Simultaneous Amplicon Sequencing to Explore Co-Occurrence Patterns of Bacterial, Archaeal and Eukaryotic Microorganisms in Rumen Microbial Communities

Sandra Kittelmann1,2, Henning Seedorf2, William A. Walters3, Jose C. Clemente4, Rob Knight4,5, Jeffrey I. Gordon2, Peter H. Janssen1*

1 AgResearch Ltd, Grasslands Research Centre, Palmerston North, New Zealand, 2 Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, Missouri, United States of America, 3 Department of Molecular, Cellular, and Developmental Biology, University of Colorado, Boulder, Colorado, United States of America, 4 Department of Chemistry and Biochemistry and BioFrontiers Institute, University of Colorado, Boulder, Colorado, United States of America, 5 Howard Hughes Medical Institute, Boulder, Colorado, United States of America

Abstract

Ruminants rely on a complex rumen microbial community to convert dietary plant material to energy-yielding products. Here we developed a method to simultaneously analyze the community’s bacterial and archaeal 16S rRNA genes, ciliate 18S rRNA genes and anaerobic fungal internal transcribed spacer 1 genes using 12 DNA samples derived from 11 different rumen samples from three host species (Ovis aries, Bos taurus, Cervus elephas) and multiplexed 454 Titanium pyrosequencing. We show that the mixing ratio of the group-specific DNA templates before emulsion PCR is crucial to compensate for differences in amplicon length. This method, in contrast to using a non-specific universal primer pair, avoids sequencing non-targeted DNA, such as plant- or endophyte-derived rRNA genes, and allows increased or decreased levels of community structure resolution for each microbial group as needed. Communities analyzed with different primers always grouped by sample origin rather than by the primers used. However, primer choice had a greater impact on apparent archaeal community structure than on bacterial community structure, and biases for certain methanogen groups were detected. Co-occurrence analysis of microbial taxa from all three domains of life suggested strong within- and between-domain correlations between different groups of microorganisms within the rumen. The approach used to simultaneously characterize bacterial, archaeal and eukaryotic components of a microbiota should be applicable to other communities occupying diverse habitats.

Introduction

Globally, ruminant-derived methane (CH4) contributes 8% of total anthropogenic greenhouse gas emissions and represents an energy loss from feed of up to 10% for the animal [1,2]. Despite ongoing efforts, no viable CH4 mitigation technology exists for pasture-fed ruminants [3]. Ruminant animals rely on a complex intestinal microbiota to convert ingested plant material into energy-yielding products. This symbiosis between the host and rumen microbiota has evolved to make use of grasses with a high content of structural carbohydrates and low nutritional value. Cellulolytic bacteria, anaerobic fungi and some ciliate protozoa are the initial colonisers and decomposers of such fibrous feed [4,5]. The breakdown of plant polymers is followed by fermentation of the resulting monomers and oligomers. Fermentation end products, such as short-chain fatty acids, are then absorbed by the host across the rumen wall. Methanogenic archaea are beneficiaries of the hydrogen (H2) generated during fermentation of feed. Their metabolism allows them to gain energy by reducing carbon dioxide (CO2) with electrons from H2 oxidation, thereby...
producing CH₄. CH₄ cannot be used by the ruminant and is removed from the rumen by eructation.

To date, deep-sequencing studies of the rumen using phylogenetic marker genes have concentrated on investigating bacterial community composition [6,7,8,9,10,11,12,13]. Although scientific interest in the structure and function of eukaryotic microbes is increasing [14], rumen studies are largely focused on bacterial and archaecal members. Protozoa and fungi, although far less abundant than the bacteria and archaea in terms of cell numbers, can make up approximately half of the total rumen microbial biomass [15]. Until recently, species identification and physiological characteristics of fungal and ciliate communities colonising the gastrointestinal tracts of herbivores have typically been assessed by using microscopy [16,17] and cultivation-based techniques [18,19].

Ligenztoffer et al., used metagenomic methods to characterize anaerobic fungal members in the guts of herbivores: they did so by focusing on the internal transcribed spacer (ITS) domain positioned between eukaryotic ribosomal RNA genes, collecting more than 250,000 partial internal transcribed spacer 1 (ITS1) from 33 ruminant and non-ruminant herbivores by multiplex pyrosequencing [20]. Eight novel anaerobic fungal lineages were detected, which accounted for 38.3% of the total number of sequences.

In the present study, we use barcoded pyrosequencing of phylogenetic marker genes to simultaneously characterize diversity in all three domains of life represented in the rumen microbiota: bacteria, archaea, and eukarya (ciliate protozoa and anaerobic fungi). We developed a method that permits adjustment of the template concentration to compensate for species diversity, abundance, and amplicon length of the different microbial groups, and furthermore minimizes nonspecific amplification of DNA originating from ingested plant material and endophytic fungi. Clone libraries constructed in earlier studies enabled us to build phylogenetic frameworks for ciliated protozoa [21] and anaerobic fungi [22] facilitating taxonomic classification of pyrosequencing reads. Our approach maximizes the number of samples that can be analyzed while still achieving satisfactory coverage of diversity, thereby increasing the power of comparisons between experimental treatments and minimizing costs of monitoring microbial communities.

**Materials and Methods**

**Rumen sampling**

The use of animals, including welfare, husbandry, experimental procedures, and the collection of rumen samples used for this study, was approved by the AgResearch Grasslands, AgResearch Ruakura, and Massey University Animal Ethics Committees, and complied with the institutional Codes of Ethical Conduct for the Use of Animals in Research, Testing and Teaching, as prescribed in the Animal Welfare Act of 1999 and its amendments. Rumen samples were collected as part of a series of feeding trials conducted in New Zealand under permit numbers 06/119 and 06/126 [Massey University, Palmerston North], 11110 modification 775, 12174, and 11975 (all three AgResearch, Grasslands Research Centre, Palmerston North), and 11897 (AgResearch, Ruakura Research Centre, Hamilton). The animals were kept at AgResearch’s Grasslands Research Centre and at Massey University, Palmerston North, at Woodlands farm, Invercargill, and at Lyefarm, Hamilton.

Eleven rumen samples were collected from sheep (*Ovis aries*), cattle (*Bos taurus*) or deer (*Cervus elephas*) feeding on a variety of different diets by sampling through a fistula, via stomach tubing or after sacrifice under supervision of a veterinarian at AgResearch’s Ruakura Research Centre (Table 1). The rumen samples have the following naming convention: Ruminant species abbreviation (C = cow, D = deer, S = sheep), animal identity number, diet abbreviation (GR = grain, HY = hay, PA = pasture, SG = summer pasture, SI = silage, WG = winter pasture), location abbreviation (PN = Palmerston North, TK = Taranaki, RK = Ruakura); for example, sample C5SGPN was obtained from cow 5 feeding on summer pasture in Palmerston North. Additional information on diet, flock, and location of all animals sampled in this study is provided in Table 1 (for more details also see [21,23,24]).

**DNA extraction**

DNA was extracted from 30 mg of homogenized, freeze-dried and ground rumen sample after Lueders et al. [25]. Briefly, cells were disrupted by combined bead-beating and phenol-chloroform-isooamyl alcohol (25:24:1; vol:vol:vol) treatment and subsequent precipitation of proteins with chloroform-isooamyl alcohol (24:1; vol:vol). DNA was precipitated with 2 volumes of 30% (wt/vol) polyethylene glycol, washed with 70% (vol:vol) ice-cold ethanol and eluted in 100 μl elution buffer (EB; 10 mM Tris, pH 8.5 with HCl). RNA was digested with 50 μg RNase A, and DNA was subsequently cleaned up using the QIAquick PCR purification kit (Qiagen, Hilden, Germany) and eluted in 50 μl EB. Note that duplicate extractions were performed in the case of sheep S4 from Palmerston North on summer pasture in order to estimate the comparability of sub-samples analyzed by 454 Titanium pyrosequencing; these two samples were designated S4SG1PN and S4SG2PN.

**Amplification of target genes and amplicon pooling**

Primers used for PCR amplification of bacterial and archaecal 16S rRNA genes, ciliate 18S rRNA genes and fungal ITS1 genes are listed in Table 2. In addition to the group specific primer pairs, we also tested a primer pair that simultaneously amplifies bacterial and archaecal 16S rRNA genes (ArBa; Table 2; [26]). All primers contained the 454 Life Sciences (Branford, CT, USA) adaptors A ([5’-CCA TCT CAT CCC TGC GTG TCT CCG ACT CAG-3’]) or B ([5’-CTC ATC CCC TGC GTG CCT TGG CAG TCT CAG-3’]) for Titanium sequencing, and a unique 12-base error-correcting barcode was attached to adaptor A for sample identification [27]. Most primers carried a two-base linker sequence between the barcode and the group-specific primer to avoid differential amplification (Table 2; [28]). Amplification was performed as follows: initial denaturation at 95°C for 2 min; a defined number of cycles (30 for bacteria/archaea/protozoa; 35 for fungi) of denaturation (95°C, 20 s), annealing (for temperatures see Table 2, 20 s) and elongation (65°C, 1 min); and a final 7-min extension at 65°C. For each DNA sample, 76 μl of reaction mix was prepared containing 32 μl of 5 PRIME HotMasterMix (5 PRIME, Gaithersburg, MD, USA) and barcoded and non-barcoded primer in water, each to a final concentration of 0.2 μM. Before addition of template DNA, an aliquot of 19 μl was transferred into a sterile tube to serve as a no-template negative control. The remaining 57 μl of reaction mix were spiked with a total of 120 ng of DNA contained in 3 μl of water, and then divided into 3 aliquots of 20 μl each. After amplification, the triplicate PCR products were pooled, and the correct sizes of PCR products and the absence of signal from negative controls were verified by agarose gel electrophoresis. PCR products of the ArBa primer pair simultaneously targeting bacteria and archaea were purified individually using the Agencourt AMPure XP kit (Beckman Coulter, Brea, CA, USA), quantified using the Quant-iT dsDNA BR assay kit (Invitrogen, Carlsbad, CA, USA) and a
Score was samples by examining the 8- or 12-nucleotide error-correcting (default: –H 6). Sequence reads were assigned to corresponding fungal ITS1 sequences the option –H 8 was passed to allow for a sequence length of 150 bp. For homopolymer-rich anaerobic here, filter settings were adjusted to allow for a minimum extremely short sequences obtained with the ArS primer pair.

Phylogenetic analysis of pyrosequencing reads

Samples were processed and analyzed following the procedure described by Caporaso et al. [29] using QIIME v1.2.1. Sequences >200 bp in length were truncated so that the average quality score was >25 and only sequences without ambiguous characters were included in the analysis. An exception was made for the extremely short sequences obtained with the ArS primer pair. Here, filter settings were adjusted to allow for a minimum sequence length of 150 bp. For homopolymer-rich anaerobic fungal ITS1 sequences the option –H 8 was passed to allow for a maximum length of a homopolymer run of eight base pairs (default: –H 6). Sequence reads were assigned to corresponding samples by examining the 8- or 12-nucleotide error-correcting Hamming or Golay barcodes, respectively. OTU picking was performed using uclust [30] for bacteria and archaea at 97% similarity threshold or the prefix-suffix method (QIIME team, unpublished) passing the option “–p 1000” for protozoa and fungi. Sequence data were given phylogenetic assignments as follows: bacterial 16S rRNA genes were blasted against the greengenes database (gg_97_otus_4feb2011.fasta; [31]); archaeal 16S rRNA genes [32], ciliate 18S rRNA genes [21], and fungal ITS1 genes [22]; version 2.1) were blasted against rumen specific, in-house databases derived from earlier studies (available from the authors upon request). To test whether the databases constructed for anaerobic fungi and ciliate protozoa gave correct taxonomic assignments, two sequence files were created. The only difference between the two sequence files was the order of sequences. In one file, reference sequences were sorted by accession number in alphabetical order, whereas in the other file, the sequences were sorted by accession number in reverse-alphabetical order. Blasting sequences against these two different sequence files resulted in unambiguous taxonomic assignments for almost all sequences. However, a small number of sequences gave ambiguous taxonomic assignments but the differences in abundances caused by these ambiguous assignments were negligible (Table S1).

Quantitative PCR of bacterial and archaeal 16S rRNA genes in rumen samples

Quantitative PCR (qPCR) of bacterial and archaeal 16S rRNA genes was carried out as described earlier [21,23]. Briefly, 3 dilutions (1:10, 1:20 and 1:40) were prepared for each DNA sample, and reactions were set up in duplicate for each dilution.

Statistical analyses

Differences between bacterial and archaeal communities obtained by using different sets of primers were calculated in QIIME using the Bray-Curtis dissimilarity distance metric, which
takes into account presence or absence as well as abundance of OTUs.

Unpaired and paired Student’s t-tests using a two-tailed distribution were performed in Excel (Microsoft Corp., Redmond, WA, USA) to evaluate the significance of between-primer/within-sample and within-primer/between-sample dissimilarities of all primer pairs targeting bacteria and archaea, and to evaluate bias for or against certain groups of archaea by the three primer pairs tested (ArL, ArS, and ArBa), respectively.

Comparison of results from pyrosequencing libraries constructed with the ArBa primer pair with those of qPCR of bacterial and archaeal 16S rRNA genes in the same samples was carried out using the CORREL and FDIST functions in Excel, and statistics of fit were calculated according to Goodman [33].

Statistical analyses including rarefaction curves, beta-diversity estimates (Bray-Curtis, Sørensen-Dice) and UPGMA cluster analyses were conducted using the QIIME pipeline [29]. Beta-diversity was evaluated using family level taxa for bacteria [31] and at an operational clade level for archaea [32]. For the comparison of primer pairs, groups showing an abundance of less than 1% of the total community were excluded from the analysis. Rarefaction analysis was performed using rarefied OTU tables (rarefied to the lowest number of reads obtained for any of the 12 DNA samples analyzed), 100 replications, and cut-offs of 97% and 95% sequence similarity, respectively [34].

Simpson’s index of dominance (1-D) describes the diversity of a community with 1 (= 100%) indicating maximum diversity in a sample. This index was calculated according to Simpson [35] using the software PAST [36].

Co-occurrence analysis between microbial populations was performed in the R studio v0.94.110 [37]. Only microbial groups that represented ≥1% of the total community within each of the four microbial groups (bacteria, archaea, protozoa or fungi) in at least one sample and that were detected in ≥50% of rumen samples were included in the analysis. Spearman’s rank correlations and P-values were calculated and plotted using the packages ellipse [38], hmisc [39], and corrplot [40].

### Results and Discussion

#### Diversity of microbial groups and amplicon mixing

To study the composition of bacterial, archaeal and eukaryotic components of the rumen microbiota in New Zealand ruminants, we used the data collected with the primer sets giving the longest amplicons, as these allow for the best possible phylogenetic resolution. Rarefaction analysis was carried out at arbitrary OTU cut-offs of 95% and 97% similarity to estimate the average number

<table>
<thead>
<tr>
<th>Name</th>
<th>Adaptor</th>
<th>(Linker* +Adapter) Primer sequence [5’→3’]</th>
<th>Region</th>
<th>Length [bp]</th>
<th>Annealing [°C]</th>
<th>Abbreviation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ArBa515F</td>
<td>B</td>
<td>(TA) GTG CCA GCM GGG GCG GTA A</td>
<td>16S</td>
<td>~290</td>
<td>52</td>
<td>ArBa</td>
<td>[26]</td>
</tr>
<tr>
<td>ArBa806R</td>
<td>A⁵⁺</td>
<td>(AC) GGA CTA CVS GGG TAT CTA AT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ba9F</td>
<td>B</td>
<td>GAG TTT GAT CMT GGC TCA G</td>
<td>16S</td>
<td>~525 (+73)</td>
<td>52</td>
<td>BaL</td>
<td>[58]</td>
</tr>
<tr>
<td>Ba515Rmod1</td>
<td>A⁵⁺</td>
<td>CCG CGG CKG CTG GCA C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>modified from [59]</td>
</tr>
<tr>
<td>Ba27F</td>
<td>B</td>
<td>AGA GTT TGA TCC TGG CTC AG</td>
<td>16S</td>
<td>~365</td>
<td>52</td>
<td>BaS</td>
<td>[60]</td>
</tr>
<tr>
<td>Ba338R</td>
<td>A⁵⁺</td>
<td>TGC TGC CTC CGG TAG GAG T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ar915aF</td>
<td>A⁵⁺</td>
<td>AGG AAT TGG CGG GGG AGC AC</td>
<td>16S</td>
<td>~492 (+73)</td>
<td>59</td>
<td>ArL</td>
<td>[62]</td>
</tr>
<tr>
<td>Ar1386R</td>
<td>B</td>
<td>GGC GTG TGT GCA AGG AGC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ar344F</td>
<td>A⁵⁺</td>
<td>ACG GGG YGC AGC AGG CGG GA</td>
<td>16S</td>
<td>~148 (+73)</td>
<td>60</td>
<td>ArS</td>
<td>[64]</td>
</tr>
<tr>
<td>Ar519R</td>
<td>B</td>
<td>TTA CGG CGG CKG CTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RP841F</td>
<td>B</td>
<td>(AA) GAC TAG GGA TTG GAG TGG</td>
<td>18S</td>
<td>~511 (+73)</td>
<td>54</td>
<td>RP</td>
<td>[21]</td>
</tr>
<tr>
<td>Reg1302R</td>
<td>A⁵⁺</td>
<td>(TC) AAT TGC AAA GAT CTA TCC C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MN100F</td>
<td>A⁵⁺</td>
<td>TCC TAC CCT TTG TGA ATT TG</td>
<td>ITS1</td>
<td>~250 (+73)</td>
<td>50</td>
<td>RF</td>
<td>[67]</td>
</tr>
<tr>
<td>MNGM2</td>
<td>B</td>
<td>CTG CTG CCT TCA TCG TTG CG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Linker sequences were designed according to Walters et al. [28].

Golay barcodes of 12-nucleotides in length were used in combination with this primer.

Hamming barcodes of 8-nucleotides in length were used in combination with this primer.

doi:10.1371/journal.pone.0047879.t002

---

**Table 2.** Oligonucleotide primers used to amplify bacterial and archaeal 16S rRNA genes, ciliate 18S rRNA genes and anaerobic fungal ITS1 genes.
of sequences of each microbial group that had to be sampled to describe phylogenetic diversity across all samples [34]. Rarefaction curves of the archaeal, ciliate and anaerobic fungal communities started to level off markedly after approximately 1,000 sequences had been sampled (Figure S1). Therefore, we concluded that archaeal, ciliate and anaerobic fungal diversity in rumen samples would be adequately compared by generating 1,000 sequences per sample. In contrast, rarefaction curves of bacteria never leveled, even after 5,000 sequences had been sampled (Figure S1). However, our data show that approximately 5,000 sequences are sufficient to distinguish between bacterial communities in different samples (Figures S2A, S2B and S3A).

Amplicons were mixed at different ratios to compensate for the different degrees of diversity within the four microbial groups and for the differing amplicon lengths. Mixing only the long amplicons in a ratio of 1:1:1 resulted in a similar number of assignable reads for bacteria (74,938; 598 bp amplicon) and archaea (86,502; 565 bp amplicon), but only 34,369 reads for ciliate protozoa (584 bp amplicon; Figure 1). We then added the fungal ITS1 amplicons to analyse all four groups simultaneously. To allow a more extensive sampling of their higher phylogenetic diversity, we added 5 times the amount of bacterial template DNA as compared to archaea, ciliate protozoa and anaerobic fungi (5:1:1:1). From this, we obtained 67,264 reads for bacteria, 24,881 reads for archaea, 8,917 reads for ciliate protozoa, and 23,637 reads for anaerobic fungi (210–290 bp amplicon; Figure 1). Although the fungal templates were 5 times underrepresented as compared to the bacterial templates, the sequence yield was similar for these two groups. This result is likely to be due to a preferential amplification of smaller products during emulsion PCR [41]. Fungal diversity in the rumen, however, is much lower than that of the Bacteria and comparable to that of the archaea and ciliate protozoa [22]. Because we did not need the high level of coverage of fungal sequences, we therefore performed a third mixing ratio test, in which we maintained the ratio of bacteria, archaea and ciliate protozoa but reduced the amount of fungal templates (5:1:1:0.2). This ratio resulted in 115,283 reads for bacteria, 24,881 reads for archaea, 8,917 reads for ciliate protozoa, and 23,637 reads for anaerobic fungi (Figure 1). Figure S4 shows the distribution of pyrosequencing reads across the different amplicon lengths for both mixing ratios tested.

An amplicon mixing ratio of 5:1:1:0.2 achieved an approximate average of 5,000 sequence reads for bacteria, and 1,000 sequence reads for archaea, ciliate protozoa and anaerobic fungi for the 12 rumen DNA samples analyzed in this study. This ratio would allow up to 100 rumen samples to be processed for a single 454 pyrosequencer run (≥400 amplicons; 4 microbial groups ×100 samples; assuming a total yield of 1,000,000 reads per run).

Simultaneous analysis of microbial communities from all three domains of life has been performed before [42]. However, the authors used a single universal primer pair [43] and recovered 71.9% of bacterial, 7.8% of archaeal, and 5.5% of eukaryotic sequences. 14.8% of sequences obtained were of chloroplast or mitochondrial origin. In the rumen environment, simultaneous whole community pyrosequencing using amplicon mixing holds several advantages compared to using a single universal primer pair. First, this approach allows for the selection of different target genes that provide sufficient phylogenetic resolution for all groups studied, rather than relying on one conserved gene for all groups of interest. Second, it provides specificity against amplification of unwanted co-extracted DNA, such as plant- or endophyte-derived DNA. Third, amplicon mixing allows compensation for three important variables: (i) the abundance of each microbial group in the environment studied; (ii) the degree of diversity of each of the groups analyzed; and (iii) the amplicon lengths of selected marker genes. However, simultaneous analysis of amplicons of very
different lengths requires careful control of all steps in the process, and it is easy to get massive over- or under-representation of shorter amplicons.

The between-run variation, measured as the dissimilarity between two identical DNA samples in two different runs (mean dissimilarity 5.5%±0.7% (standard deviation)), was smaller than when the same rumen sample was extracted in duplicate and amplicons analyzed in the same 454 run (Text S1; compare S4SG1PN and S4SG2PN in Figure S5; 6.3%±1.2%).

Libraries generated using three different primer pairs that amplified 16S rRNA genes from members of the domain Bacteria clustered by sample rather than by primer pair, indicating that all primer pairs resulted in similar bacterial community compositions (Text S2, Figures S2A, S2B, S3A, Tables S2, S3). In contrast, when samples were amplified with different primer pairs targeting the domain Archaea, the three libraries generated from each of the samples clustered considerably further apart from each other than did bacterial communities from the same samples (Text S2, Figures S2C, S2D, S3B, S6, Tables S2, S4).

The overall trend of bacterial, and thus archaeal abundances in the 12 pyrosequencing libraries constructed using the ArBa primer pair was correlated with the results obtained from qPCR (Pearson correlation coefficient = 0.83, P<0.001; Text S3, Figure S7), indicating that generalized between-sample comparisons can be validly made from data generated using the ArBa primer pair.

Composition of bacterial, archaeal and eukaryotic communities in the rumen

To analyse the phylogenetic composition of microbial communities in the 12 rumen DNA samples, data obtained from three sequencing reactions (sequencing runs 2 and 3) per primer set were combined into a single file for each microbial group. Data for Bacteria (BaL), Archaea (ArL) and ciliate protozoa (RP) resulted from sequencing runs 2 (half B) and 3 (both halves). Data for anaerobic fungi (RF) resulted from sequencing runs 2 (half A) and 3 (both halves). All four data files were analyzed individually using the QIIME software package [29]. In total, we analyzed 257,485 reads for Bacteria (a mean of 21,457 reads per sample; range 16,554–44,221 reads), 125,052 reads for Archaea (mean 10,421; range 7,827–15,998), 45,231 reads for ciliate protozoa (mean 3,769; range 2,785–5,374), and 186,485 reads for anaerobic fungi (mean 15,540; range 7,053–24,438).

Bacterial communities in the rumen samples were analyzed at the family level and all had a similar diversity (range of Simpson’s index of dominance (1-D) = 75.4–85.2%).

Figure 2. Relative composition of microbial communities in the 12 pyrosequencing libraries. Community composition of (A) bacteria, (B) archaea, (C) ciliate protozoa, and (D) anaerobic fungi in 12 DNA samples obtained from three different ruminant species feeding on a variety of natural diets.
doi:10.1371/journal.pone.0047879.g002
were classified into the family Curtiss). The largest groups of bacteria in these 11 rumen samples comparisons (81.2% using Sørensen-Dice; 94.6% using Bray-Curtis) shared the highest similarity of all between-sample Sørensen-Dice and S4SG2PN, extracted from the same rumen sample 49.7–94.6%; Figure 2A). The two replicates of DNA samples 53.5%–81.2%; mean similarity using Bray-Curtis 73.2%, range similarity between samples using Sørensen-Dice 66.7%, range composition was very similar across all samples analyzed (mean similarity between samples using Sørensen-Dice 79.2%, range 62.5%-100%; mean similarity using Bray-Curtis 77.3%, range 54.2–95.3%). The replicate DNA samples S4SG1PN and S4SG2PN from the same rumen sample shared the highest similarity of all between-sample comparisons (100% using Sørensen-Dice; 95.3% using Bray-Curtis). Sequences clustering within the Methanobrevibacter ruminantium clade (30.2%±10.6%), M. gottschalkii clade (32.4%±12.9%), Rumen Cluster C (18.1%±6.0%), and the genus Methanosphaera (9.3%±5.5%) were most abundant (Figure 2B). Besides these groups, sample S4SGPN from a red deer consisted of a large number of sequences derived from Methanosarcina species (12.2%). Out of nine Methanosarcina-related OTUs that had an abundance of ≥10 reads in this sample (making up 950 from a total of 1,905 Methanosarcina-related reads), 50.7% and 32.5% were ≥99% similar to homologous regions of 16S rRNA genes from M. barkeri and M. mazei, respectively. This finding agrees with the detection of a significant population of Methanosarcina spp. in the same animal, a deer on pasture in summer using different methods [25]. Methanosarcina spp. are usually rare in the rumen [32], because of their low growth rates, despite the abundance of acetate which is a growth substrate for these methanogens. Methanosarcina spp. are known to occur in large numbers when rumen liquid turnover is low [44]. Interestingly, deer are reported to slow down the passage rate of particles through the rumen in summer but not in winter, compared to other ruminants [45].

composition was very similar across all samples analyzed (mean similarity between samples using Sørensen-Dice 66.7%, range 53.5%–81.2%; mean similarity using Bray-Curtis 77.3%, range 49.7–94.6%; Figure 2A). The two replicates of DNA samples S4SG1PN and S4SG2PN, extracted from the same rumen sample S4SGPN, shared the highest similarity of all between-sample comparisons (81.2% using Sørensen-Dice; 94.6% using Bray-Curtis). The largest groups of bacteria in these 11 rumen samples were classified into the family Lachnospiraceae, which made up 32.4% of all bacterial 16S rRNA gene sequences amplified (mean = 24.0%), and the family Prevotellaceae, at 11.6 to 44.9% of all sequences (mean = 22.9%). These bacteria apparently play a major role in feed degradation in the rumen. The other abundant groups of bacteria were members of the families Ruminococcaceae (3.7 to 11.6%; mean = 8.1%) and Fibrobacteraceae (1.9 to 15.1%; mean = 6.1%), and less well classified groups of the orders Clostridiales (1.8 to 14.8%; mean = 8.6%) and Bacteroidales (3.5 to 27.3%; mean = 13.7%). These six largest groupings accounted for 77.3 to 89.2% of all sequences in all samples (mean = 83.6%), and were detected in all 11 rumen samples. However, in all cases, their relative abundances varied, by up to 8-fold over the samples, showing that there were some differences in bacterial community structure between the different ruminants, herds or flocks, and diets.

There was an apparent systematic tradeoff between the abundance of members of the order Clostridiales (families Lachnospiraceae, Ruminococcaceae, Veillonellaceae, and Clostridiales Family XIII plus less well defined members of the order Clostridiales) on the one hand and members of the order Bacteroidales (families Prevotellaceae and Porphyromonadaceae plus less well defined Bacteroidales; Figure 3A). This limited diversity of rumen bacteria mirrors that found by others. Using 16S rRNA gene amplicon pyrosequencing, Pitta et al. detected bacterial species belonging to a total of only 13 different family-level taxa in cows feeding on Bermudagrass hay or wheat [8]. Similarly, bacterial communities in rumen fluid of cattle feeding on a high grain diet were composed of species belonging to only 13 different families [7]. In agreement with our study, microorganisms belonging to the Clostridiales and Bacteroidales were also the predominant bacterial groups in steers and cows feeding on grass-legume hay or on switchgrass [6,11].

In the present study, minor groups detected included Cyanobacteria (2.2%±2.1%, mean ± standard deviation), Proteobacteria (1.5%±0.9%), Actinobacteria (1.4%±0.7%), and Spirochaetes (1.3%±1.4%). Chloroplasts contributed 91.1%±5.2% to total cyanobacterial sequence types (in samples where Cyanobacteria made up ≥1% of the total bacterial community). No chloroplasts were detected in samples of animals fed a grain- or silage-based diet, in which fewer chloroplasts were expected anyway because of the nature or the treatment of the feed.

The diversity of methanogenic archaea was measured at “operational clade” level, and 1-D ranged from 58.0–73.5%. As with Bacteria, archaeal community composition was similar across all samples analyzed (mean similarity between samples using Sørensen-Dice 79.2%, range 62.5%-100%; mean similarity using Bray-Curtis 77.9%, range 54.2–95.3%). The replicate DNA samples S4SG1PN and S4SG2PN from the same rumen sample shared the highest similarity of all between-sample comparisons (100% using Sørensen-Dice; 95.3% using Bray-Curtis). Sequences clustering within the Methanobrevibacter ruminantium clade (30.2%±10.6%), M. gottschalkii clade (32.4%±12.9%), Rumen Cluster C (18.1%±6.0%), and the genus Methanosphaera (9.3%±5.5%) were most abundant (Figure 2B). Besides these groups, sample S4SGPN from a red deer consisted of a large number of sequences derived from Methanosarcina species (12.2%). Out of nine Methanosarcina-related OTUs that had an abundance of ≥10 reads in this sample (making up 950 from a total of 1,905 Methanosarcina-related reads), 50.7% and 32.5% were ≥99% similar to homologous regions of 16S rRNA genes from M. barkeri and M. mazei, respectively. This finding agrees with the detection of a significant population of Methanosarcina spp. in the same animal, a deer on pasture in summer using different methods [25]. Methanosarcina spp. are usually rare in the rumen [32], because of their low growth rates, despite the abundance of acetate which is a growth substrate for these methanogens. Methanosarcina spp. are known to occur in large numbers when rumen liquid turnover is low [44]. Interestingly, deer are reported to slow down the passage rate of particles through the rumen in summer but not in winter, compared to other ruminants [45].
Of the 11 rumen samples analyzed in this study, 8 were also analyzed using denaturing gradient gel electrophoretic separation of partial 16S rRNA amplicons and by clone libraries using different primers by Jeyanathan et al. [23]. The pyrosequencing approach showed that the same methanogen groups were present in these samples. The increased sample size using pyrosequencing indicated the presence of other groups not detected by Jeyanathan et al. [23], but these made up only 0.63% of all pyrosequencing reads in those libraries.

Ciliate communities in the samples analyzed showed a broad range of diversity at the genus level (range 1-D = 0.4–76.7%), and ciliate community structure was very different between the different samples analyzed (41.4% average between-sample similarity using Bray-Curtis; Figure 2C). Interestingly, the highest similarity between samples was not obtained for the two replicates of sample S4SG1PN (92.3%), but for samples S19PAPN and S4SGPN (96%). This result can be explained by the low diversity of samples S19PAPN (1-D = 0.4%) and S5GRRK (9.6%) as compared to the much higher diversity in samples S4SG1PN (73.1%) and S4SG2PN (71.1%). Species of the genus Entodinium were the most abundant group of ciliates, with an average abundance of 40%, confirming earlier reports based on microscopic counting [46]. Four of the 11 rumen samples analyzed showed A-type ciliate communities [47], which are characterized by the presence of Polyplastron multivesiculatum (C5SGPN, C5SIPN, C6HYPN, and S2SIPN). These communities were dominated by Ostracodinium, Dasytricha, and Entodinium species. B-type ciliate communities, characterized by the presence and dominance of Epidinium and Eudiplodinium spp. [47], were observed in six rumen samples (C7PATK, D3WGPN, D4SGPN, S4SGPN, and S4WGPN). The two remaining rumen samples harboured O-type ciliate communities [47], consisting almost entirely of Entodinium species (≥94.4%; S19PAPN and S5GRRK). These findings confirm that the ciliate community types identified by Eadie are robust and reproducible across different samples.

Figure 4. Spearman’s rank correlation matrix of the dominant microbial populations across domains in analyzed rumen samples.
Microbial populations listed represent at least 1% of the bacterial, archaeal, ciliate, or fungal communities in at least one sample and were detected in at least 50% of the rumen samples analyzed. Strong correlations are indicated by large squares, whereas weak correlations are indicated by small squares. The colours of the scale bar denote the nature of the correlation with 1 indicating perfect positive correlation (dark blue) and -1 indicating perfect negative correlation (dark red) between two microbial populations. Correlations marked with circles are discussed in the text.
that a bacterial flagellum protein induced a highly specific up-
variation could also be the result of highly-specific interactions
Methanobrevibacter gottschalkii
these methanogens are being sequenced, including members of the
isoenzymes [53]. Genomes of more ruminal representatives of
Caecomyces
PLOS ONE | www.plosone.org 9 February 2013 | Volume 8 | Issue 2 | e47879
spp. and Fibrobacter spp. are both primary cellulose degraders in the
Ruminococcus spp. produce large amounts of H2, while
methanol ([56]; J. Jeyanathan, R. S. Ronimus, S. O. Hoskin, P. H. Jansen, unpublished data). It seems likely that both groups proliferate when methanol availability is high, and that they somehow partition this niche, suggesting co-existence when a particular resource (methanol) is available.
As expected, co-occurrence analysis revealed a strong positive relationship between the rumen ciliate genera Polyplastron and Ostrocodinium (R = 0.95, P = 1.8×10−11; Figure 4), two typical members of A-type ciliate communities in New Zealand ruminants [21]. Significant positive correlation was also observed between Endotricha, a key genus of B-type ciliate communities and the Epidinium-Diplodinium cluster, species of which are commonly found in B-type ciliate communities (R = 0.76, P = 1.4×10−6; Figure 4). Epidinium was also positively correlated with Epistylus (R = 0.58, P = 7.7×10−8; Figure 4), another key genus of B-type ciliate communities.
Significant correlations were further observed between certain ciliate proteozoa and anaerobic fungi. Ciliates of the genus Isotricha appeared to be negatively correlated with anaerobic fungi of the BlackRhino group (R = −0.87, P = 8.1×10−5; Figure 4). In contrast, ciliate proteozoa belonging to the genus Epistylus showed a strong positive relationship with Neocallimastix 1-group fungi (R = 0.81, P = 6.4×10−8; Figure 4).

Co-occurrence of microbial populations in the rumen
Microorganisms from all three domains of life form a complex network in the rumen ecosystem to ferment the fibrous plant material ingested by the ruminant. The different microbial groups in the rumen have been studied separately for various aspects such as the influences of diet, ruminant host species and other factors on the microbial populations present. To the best of our knowledge, only a single study exists where community composition of bacteria, archaea, ciliate proteozoa and anaerobic fungi in the rumen was studied as a whole [48], but the use of different analysis techniques for each microbial group hindered the detection of potential cross-domain interactions.
Here, we describe microbial co-occurrence patterns in the analyzed rumen samples using the combined data sets (as described in the previous section). Across all samples, our analysis showed a general negative correlation between methanogens of the Methanobrevibacter ruminantium clade and those of the Methanobrevibacter gottschalkii clade (R = −0.51 [Spearman’s rank correlation coefficient], P = 0.023 [Spearman’s p-value]; Figure 3B), i.e., a relative increase in one group resulted in the decrease of the other group. Both of these groups are H2-utilizing methanogens, and presumably compete within the rumen [32]. Interestingly, there was a good positive correlation (R = 0.72, P = 6.2×10−5; Figure 4) between the occurrence of methanogens in the Methanobrevibacter ruminantium clade and bacteria in the family Fibrobacteraceae. In contrast, occurrence of methanogens in the Methanobrevibacter gottschalkii clade was positively correlated (R = 0.90, P = 4.9×10−14; Figure 4) with bacteria in the family Ruminococcaceae. Ruminococcus spp. and Fibrobacter spp. are both primary cellulose degraders in the rumen [49]. Ruminococcus spp. produce large amounts of H2 while Fibrobacter spp. produce formate [50]. Both H2 and formate are substrates for ruminal methanogens [51]. The co-variation of the two groups of methanogens with different bacteria suggests that the methanogens may be adapted to different ruminal H2 regimes or respond to differences in available CH4 precursors. Methanobrevibacter ruminantium M1, for example, appears to be specialized to low H2 concentrations based on the presence of only methyl coenzyme M reductase (Mer) and the absence of the methyl coenzyme M reductase II (MerII) soenzyme [52]. MerI and MerII are expressed at low- and high levels of H2, respectively, and members of the genus Methanobrevibacter usually contain both soenzymes [33]. Genomes of more ruminal representatives of these methanogens are being sequenced, including members of the Methanobrevibacter gottschalkii clade [54]; the resulting genomic data may provide further insights into these occurrence patterns. Co-variation could also be the result of highly-specific interactions between certain methanogens and bacteria. It was recently shown that a bacterial flagellum protein induced a highly specific up-
regulation of genes in methanogenesis [53], indicating that such targeted interactions between methanogens and bacteria may exist. Such interactions could facilitate a better transfer of H2 between the partners, benefitting both.
We also detected a strong positive correlation (R = 0.66, P = 2.5×10−13; Figure 4) between the occurrence of members of the archaeal Rumen Cluster C group (as defined by Janssen and Kirs [32]) and members of the methanogen genus Methanosphaera. This could indicate cooperation between the groups, or that they can coexist when their niche is expanded. Methanosphaera spp. and Rumen Cluster C archaea both produce CH4 from H2 plus methanol ([56]; J. Jeyanathan, R. S. Ronimus, S. O. Hoskin, P. H. Jansen, unpublished data). It seems likely that both groups proliferate when methanol availability is high, and that they somehow partition this niche, suggesting co-existence when a particular resource (methanol) is available.

Conclusions
The approach we have described for analyzing the structure of bacterial, archaeal and eukaryotic microbial communities in the rumen ecosystem, in which all members of the community are simultaneously detected by large-scale pyrosequencing should have broad applications. It allows adjustments to accommodate differences in the abundance and diversity of different groups of microorganisms. Moreover, amplicons of different lengths can be combined by compensating for the different lengths, since shorter fragments appear to amplify preferentially. This type of study is not restricted to the analysis of small subunit ribosomal RNA genes, but can also be applied to investigate the environmental distribution of functional marker genes within the read length limit of available sequencing platforms.
Further investigation of the observed patterns of co-occurrence in the rumen may provide new clues about metabolic networks between rumen-inhabiting microbial groups and may resolve their individual contributions to overall rumen functioning. We also believe that looking beyond the bacteria and archaea and into the entirety of microbial communities, including eukaryotic microorganisms, will provide valuable new insights into the functioning of this and other microbial communities occupying diverse habitats.

Supporting Information
Figure S1 Rarefaction analysis of rumen microbial groups.
(PDF)
Figure S2 Comparison of primers for amplification of bacterial and archaeal 16S rRNA genes from rumen samples.

Table S3 Comparison of abundances of bacterial families detected by using 3 different primer sets (BaL, BaS, and ArBa) in all analyzed samples.

Figure S4 Lengths of obtained sequencing reads for two different mixing ratios of the microbial groups analyzed.

Table S4 Comparison of archaeal clades detected using 3 different primer sets (ArL, ArS, and ArBa) in all analyzed samples.

Figure S5 Bacterial community comparison of samples analyzed in two different sequencing runs using primer set BaL.

Text S1 Run-to-run variation.

Table S2 Ranges and averages of the numbers of phyla, classes, orders, and families detected across all DNA samples.

Text S2 Diversity and coverage of bacterial and archaeal communities analyzed with different primer combinations.

Text S3 Estimation of the proportion of archaea obtained with the non-specific ArBa primer pair using qPCR and pyrosequencing libraries.

Acknowledgments

We are indebted to Jill Manchester (Center for Genome Sciences and Systems Biology; 454 Titanium pyrosequencing) and Michelle Kirk (AgResearch; rumen sample preparation) for excellent technical assistance.

Author Contributions

Conceived and designed the experiments: SK PHJ. Performed the experiments: SK HS. Analyzed the data: SK HS WAW JCC PHJ. Wrote the paper: SK HS WAW JCC RK JIG PHJ.

References


