecal communities have also been observed in bacteria and archaea, where a similar pattern of lower diversity in feces was observed, as well as a distinct preference for fiber-degrading taxa (e.g., *Fibrobacter*) in rumen as opposed to sugar-degrading taxa (e.g., *Tenericutes*) in feces (43, 44).

## ACKNOWLEDGMENTS

We thank the following members of the GRC project for contributing samples used in this study:

Olubukola Ajike Isah: Department of Animal Nutrition, Federal University of Agriculture, Abeokuta (FUNAAB), Nigeria. Jorge Avila-Stagno: Facultad de Ciencias Veterinarias, Universidad de Concepción, Chillan, Chile. Kasper Dieho, Jan Dijkstra, and Andre Bannink: Animal Nutrition Group, Wageningen University, 6700 AH Wageningen, The Netherlands. Fabian N. Fon: Department of Agriculture, University of Zululand, KwaDlangezwa, Empangeni, 3886, South Africa. Hilario Mantovani: Departamento de Microbiología, Universidade Federal de Viçosa, Campus UFV, 36570-000 Viçosa, Minas Gerais, Brazil. Martin Fraga: Departamento de Microbiología, Instituto de Investigaciones Biológicas Clemente Estable, Av. Italia 3318, CP 11600, Montevideo, Uruguay. Francisco E. Franco: VITA Marangani, Universidad Nacional Mayor de San Marcos, Lima, Perú. Chris Friedman: Ministry for Primary Industries Verification Services Hawkes Bay, Silver Fern Farms—Pacific, Whakatu, Hastings, New Zealand. Arjan Jonker and Cesar S. Pinares-Patino: AgResearch Limited, Grasslands Research Centre, Palmerston North 4442, New Zealand. Sophie Krizsan: Department of Agricultural Research for Northern Sweden, Swedish University of Agricultural Sciences, SE-901 83 Umea, Sweden. Jan Lassen: Department of Molecular Biology and Genetics, Aarhus University, DK-8830 Tjele, Denmark. Satoshi Koike and Yasuo Kobayashi: Laboratory of Animal Nutrition, Research Faculty of Agriculture, Hokkaido University, N9W9, Kita-ku, Sapporo 060-8589, Japan. Sang Suk Lee and Lovelia L. Mamuad: Department of Animal Science & Technology, Sunchon National University, Suncheon, Jeonnam 540-742, Korea. Cécile Martin, Diego Morgavi, and Milka Popova: Institut National de la Recherche Agronomique, UMR1213 Herbivores, Saint-Genes-Champanelle, F-63122, France. Andreas Muenger: Agroscope, Institute for Livestock Sciences ILS, CH-1725 Posieux, Switzerland. Camila Muñoz, Rodrigo de la Barra, María Eugenia Martínez, and Andrés M. Carvajal: Instituto de Investigaciones Agropecuarias, INIA Remehue, Osorno, Región de Los Lagos, Chile. Mario A. Cobos-Peralta: Colegio de Postgraduados, Institución de Enseñanza e Investigación en Ciencias Agrícolas, CP 56230, Montecillo, Mexico. Tasia Taxis: Animal Science Genetics, University of Missouri-Columbia, Columbia, Missouri 65211, USA. Emilio M. Ungerfeld: Instituto de Investigaciones Agropecuarias, INIA Carillanca, Temuco, Chile. Min Wang and Zhi Liang Tan: Institute of Subtropical Agriculture, The Chinese Academy of Sciences, Changsha, China. Tianhai Yan: Agri-Food and Biosciences Institute, Hillsborough, County Down BT26 6DR, Northern Ireland.

This work has been supported by the NSF grant number 2029478 to M.S.E. and N.H.Y., and the New Zealand Ministry of Business, Innovation and Employment Strategic Science Investment Fund AgResearch Microbiomes programme to C.D.M. The collection of the Global Rumen Census samples was supported by the New Zealand Government as part of its support for the Global Research Alliance on Agricultural Greenhouse Gases to P.H.J. Montana wild ruminant samples were collected with support of the Bair Ranch Foundation and the Montana Agricultural Experiment Station. Some of the

computing for this project was performed at the High-Performance Computing Center at Oklahoma State University supported in part through the National Science Foundation grant OAC-1531128.

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, Oklahoma, USA

<sup>2</sup>Department of Microbiology and Immunology, Faculty of Pharmacy, Cairo University, Cairo, Egypt

<sup>3</sup>AgResearch Ltd, Grasslands Research Centre, Palmerston North, New Zealand

<sup>4</sup>Department of Animal and Range Sciences, Montana State University, Bozeman, Montana, USA

<sup>5</sup>Department of Animal and Food Sciences, Oklahoma State University, Stillwater, Oklahoma, USA

## AUTHOR ORCIDs

Casey H. Meili  <http://orcid.org/0000-0003-0379-3440>

Christina D. Moon  <http://orcid.org/0000-0001-9692-9559>

Peter H. Janssen  <http://orcid.org/0000-0002-1022-3502>

Joanna-Lynn C. Borgogna  <http://orcid.org/0000-0002-4488-4042>

Noha H. Youssef  <http://orcid.org/0000-0001-9159-3785>

## FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	2029478	Andrew P. Foote Mostafa S. Elshahed Noha H. Youssef

## AUTHOR CONTRIBUTIONS

Casey H. Meili, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review and editing | Moustafa A. TagElDein, Data curation, Formal analysis, Methodology, Writing – review and editing | Adrienne L. Jones, Data curation, Formal analysis, Writing – review and editing | Christina D. Moon, Data curation, Formal analysis, Funding acquisition, Methodology, Visualization, Writing – review and editing | Catherine Andrews, Methodology, Resources | Michelle R. Kirk, Data curation, Methodology | Peter H. Janssen, Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Writing – review and editing | Carl J. Yeoman, Data curation, Formal analysis, Funding acquisition, Methodology, Writing – review and editing | Savannah Grace, Data curation, Formal analysis | Andrew P. Foote, Data curation, Funding acquisition, Methodology, Resources, Writing – review and editing | Yosra I. Nagy, Data curation | Mona T. Kashef, Data curation.

## DATA AVAILABILITY

Illumina amplicon reads generated in this study have been deposited in GenBank SRA under BioProject accession number [PRJNA1008183](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1008183) and Biosample accessions numbers [SAMN37111842](https://www.ncbi.nlm.nih.gov/biosample/SAMN37111842)-[SAMN37112060](https://www.ncbi.nlm.nih.gov/biosample/SAMN37112060).

## ETHICS APPROVAL

Sample collection and handling in Egypt was approved by the Committee for Safe Handling and Disposal of Chemical and Biological Materials, Faculty of Pharmacy Cairo University # MI3011, June 2021. The sampling procedure at Oklahoma State University

was reviewed and approved by the Oklahoma State University Institutional Animal Care and Use Committee (Protocol #21-03)

## ADDITIONAL FILES

The following material is available [online](#).

### Supplemental Material

**Supplementary document (AEM01492-23-S0001.pdf).** Entire supplementary document including Figures S1-S5 and Tables S1-S3.

**Table S1 (AEM01492-23-S0002.xlsx).** Supplemental Table S1 in Excel format, samples used in this study.

**Table S2 (AEM01492-23-S0003.xlsx).** Supplemental Table S2 in Excel format, microbial function composition.

### REFERENCES

1. Hackmann TJ, Spain JN. 2010. Invited review: ruminant ecology and evolution: perspectives useful to ruminant livestock research and production. *J Dairy Sci* 93:1320–1334. <https://doi.org/10.3168/jds.2009-2071>
2. Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen K, Cooper A, Vilki J, Seabury CM, Caetano AR, et al. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc Natl Acad Sci USA* 106:18644–18649. <https://doi.org/10.1073/pnas.0904691106>
3. Heller R, Frandsen P, Lorenzen ED, Siegmund HR. 2013. Are there really twice as many bovid species as we thought?. *Syst Biol* 62:490–493. <https://doi.org/10.1093/systbio/syt004>
4. Moraïs S, Mizrahi I. 2019. Islands in the stream: from individual to communal fiber degradation in the rumen ecosystem. *FEMS Microbiol Rev* 43:362–379. <https://doi.org/10.1093/femsre/fuz007>
5. Matthews C, Crispie F, Lewis E, Reid M, O'Toole PW, Cotter PD. 2019. The rumen microbiome: a crucial consideration when optimising milk and meat production and nitrogen utilisation efficiency. *Gut Microbes* 10:115–132. <https://doi.org/10.1080/19490976.2018.1505176>
6. Gruninger RJ, Puniya AK, Callaghan TM, Edwards JE, Youssef N, Dagar SS, Fliegerova K, Griffith GW, Forster R, Tsang A, McAllister T, Elshahed MS. 2014. Anaerobic fungi (Phylum *Neocallimastigomycota*): advances in understanding their taxonomy, life cycle, ecology, role and biotechnological potential. *FEMS Microbiol Ecol* 90:1–17. <https://doi.org/10.1111/1574-6941.12383>
7. Yan M, Pratama AA, Somasundaram S, Li Z, Jiang Y, Sullivan MB, Yu Z. 2023. Interrogating the viral dark matter of the rumen ecosystem with a global virome database. *Nat Commun* 14:5254. <https://doi.org/10.1038/s41467-023-41075-2>
8. Couger MB, Youssef NH, Struchtemeyer CG, Liggensstoffer AS, Elshahed MS. 2015. Transcriptomic analysis of lignocellulosic biomass degradation by the anaerobic fungal isolate *Orpinomyces* sp. strain C1A. *Biotechnol Biofuels* 8:208. <https://doi.org/10.1186/s13068-015-0390-0>
9. Hagen LH, Brooke CG, Shaw CA, Norbeck AD, Piao H, Arntzen MØ, Olson HM, Copeland A, Isern N, Shukla A, Roux S, Lombard V, Henrissat B, O'Malley MA, Grigoriev IV, Tringe SG, Mackie RI, Pasa-Tolic L, Pope PB, Hess M. 2021. Proteome specialization of anaerobic fungi during ruminal degradation of recalcitrant plant fiber. *ISME J* 15:421–434. <https://doi.org/10.1038/s41396-020-00769-x>
10. Youssef NH, Couger MB, Struchtemeyer CG, Liggensstoffer AS, Prade RA, Najar FZ, Atiyeh HK, Wilkins MR, Elshahed MS. 2013. The genome of the anaerobic fungus *Orpinomyces* sp. strain C1A reveals the unique evolutionary history of a remarkable plant biomass degrader. *Appl Environ Microbiol* 79:4620–4634. <https://doi.org/10.1128/AEM.00821-13>
11. Elliott R, Ash AJ, Calderon-Cortes F, Norton BW, Bauchop T. 1987. The influence of anaerobic fungi on rumen volatile fatty acid concentrations *in vivo*. *J Agric Sci* 109:13–17. <https://doi.org/10.1017/S0021859600080928>
12. Huws SA, Creevey CJ, Oyama LB, Mizrahi I, Denman SE, Popova M, Muñoz-Tamayo R, Forano E, Waters SM, Hess M, et al. 2018. Addressing global ruminant agricultural challenges through understanding the rumen microbiome: past, present, and future. *Front Microbiol* 9:2161. <https://doi.org/10.3389/fmicb.2018.02161>
13. Hanafy RA, Johnson B, Youssef NH, Elshahed MS. 2020. Assessing anaerobic gut fungal diversity in herbivores using D1/D2 large ribosomal subunit sequencing and multi-year isolation. *Environ Microbiol* 22:3883–3908. <https://doi.org/10.1111/1462-2920.15164>
14. Liggensstoffer AS, Youssef NH, Couger MB, Elshahed MS. 2010. Phylogenetic diversity and community structure of anaerobic gut fungi (Phylum *Neocallimastigomycota*) in ruminant and non-ruminant herbivores. *ISME J* 4:1225–1235. <https://doi.org/10.1038/ismej.2010.49>
15. Meili CH, Jones AL, Arreola AX, Habel J, Pratt CJ, Hanafy RA, Wang Y, Yassin AS, TagElDein MA, Moon CD, et al. 2023. Patterns and determinants of the global herbivorous mycobiome. *Nat Commun* 14:3798. <https://doi.org/10.1038/s41467-023-39508-z>
16. Young D, Joshi A, Huang L, Munk B, Wurzbacher C, Youssef NH, Elshahed MS, Moon CD, Ochsenreither K, Griffith GW, Callaghan TM, Sczryba A, Lebuhn M, Flad V. 2022. Simultaneous metabarcoding and quantification of *Neocallimastigomycetes* from environmental samples: insights into community composition and novel lineages. *Microorganisms* 10:1749. <https://doi.org/10.3390/microorganisms10091749>
17. Azad E, Fehr KB, Derakhshani H, Forster R, Acharya S, Khafipour E, McGeough E, McAllister TA. 2020. Interrelationships of fiber-associated anaerobic fungi and bacterial communities in the rumen of bloated cattle grazing alfalfa. *Microorganisms* 8:1543. <https://doi.org/10.3390/microorganisms8101543>
18. Guo W, Wang W, Bi S, Long R, Ullah F, Shafiq M, Zhou M, Zhang Y. 2020. Characterization of anaerobic rumen fungal community composition in yak, Tibetan sheep and small tail Han sheep grazing on the Qinghai-Tibetan plateau. *Animals* 10:144. <https://doi.org/10.3390/ani10010144>
19. Huuki H, Tapio M, Mäntysaari P, Negussie E, Ahvenjärvi S, Vilki J, Vanhatalo A, Tapio I. 2022. Long-term effects of early-life rumen microbiota modulation on dairy cow production performance and methane emissions. *Front Microbiol* 13:983823. <https://doi.org/10.3389/fmicb.2022.983823>
20. Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JL, Janssen PH. 2013. Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS One* 8:e47879. <https://doi.org/10.1371/journal.pone.0047879>
21. Kumar S, Indugu N, Vecchiarelli B, Pitta DW. 2015. Associative patterns among anaerobic fungi, methanogenic archaea, and bacterial communities in response to changes in diet and age in the rumen of dairy cows. *Front Microbiol* 6:781. <https://doi.org/10.3389/fmicb.2015.00781>
22. Moon CD, Carvalho L, Kirk MR, McCulloch AF, Kittelmann S, Young W, Janssen PH, Leathwick DM. 2021. Effects of long-acting, broad spectra anthelmintic treatments on the rumen microbial community compositions of grazing sheep. *Sci Rep* 11:3836. <https://doi.org/10.1038/s41598-021-82815-y>

23. Hartinger T, Zebeli Q. 2021. The present role and new potentials of anaerobic fungi in ruminant nutrition. *J Fungi (Basel)* 7:200. <https://doi.org/10.3390/jof7030200>
24. de Oliveira MNV, Jewell KA, Freitas FS, Benjamin LA, Tótola MR, Borges AC, Moraes CA, Suen G. 2013. Characterizing the microbiota across the gastrointestinal tract of a Brazilian Nelore steer. *Vet Microbiol* 164:307–314. <https://doi.org/10.1016/j.vetmic.2013.02.013>
25. Wang K, Zhang H, Hu L, Zhang G, Lu H, Luo H, Zhao S, Zhu H, Wang Y. 2022. Characterization of the microbial communities along the gastrointestinal tract in crossbred cattle. *Animals (Basel)* 12:825. <https://doi.org/10.3390/ani12070825>
26. Swift CL, Louie KB, Bowen BP, Hooker CA, Solomon KV, Singan V, Daum C, Pennacchio CP, Barry K, Shuthanandan V, Evans JE, Grigoriev IV, Northen TR, O’Malley MA. 2021. Cocultivation of anaerobic fungi with rumen bacteria establishes an antagonistic relationship. *mBio* 12:e01442-21. <https://doi.org/10.1128/mBio.01442-21>
27. Henderson G, Cox F, Ganesh S, Jonker A, Young W, AbeciaL, Angarita E, Aravena P, et al. 2015. Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci Rep* 5:14567. <https://doi.org/10.1038/srep14567>
28. Elshahed MS, Hanafy RA, Cheng Y, Dagar SS, Edwards JE, Flad V, Fliegerová KO, Griffith GW, Kittelmann S, Lebuhn M, O’Malley MA, Podmirseg SM, Solomon KV, Vinzelj J, Young D, Youssef NH. 2022. Characterization and rank assignment criteria for the anaerobic fungi (*Neocallimastigomycota*). *Int J Syst Evol Microbiol* 72:005449. <https://doi.org/10.1099/ijsem.0.005449>
29. McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. <https://doi.org/10.1371/journal.pone.0061217>
30. Dixon P. 2003. VEGAN, a package of R functions for community ecology. *J Veg Sci* 14:927–930. <https://doi.org/10.1111/j.1654-1103.2003.tb02228.x>
31. Goslee SC, Urban DL. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *J Stat Soft* 22:1–9. <https://doi.org/10.18637/jss.v022.i07>
32. Maechler M, RousseeuwP, StruyfA, HubertM, Hornik K. 2023. Cluster: cluster analysis basics and extensions. Available from: <https://CRAN.R-project.org/package=cluster>
33. Keck F, Rimet F, Bouchez A, Franc A. 2016. Phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol Evol* 6:2774–2780. <https://doi.org/10.1002/ece3.2051>
34. White JR, Nagarajan N, Pop M, Ouzounis CA. 2009. Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* 5:e1000352. <https://doi.org/10.1371/journal.pcbi.1000352>
35. Mallick H, Rahnavard A, McIver LJ, Ma S, Zhang Y, Nguyen LH, Tickle TL, Weingart G, Ren B, Schwager EH, Chatterjee S, Thompson KN, Wilkinson JE, Subramanian A, Lu Y, Waldron L, Paulson JN, Franzosa EA, Bravo HC, Huttenhower C, Coelho LP. 2021. Multivariable association discovery in population-scale meta-omics studies. *PLoS Comput Biol* 17:e1009442. <https://doi.org/10.1371/journal.pcbi.1009442>
36. Adrienne LJ, Carrie JP, Casey HM, Rochelle MS, Philip H, Mostafa SE, Noha HY. 2023. Anaerobic gut fungal communities in marsupial hosts. *bioRxiv*. <https://doi.org/10.1101/2023.05.31.543067:2023.05.31.543067>
37. Fliegerová KO, Podmirseg SM, Vinzelj J, Grilli DJ, Kvasnová S, Schierová D, Sechovcová H, Mrázek J, Siddi G, Arenas GN, Moniello G. 2021. The effect of a high-grain diet on the rumen microbiome of goats with a special focus on anaerobic fungi. *Microorganisms* 9:157. <https://doi.org/10.3390/microorganisms9010157>
38. Tapió I, Fischer D, Blasco L, Tapió M, Wallace RJ, Bayat AR, Ventto L, Kahala M, Negussie E, Shingfield KJ, Vilki J. 2017. Taxon abundance, diversity, co-occurrence and network analysis of the ruminal microbiota in response to dietary changes in dairy cows. *PLoS One* 12:e0180260. <https://doi.org/10.1371/journal.pone.0180260>
39. Youngblut ND, Reischer GH, Dauser S, Maisch S, Walzer C, Stalder G, Farnleitner AH, Ley RE. 2021. Vertebrate host phylogeny influences gut archaeal diversity. *Nat Microbiol* 6:1443–1454. <https://doi.org/10.1038/s41564-021-00980-2>
40. Castillo C, Hernández J. 2021. Ruminal fistulation and cannulation: a necessary procedure for the advancement of biotechnological research in ruminants. *Animals (Basel)* 11:1870. <https://doi.org/10.3390/ani11071870>
41. Hagey JV, Laabs M, Maga EA, DePeters EJ. 2022. Rumen sampling methods bias bacterial communities observed. *PLoS One* 17:e0258176. <https://doi.org/10.1371/journal.pone.0258176>
42. Selden LM, Naas AE, Roux S, Daly RA, Collins WB, Nicora CD, Purvine SO, Hoyt DW, Schückel J, Jørgensen B, Willats W, Spalinger DE, Firkins JL, Lipton MS, Sullivan MB, Pope PB, Wrighton KC. 2018. Interspecies cross-feeding orchestrates carbon degradation in the rumen ecosystem. *Nat Microbiol* 3:1274–1284. <https://doi.org/10.1038/s41564-018-0225-4>
43. Liu J, Zhang M, Zhang R, Zhu W, Mao S. 2016. Comparative studies of the composition of bacterial microbiota associated with the ruminal content, ruminal epithelium and in the faeces of lactating dairy cows. *Microb Biotechnol* 9:257–268. <https://doi.org/10.1111/1751-7915.12345>
44. Williamson JR, Callaway TR, Lourenco JM, Ryman VE. 2022. Characterization of rumen, fecal, and milk microbiota in lactating dairy cows. *Front Microbiol* 13:984119. <https://doi.org/10.3389/fmicb.2022.984119>