Buccal swabbing as a non-invasive method to determine bacterial, archaeal, and eukaryotic microbial community structure in the rumen

SUPPLEMENTAL MATERIAL

Sandra Kittelmann,[a]# Michelle R. Kirk,[a] Arjan Jonker[a], Alan McCulloch,[b] Peter H. Janssen[a]#

AgResearch Ltd. Grasslands Research Centre, Palmerston North, New Zealand[a]; AgResearch Ltd. Invermay Agricultural Centre, Mosgiel, New Zealand[b]

Running head: Buccal swabbing to analyze rumen microbial communities

Address correspondence to Peter H. Janssen, peter.janssen@agresearch.co.nz or Sandra Kittelmann, sandra.kittelmann@agresearch.co.nz.

Text S1

Sequencing results from high-throughput sequencing of microbial marker genes

Using 454 Titanium pyrosequencing to sequence microbial communities in 96 samples, we obtained sufficient data to analyze bacteria, archaea, and protozoa in 90, 78 and 54 samples, respectively, at our selected sequence cut-off. The total numbers of sequencing reads were 225,836 partial bacterial 16S rRNA genes (mean number of reads per sample ± standard deviation: 2,509 ± 743), 129,754 partial archaeal 16S rRNA genes (mean of 1,664 ± 1,026 per sample), and 40,753 partial ciliate protozoa 18S rRNA genes (mean of 755 ± 737 per sample). The number of bacterial 16S rRNA genes reads were similar between samples collected via buccal swabs and stomach tubing (average ± SD: 2433 ± 879, and 2262 ± 791, respectively). In contrast, DNA samples collected using buccal swabs yielded notably lower numbers of reads for archaeal 16S rRNA genes and protozoal 18S genes (average ± SD: 1062 ± 1082 and 285 ± 578, respectively) than those collected using stomach tubing (average ± SD: 2242 ± 722 and 871 ± 717 for archaea and protozoa, respectively).

To evaluate llumina MiSeq sequencing technology and its applicability to analysing microbial community structures from DNA obtained with buccal swabs, data from two Illumina MiSeq runs were combined. Both runs used the same amplicon libraries. The two-step PCR approach used in this study, in combination with the deliberately lower starting concentration of the shorter anaerobic fungal ITS1 amplicons (approximately 250 bp in length), resulted in an targeted distribution of reads across the four microbial groups analyzed, and so allowed evaluation of community structure in nearly all 384 amplicon libraries. The exceptions in these 384 were two ciliate protozoal 18S rRNA gene and twelve anaerobic fungal ITS1 libraries generated from samples collected *via* buccal swabs, in which the number of reads was lower than the threshold we set for inclusion (100 reads/library for ciliate protozoa and 100 reads/library for anaerobic fungi). In general, a larger number of

reads ≥200 bp was obtained from the data of Read 1 than from that of Read 2 within a run. For Read 1, a total of more than 3.4×10^6 bacterial 16S rRNA gene sequences were obtained for the 96 samples analyzed (mean ± standard deviation: 35,565 ± 19,853 per sample). Archaeal 16S rRNA genes accounted for a total of 2.0×10^6 reads (mean of 21,046 ± 13,650 per sample), while protozoal 18S rRNA genes and anaerobic fungal ITS1 genes totalled 2.7×10^6 (mean of 28,252 ± 22,247 per sample) and 1.8×10^6 reads (mean of 18,503 ± 19,214 per sample), respectively. Sequencing depth obtained for Read 2 is provided separately (Table S2). Because we sequenced gene regions for bacteria, archaea, and protozoa that were approximately 500 bp long, only 8.3×10^5 (mean of 8,669 ± 4,695 per sample), 2.2×10^5 (mean of 2,285 ± 1,682 per sample), and 4.9×10^5 reads (mean of 5,086 ± 3,993 per sample) overlapped for bacteria, archaea, and protozoa, respectively. The amplicon used for anaerobic fungi was considerably shorter (approx. 250 bp), resulting in a total of 1.9×10^6 of overlapping sequences (mean of 19,917 ± 20,455 per sample).

Average Good's coverage (1) was high for both types of samples (Buccal and Rumen) and both sequencing technologies for all four microbial groups (on average > 98.5%; Table S1). In all cases where valid comparisons were possible, Illumina MiSeq technology (Read 1) provided slightly higher coverage than 454 Titanium sequencing. Both sequencing technologies generally achieved slightly higher coverage of samples collected *via* stomach tubing. These data indicated that adequate coverage was achieved to capture a large part of the bacterial, archaeal, ciliate protozoal and anaerobic fungal diversity in the analysed samples.

Reference

1. **Good IJ.** 1953. The population frequencies of species and the estimation of population parameters. Biometrika **40:**237–264.

Text S2

Comparison of 454 Titanium and Illumina MiSeq PE300 chemistry

For the comparison of sequencing technologies, we used 24 samples collected *via* stomach tubing that had been sequenced using both 454 Titanium and Illumina MiSeq PE300 chemistries. Bacterial 16S rRNA gene data for MiSeq Read 1 showed that the region downstream of primer Ba9F provided results that were not significantly different to the data obtained from full-length reads of the amplicons using 454 Titanium chemistry ($p = 0.33$, Fig. S1A). Read 2 data were also comparable to those obtained by 454 sequencing (downstream of primer Ba515Rmod1; $p = 0.11$; Fig. S1A). Irrespective of the sequence data type (454 Titanium, MiSeq Read 1, MiSeq Read 2), samples from animals that had been feeding on the same diet clustered together (Fig. S1B) and were clearly distinguishable from the samples from animals that had been feeding on different diets ($p < 0.001$, Fig. S1B). Similarly, Read 1 data of archaeal 16S rRNA genes (downstream of primer Ar1386R) resulted in community profiles highly similar to those obtained using 454 Titanium chemistry ($p = 0.84$ in test for difference, Fig. S1C). Read 2 data for archaea were, however, significantly different to those obtained by 454 sequencing (downstream of primer Ar915aF; $p < 0.001$; Fig. S1C). Significant differences between the diets were detected for all three types of sequence data ($p < 0.001$, Fig. S1D). However, samples analyzed based on MiSeq Read 2 did not cluster with samples obtained from sheep on the same diet analysed using either 454 or MiSeq Read 1 (Fig. S1D).

Analysis of ciliate protozoal communities using Read 1 data provided most similar results to 454 sequencing (downstream of primer Reg1302R, $p = 0.14$ in test for difference, Fig. S1E). In contrast, results based on Read 2 differed significantly from those based on 454 sequencing, with completely different taxa being detected (downstream of primer RP841F; $p < 0.001$; Fig. S1E). This was likely due to the fact that the region immediately following

primer RP841F is highly conserved and provides limited taxonomic resolution at the genus level. The three types of sequence data all indicated significant differences between ciliate communities derived from the different dietary treatment groups, but these were less pronounced than for bacterial and archaeal communities ($p < 0.001$, Fig. S1F).

Read 1 and Read 2 generated from fungal ITS1 sequences provided highly similar community structure data to each other ($p > 0.9$; Fig. S1G). Similar to the rumen ciliate communities, anaerobic fungi appeared to show only weak differences between diets ($p = 0.003$, Fig S1H), which tended to be not significant when testing for diet differences within individual sequence data types (MiSeq Read 1: $p = 0.06$, MiSeq Read 2: $p = 0.06$).

Based on these findings, comparisons between the microbial community structures in samples collected using the stomach tube method (Rumen) and the three different buccal swab methods (OM, PG, and SD) were made using the data set obtained from Read 1 of the Illumina MiSeq data for all four microbial groups.

Comparisons between the Illumina MiSeq (paired-end or single read) and Roche 454 GS FLX sequencing platforms have been performed previously, with the majority of studies analysing bacterial 16S rRNA gene (1-3), but also fungal ITS1 (1) diversity in various environments, and shotgun libraries for metagenomic surveys (4). Limited information is currently available on the comparability of archaeal 16S rRNA and rumen ciliate 18S rRNA gene amplicon data stemming from both systems. We also tested simultaneous sequencing of bacterial, archaeal and eukaryotic amplicons in the same Illumina MiSeq sequencing run, as already established for the Roche 454 GS FLX system and Titanium chemistry (5).

Our results confirmed that even single read sequence data from each potential pair obtained with the Illumina MiSeq PE300 chemistry was comparable to data obtained using Roche 454 GS FLX sequencing. However, the two reads that are obtained from paired-end sequencing differ in the taxonomic resolution they achieve. The higher taxonomic resolution of the

region following the primer sequenced in Read 1 for all microbial groups analyzed in this study is likely to be the reason for better comparability of Read 1 with 454 sequencing, rather than inadequate coverage with Illumina MiSeq Read 2, as Read 2 still achieved a higher coverage per sample than sequencing with 454 Titanium chemistry. Since data obtained from Read 1 currently out-performs that obtained from Read 2 in quality, it is important to partner the primer that is followed by the taxonomically more variable region of the gene with Illumina MiSeq adapter A (the binding region of the sequencing primer for Read 1). Doing so will provide highest coverage for the taxonomically most relevant sequence region, so that even single-end sequencing may already provide sufficient discriminating power between microbial communities from different environments or treatment groups. Obviously, as sequencing quality with increasing read length improves, the two paired reads will be able to be assembled with higher frequency, and this consideration will no longer be necessary. We found that the number of paired-end reads that assembled for the bacterial, archaeal, and protozoal markers (approx. 500 bp amplicons) was only 1/10th of the number of single reads ≥200 bp, necessitating the use of the single reads to maximize both the number of samples per run and the number of reads per sample. In contrast, the shorter (approx. 250 bp) amplicons from the fungal ITS1 resulted in a similar number of paired-end reads compared to single reads.

The Illumina MiSeq platform performed well when amplicons from the four different microbial groups were sequenced simultaneously. Since the quality of Illumina MiSeq technology relies on base heterogeneity, especially in the initial cycles of sequencing, this simultaneous multi-domain sequencing approach overcomes the low sequence diversity problem that arises from sequencing 16S rRNA gene amplicons alone. Therefore, in contrast to standard Illumina protocols, which commonly co-sequence the control library consisting of phage PhiX at up to 50% volume to guarantee optimal cluster identification and phasing/pre-

phasing rate determination during the initial cycles of the run, the sequencing provider only used 20%, resulting in a higher number of reads from the libraries being analyzed. Simultaneously sequencing of multiple marker genes for broad assessment of microbiota is widely applicable to other phylogenetic or functional marker genes and may be applied beyond the rumen to other ecosystems of interest. Coverage for each amplicon or even library may be individually adjusted as required. Combined with the dual-index sequencing protocol described by Kozich *et al.* (6), up to 1,536 amplicon libraries could potentially be sequenced simultaneously in a single run, and a recently proposed heterogeneity spacer approach could improve the obtained sequence yield even further (7). Our results suggest that single-direction Illumina MiSeq sequencing provides sufficient data and resolution to simultaneously identify differences in rumen-inhabiting prokaryotic and eukaryotic microbial communities between treatment groups.

References

1. **McNally R, Johnson M, Kasinadhuni N, Tinning M, McGrath K.** 2014. Changing of the guard: Migrating 16S and ITS diversity profiling from 454 onto alternate NGS platforms. J Biomol Tech **25:**S17.
2. **Sinclair L, Osman OA, Bertilsson S, Eiler A.** 2014. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. PLoS ONE **10:**e0116955.
3. **Werner JJ, Zhou D, Caporaso JG, Knight R, Angenent LT.** 2012. Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. ISME J **6:**1273.

4. **Luo C, Tsementzi D, Kyrpides N, Read T, Konstantinidis KT.** 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS ONE **7:**e30087.

5. **Kittelmann S, Seedorf H, Walters WA, Clemente JC, Knight R, Gordon JI, 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.

6. **Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD.** 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Appl Environ Microbiol **79:**5112–5120.

7. **Fadrosh DW, Ma B, Gajer P, Sengamalay N, Ott S, Brotman RM, Ravel J.** 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome **2:**1–7.

TABLE S1 Average Good's coverage for samples collected from 24 sheep *via* three different buccal swab methods (Buccal) or stomach tubing (Rumen) and sequenced using either 454 Titanium (454) or Illumina MiSeq technology (Read 1; IR1).

Microbial group	Sampling	Sequencing	Average ± SD	Minimum	Maximum
Bacteria	Buccal	454	98.55% ± 0.71%	95.20%	99.55%
	Buccal	IR1	99.78% ± 0.13%	99.39%	99.97%
	Rumen	454	99.28% ± 0.27%	98.47%	99.64%
	Rumen	IR1	99.84% ± 0.08%	99.64%	99.94%
Archaea	Buccal	454	99.78% ± 0.40%	98.02%	100%
	Buccal	IR1	99.98% ± 0.06%	99.61%	100%
	Rumen	454	99.89% ± 0.07%	99.76%	100%
	Rumen	IR1	99.99% ± 0.01%	99.96%	100%
Protozoa	Buccal	454	99.89% ± 0.23%	99.25%	100%
	Buccal	IR1	99.98% ± 0.06%	99.59%	100%
	Rumen	454	99.84% ± 0.27%	98.86%	100%

		Rumen	IR1	100% ± 0.00%	99.99%	100%
	Fungi	Buccal	454	n.a.[a]	n.a.	n.a.
		Buccal	IR1	99.94% ± 0.20%	98.74%	100%
210		Rumen	454	n.a.	n.a.	n.a.
		Rumen	IR1	99.99% ± 0.01%	99.96%	100%

---

[a]not applicable, as anaerobic fungal ITS1 amplicons were not analysed using 454 Titanium sequencing technology.

TABLE S2 Number of Illumina MiSeq PE300 sequencing reads obtained from Read 2 that were ≥200 bp in length.

Microbial group	Total number of reads	Reads per sample (mean ± SD[a])
Bacteria	873,331	9,097 ± 6,317
Archaea	344,962	3,593 ± 2,330
Ciliate protozoa	1,074,098	11,189 ± 8,284
Anaerobic fungi	1,644,437	17,130 ± 17,473

[a]Standard deviation.

TABLE S3 Relative abundances of bacterial taxa that were more than 1% higher in maximum relative abundance in samples collected *via* buccal swabs than in samples collected *via* stomach tubing and were thus eliminated from the dataset (mathematical filtering approach). Taxa indicated in bold were also eliminated in the manual filtering approach.

Taxon	Relative abundance			
	Buccal swabs		Rumen	
	Maximum	Average ± SD[a]	Maximum	Average ± SD
***Acaryochloris*, unknown sp. affil.**	0.03	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
Acinetobacter guillouiae[b]	0.01	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
Acinetobacter lwoffii	0.15	0.01 ± 0.02	<0.01	<0.01 ± <0.01
***Acinetobacter*, unknown sp. affil.**	0.03	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
Actinobacillus parahaemolyticus	0.03	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
***Actinobacillus*, unknown sp. affil.**	0.07	0.01 ± 0.01	<0.01	<0.01 ± <0.01

	***Actinomyces*, unknown sp. affil.**	0.03	0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Aerococcus*, unknown sp. affil.**	0.03	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Aggregatibacter*, unknown sp. affil.**	0.03	0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Bacillus*, unknown sp. affil.**	0.10	<0.01 ± 0.02	<0.01	<0.01 ± <0.01
	Bacteroidaceae, 5-7N15, unknown sp. affil.	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Bacteroides*, unknown sp. affil.**	0.08	0.01 ± 0.01	0.01	<0.01 ± <0.01
	Bibersteinia trehalosi	0.13	0.02 ± 0.03	<0.01	<0.01 ± <0.01
	***Campylobacter*, unknown sp. affil.**	0.03	<0.01 ± <0.01	0.01	<0.01 ± <0.01
	***Capnocytophaga*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Cellulomonadaceae, unknown gen./sp. affil.	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Comamonadaceae, unknown gen./sp. affil.	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Conchiformibius*, unknown sp. affil.**	0.05	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Corynebacterium*, unknown sp. affil.**	0.13	<0.01 ± 0.02	<0.01	<0.01 ± <0.01
	***Dietzia*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Enterobacteriaceae, unknown gen./sp. affil.	0.18	<0.01 ± 0.02	<0.01	<0.01 ± <0.01
	***Enterococcus*, unknown sp. affil.**	0.23	<0.01 ± 0.03	<0.01	<0.01 ± <0.01
	***Filifactor*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01

	***Fusobacterium*, unknown sp. affil.**	0.22	0.03 ± 0.04	<0.01	<0.01 ± <0.01
	Haemophilus influenzae	0.04	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Haemophilus*, unknown sp. affil.**	0.19	0.02 ± 0.04	<0.01	<0.01 ± <0.01
	***Halomonas*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Intrasporangiaceae, unknown gen./sp. affil.	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Jeotgalicoccus psychrophilus	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Kingella*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Kocuria*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Lachnospiraceae, unknown gen./sp. affil.	0.15	0.03 ± 0.03	0.08	0.04 ± 0.02
	Lactobacillus, unknown sp. affil.	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Leptotrichiaceae, unknown gen./sp. affil.	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Leptotrichia*, unknown sp. affil.**	0.06	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	Leuconostocaceae, unknown gen./sp. affil.	0.08	0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Mannheimia*, unknown sp. affil.**	0.12	0.02 ± 0.02	<0.01	<0.01 ± <0.01
	Microbacteriaceae, unknown gen./sp. affil.	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Mogibacteriaceae, unknown gen./sp. affil.	0.03	0.01 ± 0.01	0.01	<0.01 ± <0.01
	Mogibacterium, unknown sp. affil.	0.03	0.01 ± 0.01	0.01	<0.01 ± <0.01

	***Moraxella*, unknown sp. affil.**	0.11	0.02 ± 0.03	<0.01	<0.01 ± <0.01
	***Mycoplasma*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Neisseria*, unknown sp. affil.**	0.09	0.02 ± 0.02	<0.01	<0.01 ± <0.01
	Neisseriaceae, unknown gen./sp. affil.	0.37	0.06 ± 0.08	<0.01	<0.01 ± <0.01
	Paenibacillus amylolyticus	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Paraprevotellaceae, *Prevotella*, unknown sp. affil.	0.02	<0.01 ± <0.01	0.01	<0.01 ± <0.01
	***Parvimonas*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Pasteurellaceae, unknown gen./sp. affil.	0.27	0.05 ± 0.07	<0.01	<0.01 ± <0.01
	***Pasteurella*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Peptostreptococcaceae, unknown gen./sp. affil.	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Peptostreptococcus anaerobius	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Plesiomonas*, unknown sp. affil.**	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Porphyromonas endodontalis	0.04	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Porphyromonas*, unknown sp. affil.**	0.15	0.02 ± 0.03	<0.01	<0.01 ± <0.01
	Propionibacteriaceae, unknown gen./sp. affil.	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Psychrobacter*, unknown sp. affil.**	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Rhodococcus fascians	0.09	<0.01 ± 0.01	<0.01	<0.01 ± <0.01

	Rikenellaceae, unknown gen./sp. affil.	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Ruminococcaceae, unknown gen./sp. affil.	0.14	0.03 ± 0.03	0.11	0.05 ± 0.03
	Sharpea, unknown sp. affil.	0.04	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	***Sneathia*, unknown sp. affil.**	0.03	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Sporosarcina*, unknown sp. affil.**	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Staphylococcus aureus	0.16	<0.01 ± 0.02	<0.01	<0.01 ± <0.01
	Staphylococcus equorum	0.08	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Staphylococcus*, unknown sp. affil.**	0.07	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	***Stenotrophomonas*, unknown sp. affil.**	0.08	<0.01 ± 0.01	<0.01	<0.01 ± <0.01
	Streptococcus alactolyticus	0.01	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Streptococcus infantis	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Streptococcus luteciae	0.19	0.03 ± 0.04	0.07	<0.01 ± 0.01
	***Streptococcus*, unknown sp. affil.**	0.22	0.05 ± 0.05	<0.01	<0.01 ± <0.01
	Streptophyta, unknown fam./gen./sp. affil.	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Succinivibrio, unknown sp. affil.	0.43	0.02 ± 0.07	0.39	0.04 ± 0.11
	Veillonella parvula	0.02	<0.01 ± <0.01	<0.01	<0.01 ± <0.01
	Weeksellaceae, unknown gen./sp. affil.	0.07	0.01 ± 0.02	<0.01	<0.01 ± <0.01

---

[a]Standard deviation.

[b]This taxon was only eliminated in the manual filtering approach.

TABLE S4 Average relative abundance of OTUs assigned to the genus *Streptococcus* in samples collected *via* buccal swabs (Buccal) or stomach tubing (Rumen). Significant differences were tested for using Student's t-test. Only OTUs that occurred at ≥0.5% in at least one of the 96 samples are shown.

OTU ID	Maximum relative abundance [%]		*P*-value	Taxonomic assignment[a]
	Buccal	Rumen		
denovo1107	0.5	0.3	0.163	*Streptococcus*, unknown sp. affil.
denovo19202	4.7	<0.1	<0.001	*Streptococcus*, unknown sp. affil.
denovo20885	0.5	<0.1	0.014	*Streptococcus*, unknown sp. affil.
denovo22711	0.5	<0.1	0.017	*Streptococcus*, unknown sp. affil.
denovo32198	1.4	0.3	0.100	*Streptococcus*, unknown sp. affil.
denovo32808	0.5	<0.1	0.001	*Streptococcus*, unknown sp. affil.
denovo36449	9.9	<0.1	<0.001	*Streptococcus*, unknown sp. affil.
denovo39368	11.4	<0.1	<0.001	*Streptococcus*, unknown sp. affil.

	denovo39854	0.5	<0.1	<0.001	*Streptococcus*, unknown sp. affil.
	denovo4301	0.5	<0.1	<0.001	*Streptococcus*, unknown sp. affil.
	denovo49218	0.6	<0.1	0.147	*Streptococcus*, unknown sp. affil.
	denovo5468	1.3	<0.1	0.036	*Streptococcus*, unknown sp. affil.
335	denovo56372	0.8	<0.1	0.023	*Streptococcus*, unknown sp. affil.
	denovo57728	1.1	<0.1	0.038	*Streptococcus*, unknown sp. affil.
	denovo63575	1.5	<0.1	0.072	*Streptococcus*, unknown sp. affil.
	denovo67920	4.3	<0.1	<0.001	*Streptococcus*, unknown sp. affil.
	denovo68253	0.5	<0.1	<0.001	*Streptococcus*, unknown sp. affil.
340	denovo69405	4.5	<0.1	0.011	*Streptococcus*, unknown sp. affil.
	denovo71261	14.3	<0.1	0.026	*Streptococcus*, unknown sp. affil.
	denovo72316	3.1	<0.1	0.010	*Streptococcus*, unknown sp. affil.
	denovo80319	1.1	<0.1	0.139	*Streptococcus*, unknown sp. affil.
	denovo14016	1.1	<0.1	0.037	*Streptococcus luteciae*
345	denovo27737	12.9	<0.1	<0.001	*Streptococcus luteciae*
	denovo61036	1.7	<0.1	0.002	*Streptococcus luteciae*

	denovo65546	3.2	<0.1	0.088	*Streptococcus luteciae*
	denovo67014	0.9	<0.1	0.020	*Streptococcus luteciae*
	denovo70070	3.7	3.7	0.241	*Streptococcus luteciae*
350	denovo8955	6.9	<0.1	<0.001	*Streptococcus luteciae*

---

[a]According to the greengenes database version gg_13_5 (ref. 1).

Reference

355

1. **McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P.** 2011. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J **6:**610–618.

TABLE S5 Kruskal-Wallis one-way ANOVA of bacterial taxa at phylum level between the four different diets administered (using data obtained from sampling *via* four different sampling methods). *P*-values were calculated using the Kruskal-Wallis Test with * = $P \leq 0.05$, ** = $P \leq 0.01$, and *** = $P \leq 0.001$. Diets that share superscript letters are not significantly different ($P > 0.05$, Bonferroni *post-hoc* test).

Taxon	*P*-value	Average relative abundance in different diet groups [%]			
		100LS	100MS	25MG	65MG
Bacteroidetes	*	43.0^{b}	48.2^{ab}	45.2^{ab}	53.6^{a}
Cyanobacteria	***	0.9^{a}	0.3^{c}	0.4^{b}	$<0.1^{d}$
Fibrobacteres	*	1.7^{ab}	3.1^{a}	1.2^{b}	2.8^{a}
Firmicutes	***	51.7^{a}	46.0^{ab}	39.1^{b}	28.3^{c}
Proteobacteria	***	0.2^{b}	0.4^{b}	12.7^{a}	14.3^{a}
Spirochaetes	***	1.0^{a}	1.0^{ab}	0.5^{b}	0.8^{b}
Synergistetes	*	0.2^{a}	0.3^{a}	0.3^{a}	0.1^{a}
Tenericutes	***	1.4^{a}	0.7^{b}	0.7^{b}	0.2^{c}

TABLE S6 Kruskal-Wallis one-way ANOVA of bacterial, archaeal, ciliate protozoal, and anaerobic fungal taxa between the four different diets administered (using data obtained from sampling *via* four different sampling methods). *P*-values were calculated from log-transformed relative abundance data. Differences between individual diets were confirmed using Bonferroni *post-hoc* tests. Taxa that showed no significant differences between treatment groups using ANOVA or subsequent Bonferroni *post-hoc* tests are not listed. *P*-values were calculated using the Kruskal-Wallis Test with * = $P \leq 0.05$, ** = $P \leq 0.01$, and *** = $P \leq 0.001$. Diets that share superscript letters are not significantly different ($P > 0.05$, Bonferroni *post-hoc* test).

Taxon	*P*-value	Average relative abundance in different diet groups [%]			
		100LS	100MS	25MG	65MG
Bacteria					
Anaeroplasma, unknown sp. affil.	***	0.6^{a}	0.2^{b}	0.4^{a}	$<0.1^{c}$
Anaerovibrio, unknown sp. affil.	***	$<0.1^{c}$	$<0.1^{bc}$	$<0.1^{ab}$	0.3^{a}
Bacteroidales, unknown fam./gen./sp. affil.	***	7.0^{a}	6.8^{ab}	5.6^{b}	3.3^{c}
BF311, unknown sp. affil.	***	0.1^{a}	0.4^{a}	0.1^{b}	$<0.1^{b}$

	BS11, unknown gen./sp. affil.	***	1.1^{a}	0.6^{b}	0.4^{b}	0.3^{b}
	Bulleidia, unknown sp. affil.	***	0.4^{a}	0.2^{b}	0.5^{a}	1.1^{a}
	Butyrivibrio, unknown sp. affil.	***	4.3^{a}	5.0^{ab}	2.6^{b}	1.3^{c}
	CF231, unknown sp. affil.	***	0.9^{a}	0.9^{a}	0.7^{a}	0.4^{b}
	Christensenellaceae, unknown gen./sp. affil.	***	0.6^{a}	0.3^{b}	0.1^{b}	0.2^{b}
	Clostridiales, unknown fam./gen./sp. affil.	***	11.2^{a}	14.5^{a}	7.7^{b}	4.3^{c}
	Clostridium, unknown sp. affil.	***	0.9^{a}	0.7^{b}	0.5^{bc}	0.3^{c}
	Coprococcus, unknown sp. affil.	**	1.3^{ab}	2.9^{a}	1.1^{b}	1.5^{ab}
	Dialister, unknown sp. affil.	***	<0.1^{bc}	<0.1^{c}	0.1^{b}	0.4^{a}
	Fibrobacter succinogenes	*	1.7^{ab}	3.1^{a}	1.2^{b}	2.8^{a}
	Lachnobacterium, unknown sp. affil.	***	<0.1^{b}	<0.1^{b}	0.1^{b}	1.2^{a}
	Lachnospiraceae, unknown gen./sp. affil.	***	7.2^{a}	4.6^{bc}	6.4^{ab}	3.6^{c}
	Megasphaera, unknown sp. affil.	***	<0.1^{c}	<0.1^{bc}	0.2^{b}	0.7^{a}
	Mogibacteriaceae, unknown gen./sp. affil.	***	2.1^{a}	0.8^{bc}	1.3^{b}	0.7^{c}
	Mogibacterium, unknown sp. affil.	***	1.8^{a}	1.1^{bc}	1.0^{b}	0.6^{c}
	Oscillospira, unknown sp. affil.	***	0.4^{a}	0.4^{ab}	0.5^{a}	0.3^{b}

	p.2534.18B5, unknown gen./sp. affil.	***	0.6^{a}	<0.1^{b}	<0.1^{b}	<0.1^{b}
	Paraprevotella, unknown sp. affil.	**	0.1^{a}	<0.1^{b}	<0.1^{ab}	<0.1^{b}
	Paraprevotellaceae, unknown gen./sp. affil.	***	0.8^{a}	0.6^{b}	0.7^{a}	0.4^{b}
	Prevotella ruminicola	***	0.9^{a}	0.4^{b}	0.6^{a}	<0.1^{c}
	Prevotella, unknown sp. affil.	***	25.6^{b}	28.6^{b}	29.4^{b}	43.9^{a}
	Prevotellaceae, unknown gen./sp. affil.	***	0.1^{c}	0.8^{ab}	0.5^{a}	0.4^{b}
	RF16, unknown gen./sp. affil.	**	1.2^{ab}	1.8^{b}	2.1^{a}	0.9^{b}
	RF39, unknown fam./gen./sp. affil.	***	0.7^{a}	0.5^{b}	0.3^{bc}	0.2^{c}
	Ruminobacter, unknown sp. affil.	***	<0.1^{b}	<0.1^{b}	11.6^{a}	<0.1^{b}
	Ruminococcaceae, unknown gen./sp. affil.	***	10.0^{a}	6.2^{b}	7.8^{b}	4.4^{b}
	Ruminococcus flavefaciens	***	1.2^{a}	1.1^{a}	0.7^{ab}	0.5^{b}
	Ruminococcus, unknown sp. affil.	*	1.9^{b}	3.5^{ab}	3.3^{ab}	4.0^{a}
	S24.7, unknown gen./sp. affil.	**	1.1^{b}	2.4^{a}	2.1^{ab}	2.0^{a}
	Selenomonas ruminantium	***	0.1^{b}	0.1^{b}	0.3^{a}	0.6^{a}
	Sharpea, unknown sp. affil.	**	0.1^{b}	0.6^{a}	0.6^{ab}	0.4^{ab}
	Succinivibrio, unknown sp. affil.	***	0.1^{c}	0.2^{b}	0.3^{b}	14.2^{a}

	Succinivibrionaceae, unknown gen./sp. affil.	***	0.1^{b}	0.1^{b}	0.8^{a}	0.1^{b}
	Treponema, unknown sp. affil.	***	1.0^{a}	1.0^{ab}	0.5^{b}	0.8^{b}
	Veillonellaceae, unknown gen./sp. affil.	**	1.3^{ab}	0.4^{b}	1.4^{a}	0.5^{ab}
	YRC22, unknown sp. affil.	***	2.3^{a}	0.9^{b}	0.6^{b}	0.8^{b}
	YS2, unknown fam./gen./sp. affil.	***	0.9^{a}	0.3^{c}	0.4^{b}	<0.1^{d}
	Archaea					
	Methanimicrococcus blatticola	*	0.8^{ab}	0.7^{ab}	0.1^{b}	1.7^{a}
	Methanobrevibacter gottschalkii clade	***	44.1^{b}	63.9^{a}	46.0^{b}	66.7^{a}
	Methanomassiliicoccaceae Group10 sp.	***	2.7^{a}	0.7^{c}	1.1^{b}	0.7^{c}
	Methanomassiliicoccaceae Group11 sp. BRNA1	**	0.3^{a}	0.1^{ab}	0.3^{a}	0.1^{b}
	Methanomassiliicoccaceae Group12 sp. ISO4H5	**	2.5^{ab}	3.0^{ab}	5.0^{a}	2.4^{b}
	Methanomassiliicoccaceae Group3b sp.	***	0.1^{b}	<0.1^{c}	1.9^{a}	<0.1^{bc}
	Methanomassiliicoccaceae Group4 sp. MpT1	***	2.3^{a}	0.1^{b}	3.2^{a}	0.2^{b}
	Methanomassiliicoccaceae Group9 sp. ISO4G1	**	3.3^{a}	1.5^{b}	2.1^{ab}	2.1^{b}
	Methanosphaera sp. A4	***	0.7^{a}	0.3^{c}	0.7^{bc}	1.8^{ab}
	Methanosphaera sp. Group5	***	11.6^{a}	2.1^{c}	4.2^{bc}	4.8^{b}

	Methanosphaera sp. ISO3F5	***	15.2^{a}	0.8^{c}	14.6^{a}	3.8^{b}
	Methanosphaera stadtmanae	*	0.1^{a}	0.1^{b}	0.2^{ab}	0.6^{ab}
	Ciliate protozoa					
	Anoplodinium-Diplodinium	***	<0.1^{b}	<0.1^{b}	5.4^{a}	0.1^{b}
	Dasytricha	**	1.0^{b}	0.1^{b}	1.2^{a}	0.3^{b}
	Entodinium	***	52.4^{c}	90.2^{a}	41.5^{c}	72.0^{b}
	Epidinium	***	34.1^{a}	6.2^{b}	41.7^{a}	27.4^{a}
	Eudiplodinium	***	12.4^{a}	3.4^{b}	10.0^{a}	0.2^{c}
	Isotricha	*	<0.1^{ab}	<0.1^{ab}	0.1^{a}	<0.1^{b}
	Anaerobic fungi					
	BlackRhino	***	45.1^{a}	8.8^{b}	49.2^{a}	29.9^{a}
	JF423626	***	20.1^{b}	66.2^{a}	27.6^{b}	33.7^{b}
	Piromyces 2	**	<0.1^{a}	<0.1^{b}	<0.1^{b}	1.0^{b}
	Piromyces 3	***	5.3^{a}	0.1^{b}	0.1^{b}	5.0^{b}
	Piromyces 6	***	4.8^{a}	0.3^{b}	<0.1^{b}	0.6^{b}

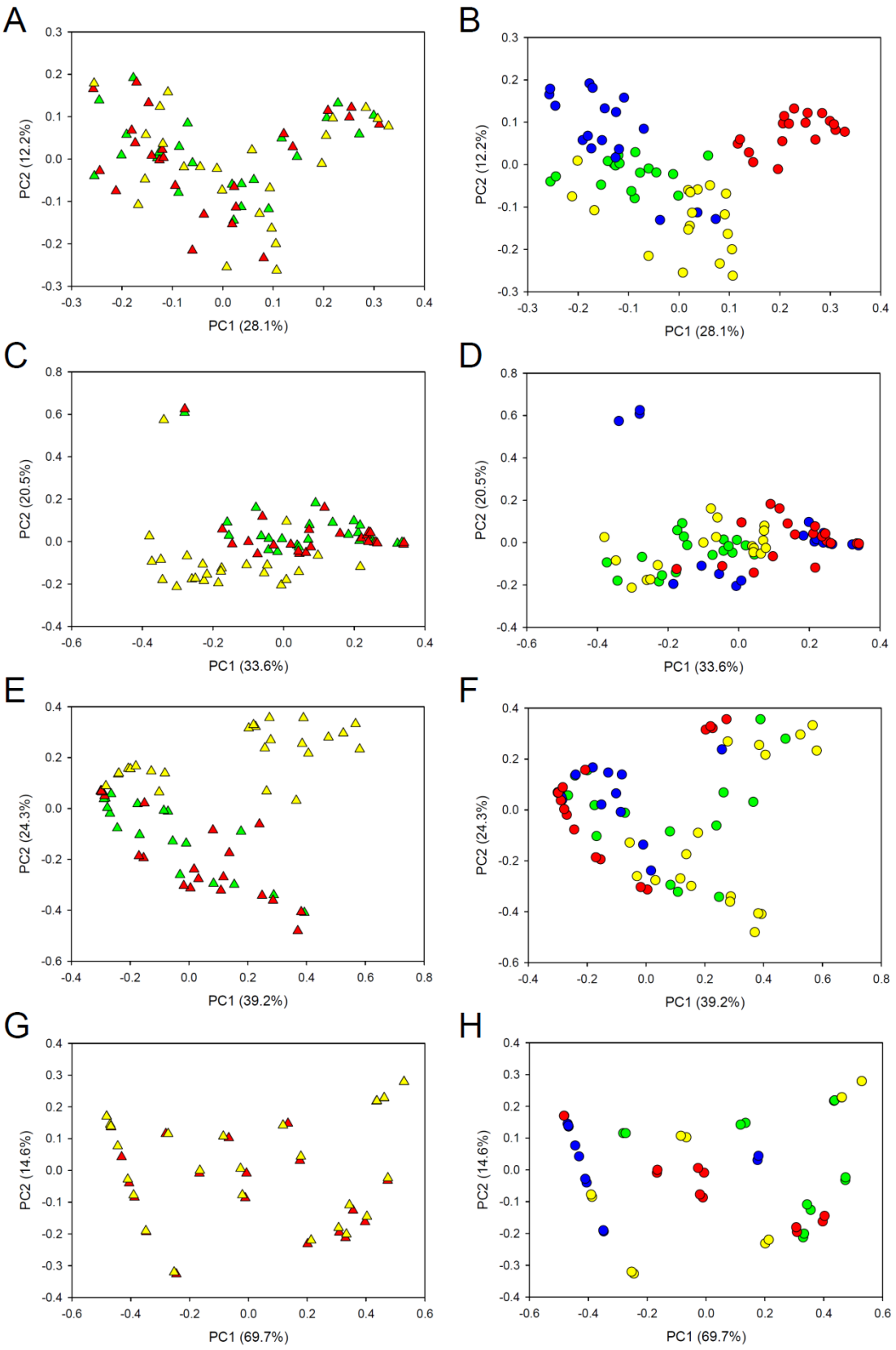
TABLE S7 Kruskal-Wallis one-way ANOVA of bacterial, archaeal, ciliate protozoal, and anaerobic fungal taxa reveals the taxa that were significantly different between the four different sampling methods. *P*-values were calculated from log-transformed relative abundance data. Taxa that showed no significant differences between treatment groups using Bonferroni *post-hoc* tests are not listed. *P*-values were calculated using the Kruskal-Wallis Test with * = $P \leq 0.05$, ** = $P \leq 0.01$, and *** = $P \leq 0.001$. Sampling methods that share superscript letters are not significantly different ($P > 0.05$, Bonferroni *post-hoc* test).

Taxon	*P*-value	Average relative abundance with different sampling methods [%]			
		Buccal OM	Buccal PG	Buccal SD	Rumen
Bacteria					
BS11, unknown gen./sp. affil.	***	0.6^{ab}	0.8^{a}	0.7^{bc}	0.3^{c}
Bulleidia, unknown sp. affil.	**	0.6^{ab}	0.7^{a}	0.5^{b}	0.4^{b}
Christensenellaceae, unknown gen./sp. affil.	**	0.3^{a}	0.4^{a}	0.2^{ab}	0.2^{b}
Clostridium, unknown sp. affil.	**	0.7^{a}	0.7^{a}	0.6^{ab}	0.3^{b}
Coprococcus, unknown sp. affil.	**	2.0^{a}	1.9^{a}	1.7^{a}	1.3^{b}

	Lachnospiraceae, unknown gen./sp. affil.	**	6.3^{a}	6.3^{a}	5.2^{ab}	3.8^{b}
	Lactobacillus, unknown sp. affil.	***	0.8^{a}	0.7^{a}	1.7^{a}	<0.1^{b}
	Mogibacteriaceae, unknown gen./sp. affil.	***	1.6^{a}	1.5^{a}	1.3^{a}	0.5^{b}
	Mogibacterium, unknown sp. affil.	***	1.4^{a}	1.4^{a}	1.3^{a}	0.5^{b}
	Oscillospira, unknown sp. affil.	***	0.5^{ab}	0.6^{a}	0.3^{bc}	0.1^{c}
	Paludibacter, unknown sp. affil.	***	1.5^{a}	1.3^{a}	1.3^{ab}	0.1^{b}
	Paraprevotellaceae, *Prevotella*, unknown sp. affil.	***	1.4^{ab}	2.6^{a}	1.0^{bc}	0.2^{c}
	Prevotellaceae, *Prevotella*, unknown sp. affil.	***	27.0^{b}	26.0^{b}	35.4^{a}	39.1^{a}
	RF16, unknown gen./sp. affil.	*	1.0^{b}	0.9^{ab}	1.0^{b}	3.1^{a}
	RFN20, unknown sp. affil.	**	0.3^{ab}	0.3^{ab}	0.3^{b}	1.1^{a}
	Ruminococcaceae, unknown gen./sp. affil.	**	8.8^{a}	8.9^{a}	5.6^{ab}	5.1^{b}
	Sharpea, unknown sp. affil.	***	0.2^{ab}	0.9^{a}	0.5^{b}	<0.1^{c}
	Streptococcus alactolyticus	***	0.6^{a}	1.0^{a}	2.0^{a}	<0.1^{b}
	Succiniclasticum, unknown sp. affil.	***	1.7^{ab}	2.4^{a}	1.7^{a}	1.0^{b}
	Archaea					
	Methanimicrococcus blatticola	*	0.8^{a}	1.0^{ab}	1.3^{ab}	<0.1^{b}

	Methanobrevibacter ruminantium clade	*	20.8^{a}	18.8^{ab}	20.1^{ab}	15.6^{b}
	Methanobrevibacter smithii	***	<0.1^{b}	0.1^{b}	0.1^{b}	0.3^{a}
	Methanocorpusculum sp.	***	0.7^{a}	0.7^{a}	0.2^{b}	<0.1^{b}
	Methanomassiliicoccaceae Group8 sp. WGK1	***	0.4^{ab}	0.8^{a}	0.2^{b}	<0.1^{b}
	Methanomassiliicoccaceae Group9 sp. ISO4G1	*	1.9^{ab}	3.5^{a}	1.8^{b}	1.8^{b}
	Methanosphaera sp. A4	***	1.2^{a}	1.4^{a}	0.6^{ab}	0.2^{b}
	Methanosphaera stadtmanae	*	0.5^{a}	0.3^{ab}	0.1^{b}	0.1^{b}
	Ciliate protozoa					
	Eudiplodinium	*	9.2^{ab}	1.9^{ab}	3.2^{b}	11.7^{a}
	Anaerobic fungi					
	Piromyces 2	***	<0.1^{b}	<0.1^{b}	<0.1^{b}	1.0^{a}
	Piromyces 7	*	<0.1^{b}	0.5^{b}	<0.1^{ab}	0.1^{a}

FIG S1 Principal coordinate analysis using the Bray-Curtis dissimilarity metric of A. and B. bacterial communities, C. and D. archaeal communities, E. and F. ciliate protozoal, and G. and H. anaerobic fungal communities in 24 samples sequenced using different sequencing technologies (triangles) and obtained by stomach tubing from sheep feeding on four different diets (circles). Each point represents one sample. Red triangle = MiSeq Read 1, yellow triangle = MiSeq Read 2, green triangle = 454 Titanium, red circle = 65 MG, green circle = 100 LS, blue circle = 100 MS, yellow circle = 25 MG. No 454 Titanium sequence data was available for the comparison of anaerobic fungal communities. The left and right panels show the same plots, with the points colored in different ways.


A
PC2 (12.2%)
PC1 (28.1%)
B
PC2 (12.2%)
PC1 (28.1%)
C
PC2 (20.5%)
PC1 (33.6%)
D
PC2 (20.5%)
PC1 (33.6%)
E
PC2 (24.3%)
PC1 (39.2%)
F
PC2 (24.3%)
PC1 (39.2%)
G
PC2 (14.6%)
PC1 (69.7%)
H
PC2 (14.6%)
PC1 (69.7%)

510

Figure S2. Principal coordinate analysis using the Bray-Curtis dissimilarity metric of bacterial communities in 96 samples, collected *via* four different sampling methods (triangles, left panel) from sheep feeding on four different diets (circles, right panel) without exclusion of potential oral taxa. Sequence analysis was performed using 454 Titanium chemistry. Each point represents one sample. Red triangle = Buccal PG, green triangle = Rumen, blue triangle = Buccal OM, yellow triangle = Buccal SD, red circle = 65 MG, green circle = 100 LS, blue circle = 100 MS, yellow circle = 25 MG. The left and right panels show the same plots, with the points colored in different ways.

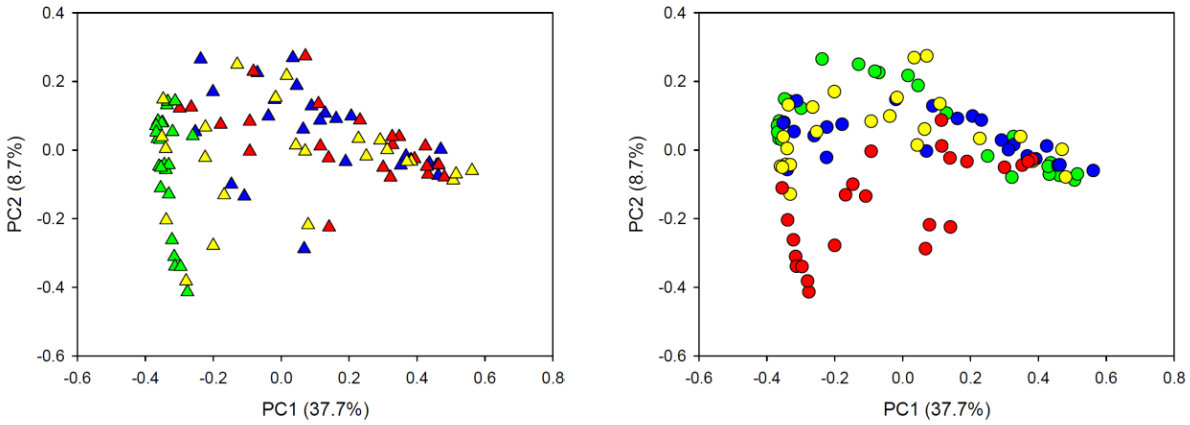


Figure S3. Heatmap of average bacterial communities in six samples collected *via* buccal swabs (OM, PG and SD) and stomach tubing (RM) from sheep feeding on four different diets (100LS, 25MG, 65MG or 100MS). Each column represents the average of six samples collected from six different animals. Strong red colors indicate high standardized relative abundance values (row Z-scores), while dark blue colors indicate low standardized relative abundance values. Samples and taxa were clustered using Pearson correlation and hierarchical clustering with the average linkage method.